



The relative efficacy of topical non-steroidal anti-inflammatory drugs and capsaicin in osteoarthritis

By Dr Monica SM Persson, BMBS

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy

April 2019

Declaration

I hereby declare that this thesis is the result of original research. It has largely been conducted by myself and any assistance received is detailed below. It has not already been accepted for any degree, diploma, or other qualification. All authors and works to which reference has been made are fully acknowledged.

Study design, literature search, application for ethics approval, recruitment, trial management, data collection, analysis, and thesis writing were conducted primarily by myself. Professors Weiya Zhang, Michael Doherty, and David Walsh provided guidance and supervision in all aspects of the present work. In addition, Professor Weiya Zhang conducted the sample size calculations for the n-of-1 trial series. Dr Joanne Stocks assisted with data validation for the meta-analysis and conducted the physical examinations for the n-of-1 trial series. Dr Aliya Sarmanova conducted the ultrasound examinations for the n-of-1 trial series and calculated the ultrasound inter-machine agreement. Dr Gwen S Fernandes conducted the radiographic scoring for the n-of-1 trial series. Drs Haruna Takahashi, Neele Dellschaft, and Yu Fu conducted data validation for non-English trials. Dr Danielle Burke provided statistical advice for the analysis of the individual patient data meta-analyses, in particular regarding the model specifications of the one-stage approach. Additional statistical support was received from Professor Richard Riley (individual patient data meta-analysis), Dr Boliang Guo (one-stage individual patient data meta-analysis), Professor Tony Ades (Bayesian network meta-analysis), Dr Kun Zou (overall treatment effect and proportion of contextual effects), Dr Andrea Venn (general), and Dr Archan Bhattacharya (general).

Abstract

Background

Pain is often the most troubling issue for people with osteoarthritis (OA) and is typically the focus of pharmacological management. A multitude of treatment options are available for OA, but these produce, on average, only small to moderate reductions in pain. Often the modest reductions in pain do not outweigh the safety concerns of traditional analgesics. Topical non-steroidal anti-inflammatory drugs (NSAIDs) and capsaicin are safe alternatives, with different mechanisms of action, that are widely used for knee and hand OA. However, there is a paucity of evidence to guide the choice between the treatments. Their relative efficacies are unknown and it is unclear if certain people would benefit more from one treatment than the other. The identification of predictors of response to the treatments is essential in order to maximise treatment effects and implement a precision medicine approach in OA care.

Objectives

[1] To determine the relative efficacy of topical NSAIDs and capsaicin in OA, [2] to identify predictors of response that may allow topical therapy to be tailored to the individual, and [3] to implement and appraise methods that guide precision medicine in OA.

Methods

Systematic literature searches were conducted for randomised controlled trials (RCTs) of topical NSAIDs or capsaicin in OA and neuropathic pain (NP). NP was chosen as a surrogate for phenotypes with predominantly neuropathic-like pain. Average treatment effects were examined using aggregate data, extracted from eligible publications. Placebo-controlled RCTs were pooled in random-effects conventional meta-analyses (CMA) for the specific (difference between treatment and placebo) and overall (improvement from baseline in the treatment group) treatment effects. Network meta-analysis (NMA) compared the treatments using placebo as a common comparator. Aggregate data effect sizes (ES) were presented as Hedges' g.

Predictors of response were examined using individual patient data meta-analysis (IPD MA) and a pilot n-of-1 trial series. Raw IPD from eligible RCTs, acquired through author contact and direct enquiry to pharmaceutical companies, were pooled in IPD MA. Predictors of the specific (treatment-by-covariate interactions) and overall (prognostic factors) treatment effect were investigated. Participants with radiographic, painful knee OA were recruited to an open label pilot n-of-1 trial series. Participants were randomly allocated to three treatment cycles consisting of alternating periods of 5% ibuprofen gel and 0.025% capsaicin (Zacin) cream. Treatment periods were four weeks. Treatment preferences were determined at an individual level and responses were pooled across participants to examine the variation in responses and potential predictors of response.

Results

Of 63 topical NSAID and 10 capsaicin RCTs identified in the systematic literature search, CMA and NMA were based on 21 (n=6,191) and five (n=415) RCTs for topical NSAIDs and capsaicin, respectively. Moderate level evidence suggested that topical NSAIDs were effective for pain relief in OA (21 RCTs, ES: 0.31, 95% confidence interval [CI] 0.20 to 0.41). Very low level evidence suggested that capsaicin was only effective when used at the 0.025% concentration (4 RCTs, ES 0.41, 95%CI 0.17 to 0.64; ES of all capsaicin RCTs: 5 RCTs, ES 0.28, 95%CI -0.04 to 0.60). Absolute levels of pain relief from the treatments, compared to baseline, were substantial (overall effect: topical NSAIDs ES 1.23, 95%CI 1.06 to 1.41; capsaicin ES 1.05, 95%CI 0.52 to 1.57). Very low level evidence in NMA indicated that, on average, topical NSAIDs and capsaicin were equally effective (ES 0.04, 95% credible interval [CrI] -0.29 to 0.37).

IPD were provided for 15 (n=3,889) topical NSAID RCTs, including 11 (n=3,140) placebo-controlled trials. No IPD were obtained for capsaicin. Nine potential predictors of response to topical NSAIDs were examined. A significant but small interaction was observed between sex and topical NSAID use, with women reporting a greater effect (11 RCTs, 2,939 participants, p=0.023). Women also reported larger

overall levels of pain relief on topical NSAIDs (15 RCTs, 1,857 participants, $p=0.008$), but this could be explained by baseline pain severity.

Twenty-two participants enrolled in the pilot n-of-1 trial series and completed 104 treatment periods. Clinically important pain relief (≥ 1 point on 0-10 numeric rating scale) was achieved in 64% of the completed periods. Women reported greater pain reduction from baseline irrespective of treatment ($p=0.031$), but this could be explained by baseline pain severity. No difference in pain relief was observed, on average, between topical ibuprofen and capsaicin ($p=0.271$). Individual responses to treatment varied, with approximately 60% of people preferring one treatment over the other, but without clear predictors for the preference. Twenty-five potential predictors of response were examined, of which fibromyalgia severity, distal pressure pain thresholds, temporal summation, synovial hypertrophy, and quadriceps strength inconsistently associated with a preference for one treatment over the other.

Very low level evidence suggested that low-dose capsaicin cream was not effective for NP (7 RCTs, ES -0.18, 95%CI -0.90 to 0.53). No data were available for the CMA of topical NSAIDs in NP. In addition, insufficient IPD were received for NP participants and it was not possible to examine for predictors of response to topical NSAIDs or capsaicin in NP.

Conclusion

Topical NSAIDs and capsaicin may both effectively relieve OA pain, but treatment effects from group comparisons do not directly translate to treatment outcomes at the individual level. Patients with different pain mechanisms may respond differently to these two agents, but definitive trial evidence is still needed. Presently, patients may benefit from trying both treatments to determine which is better for them.

To better guide the care of an individual, evidence synthesis should encompass a spectrum of methods, from study-level to participant-

level evidence. Further development of IPD MA and n-of-1 trial series methodology is required to successfully guide precision medicine.

List of publications

Publications based on PhD work

Published papers

1. Persson, M.S.M., Fu, Y., Bhattacharya, A., Goh, S.L., van Middelkoop, M., Bierma-Zeinstra, S.M., Walsh, D., Doherty, M. and Zhang, W., 2016. Relative efficacy of topical non-steroidal anti-inflammatory drugs and topical capsaicin in osteoarthritis: protocol for an individual patient data meta-analysis. *Systematic reviews*, 5(1):165.
2. Persson, M.S.M., Stocks, J., Walsh, D.A., Doherty, M. and Zhang, W., 2018. The relative efficacy of topical non-steroidal anti-inflammatory drugs and capsaicin in osteoarthritis: A network meta-analysis of randomised controlled trials. *Osteoarthritis and cartilage*, 26(12): 1575-1582

Presentations at conferences

1. 7th Annual International Conference on Evidence Based Healthcare (November, 2018): Evidence Based Medicine in osteoarthritis – from research evidence to individualised care (oral presentation)
2. UK-RiME Showcase (Oxford, 2017): Topical non-steroidal anti-inflammatory drugs and capsaicin in osteoarthritis (oral and poster presentation)

Other publications from the period of this thesis

1. Persson, M.S.M., Sarmanova, A., Doherty, M. and Zhang, W., 2018. Conventional and biologic disease-modifying anti-rheumatic drugs for osteoarthritis: a meta-analysis of randomized controlled trials. *Rheumatology*, 57(10):1830-1837
2. Persson, M.S.M., Sarmanova, A., Doherty, M. and Zhang, W., 2018. Comment on: Conventional and biologic disease-modifying anti-rheumatic drugs for osteoarthritis: a meta-analysis of randomized controlled trials: reply. *Rheumatology*. 57(11):2060-2061

3. Zeng, C., Wei, J., Persson, M.S.M., Sarmanova, A., Doherty, M., Xie, D., Wang, Y., Li, X., Li, J., Long, H. and Lei, G., 2018. Relative efficacy and safety of topical non-steroidal anti-inflammatory drugs for osteoarthritis: a systematic review and network meta-analysis of randomised controlled trials and observational studies. *Br J Sports Med*, 52(10): 642-650

Acknowledgments

First and foremost, I would like to thank my supervisors, Professors Weiya Zhang, Michael Doherty, and David Walsh, for their guidance, support, and confidence in me throughout my PhD. I couldn't have asked for a better supervisory team. You believed in me enough to offer me this PhD opportunity and it has truly changed my life – I count myself very lucky to be passionate about the work I have done.

My sincere thank you goes to the staff and PhD students in Academic Rheumatology who figuratively and literally enriched my days by sharing my interest in research and cake. In particular, I would like to thank the following people:

- Kehinde Akin-Akinyosoye for being a constant source of support, an inspiration, and a welcome distraction from work. It has been a pleasure to go through the PhD journey together
- Joanne Stocks for her help throughout my thesis and the many conversations we have had about the highs and lows of our research
- Jane Healy for her endless patience in answering all my administrative questions and for looking out for me when I was away on home visits
- Helen Richardson for sharing her past experience and guiding me through the complexities of setting up a study
- Rose Farrands-Bentley for managing my trial grant and helping me with all sorts of queries – always with a smile
- Wendy Jenkins for her advice on the practicalities of running a trial – from obtaining informed consent to safely conducting home visits

My thanks also go to the University of Nottingham and Arthritis Research UK Pain Centre for funding my PhD, the OA Trial Bank for their collaboration, the Nottingham University Hospital Charity for funding my pilot n-of-1 trial series, my individual patient data meta-analysis collaborators for generously sharing their data, and to all the participants who gave up their time to make my research a reality.

I would like to thank my friends for being supportive enough to ask about my research and kind enough to listen to my response. Special thanks should go to my family and, in particular, to my parents and brother. Your belief and pride in me is strong enough to motivate me even through the stressful times. Finally, I want to thank my boyfriend, Dave, for supporting my decision to do a PhD, for the hours spent listening to my presentations, for patiently talking through my work with me, and for being my rock throughout.

Table of Contents

Declaration	i
Abstract	ii
List of publications	vi
Acknowledgments	viii
Table of Contents	x
List of tables	xxi
List of figures	xxvi
List of abbreviations	xxx
1 CHAPTER 1: Introduction	1
1.1 Introduction to Osteoarthritis (OA)	1
1.1.1 History, definition, and nature	1
1.1.2 Epidemiology	4
1.1.2.1 Incidence and prevalence	4
1.1.2.2 Disease burden	5
1.1.2.2.1 Individual costs	5
1.1.2.2.2 Socioeconomic costs	5
1.1.2.3 Risk factors	6
1.1.2.3.1 Systemic risk factors	7
1.1.2.3.2 Local risk factors	8
1.1.3 Diagnosis	10
1.1.3.1 Clinical presentation	10
1.1.3.2 Criteria for diagnosis	13
1.1.4 Pathology	16
1.1.4.1 Pathophysiology	16
1.1.4.1.1 The role of inflammation	20
1.1.4.2 Pain mechanisms	22
1.1.4.2.1 Peripheral	23
1.1.4.2.2 Central	25
1.1.4.2.3 Neuropathic-like pain	27
1.1.4.2.3.1 Neuropathic pain (NP)	27
1.1.4.2.3.1.1 Clinical presentation of NP	28
1.1.4.2.3.1.2 Criteria for diagnosis of NP	28

1.1.4.2.3.1.3	Pain mechanisms in NP.....	29
1.1.4.2.3.1.4	Management of NP.....	29
1.1.4.2.3.1.5	Topical NSAIDs and capsaicin in NP.....	30
1.1.5	Management of OA.....	31
1.1.5.1	Non-pharmacological.....	33
1.1.5.2	Pharmacological	33
1.1.5.3	Surgical.....	35
1.1.6	Outcome measures for OA.....	37
1.1.6.1	Pain.....	38
1.1.6.2	Disability.....	40
1.1.6.3	Global assessment	41
1.1.6.4	Quality of life	42
1.1.6.5	Structural	42
1.1.6.6	Others	43
1.2	Introduction to topical NSAIDs.....	44
1.2.1	History and Definitions.....	44
1.2.2	Possible mechanism of action.....	45
1.2.2.1	Topical mechanism of action in OA	46
1.2.3	Recommendations and efficacy.....	48
1.2.3.1	Recommendations	48
1.2.3.2	Efficacy	49
1.2.4	Safety.....	53
1.2.4.1	Local AEs.....	53
1.2.4.2	Systemic AEs.....	54
1.3	Introduction to capsaicin	56
1.3.1	History and definitions	56
1.3.2	Possible mechanism of action.....	58
1.3.2.1	Mechanism of action in OA	59
1.3.3	Recommendations and efficacy.....	60
1.3.3.1	Recommendations	60
1.3.3.2	Efficacy	61
1.3.4	Safety.....	64

1.4	Introduction to precision medicine	65
1.4.1	Background	65
1.4.2	Methods for identifying treatment responders	67
1.4.3	Precision medicine in OA	69
1.4.3.1	Responders to topical NSAIDs and capsaicin in OA...	70
1.5	Justifications for thesis	75
2	CHAPTER 2: General methods	77
2.1	General meta-analysis methods	77
2.1.1	Identification of studies	77
2.1.1.1	Systematic literature search	78
2.1.1.1.1	Search terms and strategies	78
2.1.1.1.2	Databases	78
2.1.1.1.3	Analysing the literature search output	79
2.1.1.2	Identification of other studies	79
2.1.2	Inclusion and exclusion criteria	79
2.1.2.1	Study design	80
2.1.2.2	Participants	80
2.1.2.3	Interventions	81
2.1.2.4	Comparators	81
2.1.2.5	Outcomes	81
2.1.3	Data collection	81
2.1.3.1	Data extraction	82
2.1.3.1.1	Data extraction form	82
2.1.3.1.2	Outcomes extracted	82
2.1.3.1.3	Non-English published reports	83
2.1.3.2	Data validation	83
2.1.3.3	Study quality assessment	84
2.1.3.3.1	Modified risk of bias tool	85
2.1.3.4	Data management	86
2.1.3.4.1	Microsoft Access database	86
2.1.3.4.2	Data entry, checks, and calculations	86
2.1.3.5	Data for analysis	88
2.1.3.5.1	Outcome hierarchy for analysis	88

2.2	Meta-analyses of published reports.....	91
2.2.1	Conventional meta-analysis.....	91
2.2.1.1	Concept	91
2.2.1.2	Application	93
2.2.2	Network meta-analysis	94
2.2.2.1	Concept	94
2.2.2.2	Application	96
2.3	Individual patient data meta-analysis	97
2.3.1	Concept	97
2.3.2	Application	99
2.4	Pilot n-of-1 trial series	100
2.4.1	Concept	100
2.4.2	Application	101
3	CHAPTER 3: Study-level efficacy and relative efficacy of topical NSAIDs and capsaicin in OA.....	102
3.1	Overview	102
3.2	Methods	102
3.2.1	Conventional meta-analysis.....	102
3.2.1.1	Specific treatment effect	103
3.2.1.1.1	Formulae	106
3.2.1.1.1.1	Grouping multiple time points per study.....	106
3.2.1.1.1.2	Calculating ES from dichotomous outcomes ...	106
3.2.1.2	Overall treatment effect and PCE.....	108
3.2.1.2.1	Formulae	110
3.2.1.2.2	Meta-analysis methods	111
3.2.2	Network meta-analysis	112
3.2.2.1	Bayesian analysis	113
3.2.2.2	Frequentist analysis	113
3.2.3	GRADE assessment of quality	114
3.3	Results	116
3.3.1	Description of studies	116
3.3.1.1	Topical NSAIDs	116

3.3.1.1.1 Literature search	116
3.3.1.1.2 Description of the studies	118
3.3.1.1.3 Risk of bias assessment.....	124
3.3.1.2 Capsaicin	125
3.3.1.2.1 Literature search	125
3.3.1.2.2 Description of the studies	127
3.3.1.2.3 Risk of bias assessment.....	129
3.3.2 Specific treatment effects	130
3.3.2.1 Topical NSAIDs versus placebo (CMA)	130
3.3.2.1.1 At or nearest to four weeks	130
3.3.2.1.1.1 Cumulative meta-analysis	131
3.3.2.1.1.2 Publication bias and small study effects	132
3.3.2.1.1.3 Sensitivity analyses	134
3.3.2.1.1.3.1 By study quality.....	134
3.3.2.1.1.3.2 By combining continuous and dichotomous outcomes	134
3.3.2.1.1.4 Subgroup analysis by duration of treatment ...	137
3.3.2.1.1.4.1 Meta-regression of sample size and duration of treatment.....	138
3.3.2.2 Capsaicin versus placebo	139
3.3.2.2.1 At or nearest to four weeks	139
3.3.2.2.1.1 Cumulative meta-analysis	141
3.3.2.2.1.2 Publication bias and small study effects	142
3.3.2.2.1.3 Sensitivity analyses	143
3.3.2.2.1.4 Subgroup analysis by capsaicin dose	143
3.3.2.2.1.5 Subgroup analysis by duration of treatment ...	144
3.3.3 Topical NSAIDs versus capsaicin (NMA)	146
3.3.3.1 Description of the trial network	146
3.3.3.1.1 Subgroup analysis by recommendation	146
3.3.3.2 Topical NSAIDs versus capsaicin	147
3.3.3.2.1 Subgroup analysis by recommendation	148
3.3.4 Overall treatment effect and PCE (CMA)	149
3.3.4.1 Topical NSAIDs	149
3.3.4.1.1 At or nearest to four weeks	149
3.3.4.1.2 Sensitivity analyses	151

3.3.4.1.3	Subgroup analysis by drug class	152
3.3.4.1.4	Subgroup analysis by duration of treatment	153
3.3.4.2	Capsaicin	154
3.3.4.2.1	At or nearest to four weeks	154
3.3.4.2.2	Sensitivity and subgroup analyses	154
3.3.4.3	Topical NSAIDs relative to capsaicin	156
3.4	Summary of findings	157
3.5	Discussion	158
3.5.1	Key findings and interpretation.....	158
3.5.1.1	Topical NSAIDs in OA	158
3.5.1.2	Capsaicin in OA.....	160
3.5.1.3	Topical NSAIDs versus capsaicin	161
3.5.2	Caveats	162
3.5.3	Methodological comments	163
3.5.3.1	Estimation of specific treatment effects	163
3.5.3.2	Examination of duration of treatment and treatment effect.....	165
3.5.3.3	Analysis of overall treatment effect and PCE	165
4	CHAPTER 4: Predictors of response to topical NSAIDs and capsaicin in OA	167
4.1	Overview	167
4.2	Methods	167
4.2.1	Data collection and preparation for analysis.....	167
4.2.1.1	Data acquisition from trial custodians.....	167
4.2.1.1.1	Contacting authors	168
4.2.1.1.2	Contacting pharmaceutical companies	169
4.2.1.1.2.1	For published trials	169
4.2.1.1.2.2	For unpublished trials	169
4.2.1.2	Data transfer.....	169
4.2.1.2.1	Data transfer agreements	169
4.2.1.2.2	Data sharing procedures.....	170
4.2.1.3	Data clean and amalgamation	171
4.2.2	Statistical analysis	172

4.2.2.1	General approaches.....	172
4.2.2.1.1	Forest plot generation.....	173
4.2.2.1.2	One-stage modelling.....	174
4.2.2.1.2.1	Specific treatment effect	174
4.2.2.1.2.2	Predictors of the specific treatment effect	175
4.2.2.1.2.3	Overall treatment effect.....	176
4.2.2.1.2.4	Predictors of overall treatment effect	176
4.2.2.1.3	Two-stage modelling.....	177
4.2.2.2	Potential predictors	178
4.2.2.2.1	Participant-level covariates	178
4.2.2.2.2	Study-level covariates	179
4.3	Results	180
4.3.1	Literature search and data acquisition.....	180
4.3.1.1	Topical NSAIDs	180
4.3.1.2	Capsaicin	184
4.3.2	Predictors of response	185
4.3.2.1	Topical NSAIDs	185
4.3.2.1.1	Description of studies and participants	185
4.3.2.1.1.1	Specific effect trials.....	188
4.3.2.1.1.2	Overall effect trials	189
4.3.2.1.1.3	Risk of bias assessment	191
4.3.2.1.2	Treatment effect of topical NSAIDs	192
4.3.2.1.2.1	Specific effect.....	192
4.3.2.1.2.1.1	Repeated measures.....	193
4.3.2.1.2.2	Overall effect.....	195
4.3.2.1.2.2.1	Repeated measures.....	196
4.3.2.1.3	Predictors of response to topical NSAIDs	197
4.3.2.1.3.1	Predictors of specific effect.....	198
4.3.2.1.3.1.1	Repeated measures.....	199
4.3.2.1.3.2	Predictors of overall effect.....	202
4.3.2.1.3.2.1	Repeated measures.....	203
4.3.2.1.3.3	Study-level predictors.....	205
4.4	Discussion	206
4.4.1	Key findings and interpretation.....	206

4.4.1.1	Sex differences in response to topical NSAIDs	206
4.4.1.2	Baseline pain.....	210
4.4.1.3	Study-level predictors	211
4.4.2	Caveats	211
4.4.3	Methodological comments	213
4.4.3.1	IPD retrieval rates	213
4.4.3.2	Statistical considerations for IPD MA	214
4.4.3.3	Presentation and appraisal of IPD MA	216
5	CHAPTER 5: Individualising care and identifying predictors of response to topical NSAIDs and capsaicin in OA	218
5.1	Overview	218
5.2	Methods	219
5.2.1	Research ethics, registration, and funding	219
5.2.2	Trial design.....	219
5.2.2.1	Randomisation and treatment allocation	221
5.2.2.2	Blinding	221
5.2.3	Study participants	222
5.2.4	Interventions	224
5.2.4.1	Concurrent therapies.....	225
5.2.5	Baseline and outcome assessments	225
5.2.5.1	Baseline measures	225
5.2.5.1.1	Questionnaire assessments.....	226
5.2.5.1.2	Clinical assessments.....	228
5.2.5.2	Outcome measures	231
5.2.5.3	Withdrawals and incomplete data.....	232
5.2.6	Sample size justification	232
5.2.7	Statistical analysis	233
5.2.7.1	Treatment responses at the individual level	234
5.2.7.2	Aggregated treatment responses	235
5.2.7.3	Examination of predictors of response.....	235
5.2.7.4	Sensitivity analyses	236
5.2.7.5	Carryover and period effects	237
5.2.7.6	Feasibility assessments	237

5.3 Results	239
5.3.1 Study population	239
5.3.1.1 Baseline characteristics	240
5.3.2 Preferential treatment responses	243
5.3.3 Aggregated treatment effect	245
5.3.4 Predictors of response	245
5.3.4.1 Preferential treatment response	245
5.3.4.2 Overall treatment effect (irrespective of treatment)	246
5.3.4.3 Treatment-by-covariate interactions	247
5.3.5 Feasibility assessments	248
5.3.5.1 Outcome measures	248
5.3.5.2 Weekly NRS pain scores	249
5.3.5.3 Number of cycles	250
5.4 Discussion	250
5.4.1 Key findings and interpretation	250
5.4.2 Recommendations for a definitive study	254
5.4.3 Caveats	257
5.4.4 Methodological comments	261
6 CHAPTER 6: Cross disease comparison – efficacy of topical NSAIDs and capsaicin in NP	263
6.1 Overview	263
6.2 Methods	263
6.2.1 Identification of eligible trials	263
6.2.1.1 Systematic literature search	263
6.2.1.1.1 Search terms and strategies	264
6.2.1.1.2 Databases and literature search output	264
6.2.1.2 Identification of other studies	264
6.2.1.3 Study eligibility criteria	264
6.2.2 Data collection	266
6.2.3 Data analysis	266
6.3 Results	267
6.3.1 Study-level MAs	267
6.3.1.1 Description of studies	267

6.3.1.1.1	Topical NSAIDs	267
6.3.1.1.1.1	Literature search	267
6.3.1.1.2	Capsaicin	269
6.3.1.1.2.1	Literature search	269
6.3.1.1.2.2	Description of the studies.....	271
6.3.1.1.2.3	Risk of bias assessment	273
6.3.1.2	Specific treatment effect	274
6.3.1.2.1	Topical NSAIDs	274
6.3.1.2.2	Capsaicin	274
6.3.1.2.2.1	At or nearest to four weeks.....	274
6.3.1.2.2.1.1	Cumulative meta-analysis	276
6.3.1.2.2.1.2	Publication bias and small study effects..	277
6.3.1.2.2.1.3	Sensitivity analyses	278
6.3.1.2.2.1.4	Subgroup analysis by duration of treatment	279
6.3.1.3	Overall treatment effect and PCE.....	280
6.3.1.3.1	Topical NSAIDs	280
6.3.1.3.2	Capsaicin	280
6.3.1.4	Summary of findings	281
6.3.2	IPD MA	282
6.3.2.1	Description of studies.....	282
6.3.2.1.1	Topical NSAIDs	282
6.3.2.1.2	Capsaicin	282
6.4	Discussion	285
6.4.1	Key findings and interpretation.....	285
6.4.2	Caveats	287
7	DISCUSSION	289
7.1	Key findings and interpretation	289
7.2	Future work	290
	REFERENCES	293
	APPENDICES	327
	Appendix 1 – Systematic literature search strategies	327
	Appendix 2 – Data extraction form and instructions.....	330

Appendix 3 – GRADE assessment: Topical NSAIDs in OA	336
Appendix 4 – GRADE assessment: Capsaicin in OA.....	337
Appendix 5 – GRADE assessment: NMA in OA	338
Appendix 6 – GRADE assessment: Overall effect of topical NSAIDs in OA	339
Appendix 7 – GRADE assessment: Overall effect of capsaicin in OA...	340
Appendix 8 – Contact letter to authors	341
Appendix 9 – Contact email to institutions	344
Appendix 10 – Contact letter to pharmaceutical companies	345
Appendix 11 – OA Trial Bank DTA.....	348
Appendix 12 – UoN drafted DTA.....	361
Appendix 13 – List of variables from IPD MA trials	364
Appendix 14 – DTA for collaboration with VSM	365
Appendix 15 – Pilot n-of-1 trial series UoN ethics approval	367
Appendix 16 – Association between baseline characteristics and overall treatment effect of topical therapy in pilot n- of-1 trial series	369
Appendix 17 – Treatment-by-baseline characteristic interaction effects in pilot n-of-1 trial series	370
17a. Primary analysis.....	370
17b. Sensitivity analysis – baseline pain 4-8.....	373
17c. Sensitivity analysis – MCID outcome	376
Appendix 18 – Weekly NRS pain scores in pilot n-of-1 trial series	379
Appendix 19 – GRADE assessment: Capsaicin in NP	380
Appendix 20 – GRADE assessment: Overall effect of capsaicin in NP	381

List of tables

Table 1-1 - ACR classification criteria and EULAR diagnostic criteria for OA.....	13
Table 1-2 - OARSI clinical trial recommendations for outcome measures in symptom-modifying OA trials	37
Table 1-3 - Putative mechanisms of action of NSAIDs	46
Table 1-4 - Details of topical NSAID recommendations in OA.....	48
Table 1-5 - MAs of topical NSAID efficacy in OA or chronic MSK pain .	50
Table 1-6 - Details of topical capsaicin recommendations in OA	60
Table 1-7 - MAs of capsaicin efficacy in OA or chronic MSK pain.....	62
Table 1-8 - Suggested main subgroups of OA in clinical research	69
Table 1-9 - Predictors of increased response to topical capsaicin in NP	74
Table 2-1 - List of databases used for systematic literature search.....	78
Table 2-2 - Inclusion and exclusion criteria used for MAs of topical NSAIDs and capsaicin in OA	80
Table 2-3 - Modified Risk of Bias assessment.....	84
Table 2-4 - Formulae used to prepare data for analysis.....	87
Table 2-5 - Pain outcome hierarchy list used for analysis	89
Table 3-1 - Formulae used to group multiple time points per study .	106
Table 3-2 - Steps for calculating the ES for dichotomous outcomes presented as number/percent with pain reduction or OR for pain reduction	107
Table 3-3 - Steps for calculating the overall ES and PCE	110

Table 3-4 - Topical NSAID and capsaicin recommendations for topical administration in OA.....	112
Table 3-5 - Description of the studies included in the analysis of topical NSAIDs in OA.....	119
Table 3-6 - Description of the interventions and comparators in the studies included in the analysis of topical NSAIDs in OA.....	122
Table 3-7 - Description of the studies included in the analysis of capsaicin in OA.....	128
Table 3-8 - Description of the interventions and comparators in the studies included in the analysis of capsaicin in OA.....	128
Table 3-9 - Sensitivity analyses by study quality for topical NSAIDs in OA. Outcome: Pain at or nearest to four weeks	134
Table 3-10 - Median number of participants analysed per study by duration of treatment	138
Table 3-11 - Meta-regression of sample size and duration of treatment versus specific treatment effect	138
Table 3-12 - Median number of participants analysed per study by duration of treatment for capsaicin in OA	145
Table 3-13 - Bayesian and frequentist NMAs of topical NSAIDs and capsaicin	147
Table 3-14 - Subgroup Bayesian and frequentist NMAs of topical NSAIDs and capsaicin used as recommended in the BNF.....	148
Table 4-1 - Potential predictors of the specific and overall treatment effects	178
Table 4-2 - Criteria used to assess the credibility of subgroup analyses	179
Table 4-3 - Description of the studies included in the IPD MA of topical NSAIDs in OA.....	186

Table 4-4 - Description of the interventions and comparators used in the IPD MA of topical NSAIDs in OA.....	187
Table 4-5 - Distribution of baseline characteristics in the placebo-controlled trials analysed in the IPD MA of topical NSAIDs in OA	188
Table 4-6 - Distribution of baseline characteristics in the placebo- and active-controlled trials analysed in the overall treatment effect IPD MA of topical NSAIDs in OA.....	190
Table 4-7 - Specific effect of topical NSAIDs in OA at or nearest to four weeks	193
Table 4-8 - Time points (weeks) of assessment and associated specific effect of topical NSAIDs in OA.....	193
Table 4-9 - Specific effect of topical NSAIDs in OA using all time-points of assessment (repeated measures)	194
Table 4-10 - Overall effect of topical NSAIDs in OA at or nearest to four weeks	196
Table 4-11 - Time points (weeks) of assessment in one-stage IPD MA trials of the overall effect of topical NSAIDs in OA	196
Table 4-12 - Interaction effect (β) and associated 95%CI for potential predictors of the specific effect of topical NSAIDs at or nearest to four weeks	198
Table 4-13 - Specific effect of topical NSAIDs at or nearest to four weeks by sex	199
Table 4-14 - Interaction effect (β) and associated 95% CI for potential predictors of the specific effect of topical NSAIDs using repeated measures data.....	200
Table 4-15 - β coefficient and associated 95% CI for potential predictors of the overall treatment effect of topical NSAIDs at or nearest to four weeks	202

Table 4-16 - Overall effect of topical NSAIDs at or nearest to four weeks by sex	202
Table 4-17 - β coefficient and associated 95% CI for potential predictors of the overall treatment effect of topical NSAIDs using repeated measures data	204
Table 4-18 - β coefficient for potential study-level predictors of the specific and overall treatment effect of topical NSAIDs at or nearest to four weeks	205
Table 5-1 - Possible treatment sequences for each n-of-1 trial	221
Table 5-2 - Pilot n-of-1 trial series participant eligibility criteria	223
Table 5-3 - Potential predictors of treatment response and their method of assessment for the pilot n-of-1 trial series ...	226
Table 5-4 - Inter-machine agreement for the presence of US features	229
Table 5-5 - Demographic and clinical baseline characteristics in the pilot n-of-1 trial series	241
Table 5-6 - Distribution of significant baseline characteristics per preferential treatment response group (determined using NRS criteria).....	246
Table 5-7 - The number (%) of participants per treatment preference category using NRS preferential treatment response, participant's end-of-cycle pain preference criteria, or participant's end-of-study overall preference	249
Table 5-8 - The number (%) of participants per treatment preference category using primary NRS criteria and using weekly participant diary data at each duration of follow-up	250
Table 6-1 - Inclusion and exclusion criteria used for MAs of topical NSAIDs and capsaicin in NP.....	265

Table 6-2 - Description of the studies included in the analysis of capsaicin in NP	272
Table 6-3 - Description of the interventions and comparators in the studies included in the analysis of capsaicin in NP	272
Table 6-4 - Sensitivity analyses by study quality for capsaicin in NP. Outcome: Pain at or nearest to four weeks	278

List of figures

Figure 1-1 - Taxonomy of OA	2
Figure 1-2 - Systemic and local biomechanical risk factors for OA.....	7
Figure 1-3 - Clinical features of hand OA	11
Figure 1-4 - Clinical features of knee OA	12
Figure 1-5 - Kellgren-Lawrence classification of radiographic knee OA	15
Figure 1-6 - Pathophysiology of osteoarthritic knee joint	17
Figure 1-7 - Subchondral BMLs in knee OA	18
Figure 1-8 - Macroscopic pathology of OA of the femoral head	19
Figure 1-9 - US features of inflammation in OA.....	20
Figure 1-10 - NICE recommendations for a holistic assessment of a person with OA.....	32
Figure 1-11 - BMLs and signs of inflammation seen in an MRI of a knee with OA	43
Figure 2-1 - PRISMA 2009 Flow Diagram	92
Figure 2-2 - Representation of a network meta-analysis	95
Figure 3-1 - Sub-effects of the contextual effect	108
Figure 3-2 - Illustration of the overall treatment effect and PCE	109
Figure 3-3 - Flow Diagram: Literature search results for topical NSAIDs in OA	117
Figure 3-4 - Risk of bias assessment for topical NSAID versus placebo trials in OA	124
Figure 3-5 - Flow Diagram: Literature search results for capsaicin in OA	126

Figure 3-6 - Risk of bias assessment for capsaicin versus placebo trials in OA	129
Figure 3-7 - Topical NSAIDs versus placebo in OA. Outcome: pain at or nearest to four weeks	130
Figure 3-8 - Cumulative MA of topical NSAIDs versus placebo in OA. Outcome: pain at or nearest to four weeks.....	131
Figure 3-9 - Funnel plot with pseudo 95% confidence limits for topical NSAIDs in OA.....	132
Figure 3-10 - Filled funnel plot of topical NSAIDs in OA	133
Figure 3-11 - Topical NSAIDs versus placebo in OA. Sensitivity analysis combining continuous and dichotomised pain outcomes at or nearest to four weeks	136
Figure 3-12 - Topical NSAIDs versus placebo in OA. Outcome: pain by weeks of treatment.....	137
Figure 3-13 - Capsaicin versus placebo in OA. Outcome: Pain at or nearest to four weeks	139
Figure 3-14 - Cumulative MA of capsaicin versus placebo in OA. Outcome: Pain at or nearest to four weeks.....	141
Figure 3-15 - Funnel plot with pseudo 95% confidence limits for capsaicin in OA.....	142
Figure 3-16 - Subgroup analysis: Capsaicin 0.025% four times per day versus placebo in OA. Outcome: Pain at or nearest to four weeks	143
Figure 3-17 - Capsaicin versus placebo in OA. Outcome: Pain by duration of treatment	144
Figure 3-18 - Network diagram for primary NMA in OA.....	146
Figure 3-19 - Overall treatment effect of topical NSAIDs in OA. Outcome: Pain at or nearest to four weeks.....	149

Figure 3-20 - Cumulative MA: Overall treatment effect of topical NSAIDs in OA. Outcome: Pain at or nearest to four weeks	150
Figure 3-21 - Overall treatment effect (ES) and PCE of topical NSAIDs at or nearest to four weeks in OA. Sensitivity analysis by trial quality	151
Figure 3-22 - Overall treatment effect (ES) and PCE of topical NSAIDs at or nearest to four weeks in OA. Subgroup analysis by drug class	152
Figure 3-23 - Overall treatment effect (ES) and PCE of topical NSAIDs in OA. Outcome: Pain by weeks of treatment	153
Figure 3-24 - Overall treatment effect of capsaicin in OA. Outcome: Pain at or nearest to four weeks.....	154
Figure 3-25 - Overall treatment effect (ES) and PCE of capsaicin in OA. Outcome: Pain by weeks of treatment	155
Figure 3-26 - Overall treatment effect (ES) and PCE of topical NSAIDs and capsaicin in OA. Outcome: Pain at or nearest to four weeks	156
Figure 4-1 - Process of preparing original study data for merging and analysis	171
Figure 4-2 - IPD MA models and time points used for the analysis of the specific and overall treatment effect	173
Figure 4-3 - Levels in multi-level IPD MA	174
Figure 4-4 - Flow diagram of trial identification, contact, and data acquisition for IPD MA of topical NSAIDs in OA	181
Figure 4-5 - Flow diagram of trial identification, contact, and data acquisition for IPD MA of topical capsaicin in OA	184
Figure 4-6 - Risk of bias assessment for topical NSAID trials analysed in the IPD MA in OA	191

Figure 4-7 - Specific effect of topical NSAIDs versus placebo at or nearest to four weeks in two-stage IPD MA.....	192
Figure 4-8 - Overall effect of topical NSAIDs at or nearest to four weeks in two-stage IPD MA.....	195
Figure 4-9 - Treatment-by-sex interaction term for topical NSAIDs at or nearest to four weeks in two-stage IPD MA.....	199
Figure 4-10 - Specific effect of topical NSAID, presented as MD (0-100mm VAS), for men and women at all durations of assessment.....	201
Figure 5-1 - N-of-1 trial design showing a hypothetical random treatment sequence for capsaicin cream (C) and ibuprofen gel (I).....	220
Figure 5-2 - CONSORT flow diagram of participant recruitment and retention in the pilot n-of-1 trial series.....	239
Figure 5-3 - Treatment responses for topical ibuprofen (blue) and capsaicin (red).	244
Figure 6-1 - Flow Diagram: Literature search results for topical NSAIDs in NP	268
Figure 6-2 - Flow Diagram: Literature search results for capsaicin in NP	270
Figure 6-3 - Risk of bias assessment of the trials included in the analysis of capsaicin in NP	273
Figure 6-4 - Capsaicin versus placebo in NP. Outcome: Pain at or nearest to 4 weeks.....	274
Figure 6-5 - Cumulative MA of capsaicin versus placebo in NP. Outcome: Pain at or nearest to four weeks	276
Figure 6-6 - Funnel plot with pseudo 95% confidence limits for capsaicin in NP	277

Figure 6-7 - Capsaicin versus placebo in NP. Outcome: Pain by weeks of follow-up	279
Figure 6-8- Flow diagram of trial identification, contact, and data acquisition for IPD MA of topical NSAIDs in NP	282
Figure 6-9 - Flow diagram of trial identification, contact, and data acquisition for IPD MA of capsaicin in NP	284

List of abbreviations

ACR	American College of Rheumatology
ADL	Activities of daily living
AE	Adverse event
AMED	Allied and Complementary Medicine Database
ANCOVA	Analysis of covariance
AUSCAN	Australian/Canadian OA Hand Index
BMD	Bone mineral density
BMI	Body mass index
BML	Bone marrow lesion
BNF	British National Formulary
CI	Confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CMA	Conventional meta-analysis
CMC	Carpometacarpal
CNS	Central nervous system
COX	Cyclo-oxygenase
CPM	Conditioned pain modulation
CrI	Credible interval
CRP	C-reactive protein
CV	Cardiovascular
DIP	Distal interphalangeal
DLA	Disability Living Allowance
DMARD	Disease-modifying anti-rheumatic drug
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTA	Data transfer agreement
EFNS	European Federation of Neuropathic Societies
EMA	European Medicines Agency
eMC	Electronic Medicines Compendium
ES	Effect size
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
FIHOA	Functional Index for Hand OA
GI	Gastrointestinal

GSK	GlaxoSmithKline
GTN	Glyceryl trinitrate
HADS	Hospital Anxiety and Depression Scale
HIV	Human immunodeficiency virus
HR-QoL	Health-related quality of life
hsCRP	High sensitivity C-reactive protein
IASP	International Association for the Study of Pain
IPD	Individual patient data
IPD MA	Individual patient data meta-analysis
IQR	Interquartile range
ITT	Intention-to-treat
JSN	Joint space narrowing
KL	Kellgren-Lawrence
KOOS	Knee Injury and Osteoarthritis Outcome Score
KPIC	Knee pain and related health in the community study
MA	Meta-analysis
MCP	Metacarpophalangeal
MD	Mean difference
MHRA	Medicines and Healthcare Products Regulatory Agency
MJL	Medial joint line
MRI	Magnetic resonance imaging
MSK	Musculoskeletal
n	Number of participants
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NLDA	Nottingham Line Drawing Atlas
NMA	Network meta-analysis
NNT	Number needed to treat
NP	Neuropathic pain
NRS	Numeric rating scale
NSAID	Non-steroidal anti-inflammatory drug
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
OR	Odds ratio
PCE	Proportion of contextual effects
PDN	Painful diabetic neuropathy

PDQ	painDETECT questionnaire
PGA	Patient's global assessment
PHN	Post-herpetic neuralgia
PIP	Proximal interphalangeal
PIS	Patient information sheet
PP	Per protocol
PPT	Pressure pain threshold
QoL	Quality of life
QST	Quantitative sensory testing
RA	Rheumatoid arthritis
REML	Restricted maximum likelihood
RCT	Randomised controlled trial
RF	Rheumatoid factor
RR	Relative risk
SD	Standard deviation
SE	Standard error
SF-36	Short form-36
SH	Synovial hypertrophy
SMD	Standardised mean difference
SR	Systematic review
SS	Symptom severity
SYSADOA	Symptomatic slow acting drugs for osteoarthritis
TJR	Total joint replacement
TRPV1	Transient receptor potential vanilloid 1
TS	Temporal summation
UoN	University of Nottingham
US	Ultrasound
VAS	Visual analogue scale
WMD	Weighted mean difference
WOMAC	Western Ontario and McMaster Universities OA Index
WPI	Widespread pain index

1 CHAPTER 1: Introduction

1.1 Introduction to Osteoarthritis (OA)

1.1.1 History, definition, and nature

The term “osteoarthritis” was coined by A E Garrod in 1890 (Dequeker and Luyten, 2008). Previously known by a variety of other names, including osteoarthrosis, degenerative joint disease, and arthrosis deformans, these terms all describe a complex and heterogeneous disorder that has affected the synovial joints of mankind throughout our evolution (Dequeker and Luyten, 2008).

Several definitions for OA exist, and these definitions evolve as our knowledge of the condition increases. Most recently, the Osteoarthritis Research Society International (OARSI) provided a definition that encompasses the anatomic, molecular, and physiologic elements of OA (Kraus et al., 2015). Using the lay definition provided by OARSI, OA can be defined as:

“A disorder that can affect any moveable joint of the body, for example knees, hips and shoulders. It can show itself as a breakdown of tissues and abnormal changes to cell structures of joints which can be initiated by injury.” AND ‘Osteoarthritis first shows itself as a change to the biological processes within a joint, followed by abnormal changes to the joint itself (such as the breakdown of cartilage, bone reshaping, bony lumps, joint inflammation, loss of joint function). This can result in pain, stiffness and loss of movement.” (Blackburn et al., 2016).

An interesting aspect of the full definition that OARSI has developed is the differentiation between OA as a *disease* and as an *illness*. As a disease, OA encompasses the anatomic, molecular, and physiological changes that occur. However, these only truly become an illness when they manifest in an individual, causing the pain, disability, stiffness, and other symptoms that are associated with OA.

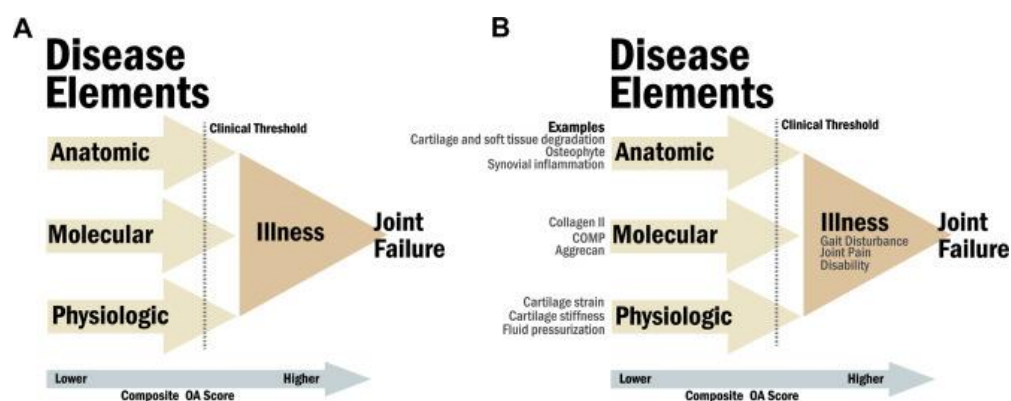


Figure 1-1 - Taxonomy of OA

Panel A: OARSI proposed taxonomy for OA that describes the molecular, anatomic, and physiologic components of disease and its progression to illness. Panel B: OA specific examples. Image reproduced with permission. Reference: (Kraus et al., 2015)

The OARSI definition describes the nature of OA (Figure 1-1). It commences with a multitude of changes within a joint that, to the individual involved, are often unnoticed. With time, these changes may reach a threshold where they manifest clinically with symptoms such as pain and functional impairments. At this point, OA transitions from a disease to an illness. However, it must be noted that a person's experience of OA does not necessarily follow the linear path illustrated. For many, OA does not cause a problem and never progresses to an illness. For others, progression of OA may halt and resolution of symptoms may be experienced, as often witnessed in hand OA.

The essence of OA as an illness was captured in 1805 by John Haygarth who described how OA may "sadly embitter the comforts of life; but I know of no instance in which they seemed to shorten its duration" (Dequeker and Luyten, 2008). Although the association between OA and increased mortality continues to be debated (Xing et al., 2016, Veronese et al., 2016, Leyland et al., 2017), current evidence indicates that OA does negatively impact the lives of those affected (Maly and Krupa, 2007). The course of the illness at this point varies greatly within individuals and for some people it culminates in total joint replacement, regarded as a marker for joint failure.

The following introduction will focus primarily on OA as an illness (symptomatic OA), as it is the illness that is most burdensome for the individual and society. In addition, pain is central to the experience of OA and this will be where much of the emphasis will be placed.

1.1.2 Epidemiology

1.1.2.1 Incidence and prevalence

It is difficult to assess the incidence rate of OA, due to the challenges of defining OA and determining its point of onset (Pereira et al., 2011). The onset of OA is insidious and to set a threshold within a continuum of structural and symptomatic changes that determines the presence or absence of OA is very difficult (Felson and Zhang, 2015). It is therefore much easier to assess the prevalence than incidence of OA.

The prevalence of OA varies greatly according to the joint site, the definition used, and the population studied. The hand, knee, and hip joints are most commonly examined separately. In addition, observational studies will often determine the prevalence of OA using three main definitions: radiographic OA, symptomatic OA (radiographic OA plus symptoms), or self-reported OA (Pereira et al., 2011). These interconnecting factors make it difficult to summarise the prevalence of OA in a simple sentence.

In a systematic review (SR) and meta-analysis (MA) of 72 cross-sectional and longitudinal studies worldwide, Pereira et al. (2011) found that the overall prevalence of hand OA (43.3%, 95% confidence interval (CI) 42.7 to 43.9) was greater than for knee (23.9%, 95%CI 23.6 to 24.2) or hip OA (10.9%, 95%CI 10.6 to 11.2). Across each joint, the prevalence of radiographic OA was higher than self-reported or symptomatic OA, which were generally similar (Pereira et al., 2011). For example, although the radiographic prevalence of hand OA was 48.7% (95%CI 48.0 to 49.3), the prevalence of symptomatic hand OA was 15.0% (95%CI 14.1 to 15.9). The prevalence of OA varied also by age, sex, and population (Pereira et al., 2011). The considerable and significant heterogeneity associated with the prevalence estimates in this MA highlights not only the difficulty in establishing the prevalence of OA, but also the difficulty in summarising it. What one can conclude is that OA is a common condition that affects a substantial proportion of the older population. Its impact can be further considered not only by the number of people affected, but also by its burden.

1.1.2.2 Disease burden

The disease burden of OA is considerable and is thought to increase dramatically as populations grow increasingly older and more overweight (Hunter et al., 2014).

1.1.2.2.1 Individual costs

On an individual level, the main burden of OA is predominantly from pain, activity limitations, and reductions in quality of life (QoL) (Hunter et al., 2014). Pain can often be distressing to people with OA, particularly that which is aching, sharp, or triggered by activity (Hawker et al., 2008a). The impact of pain is not only limited to its presence, but also its effect on mood, sleep, and limitation of activities (Hawker et al., 2008a). In fact, the risk of mobility disability (needing help walking or climbing stairs) for those over 65 years is greater in knee OA than in any other medical condition (Hunter et al., 2014). Mobility limitations may result in social isolation and reductions in self-worth (Maly and Krupa, 2007). These factors all combine to reduce the QoL of individuals living with OA. In the UK, OA is the 11th leading cause of years lived with disability, higher than conditions such as diabetes, ischaemic heart disease, and stroke (Murray et al., 2013).

1.1.2.2.2 Socioeconomic costs

The socioeconomic costs of OA are classified as direct costs, indirect costs, and intangible costs. Direct costs include the cost of treatments (non-pharmacological, pharmacological, and surgical), healthcare resources, and management of complications that arise secondary to OA (Hunter et al., 2014). Indirect costs include work-time lost, reduced employment, reduced productivity, carer costs, and disability benefits (Hunter et al., 2014). The third category, intangible costs, includes the reduction in QoL, pain, and suffering that results from OA (Hunter et al., 2014). These equate to the individual costs previously discussed.

Of the direct costs of OA, only an estimated 10% is believed to be attributable to pharmaceutical costs (Hunter et al., 2014). The majority of direct costs are due to hospital stays, particularly for total joint replacements (TJR) (Hunter et al., 2014). The use of effective and safe

pharmaceutical treatments that are able to postpone or reduce the need for TJRs therefore have the potential to significantly reduce the burden of OA on the health system. The costs of healthcare for OA vary by country, but are believed to be approximately 1% of the total healthcare expenditure (Puig-Junoy and Zamora, 2015). One review has gone as far as to estimate that the direct costs of OA amounted to over £1 billion in 2010 in the UK (Chen et al., 2012). With the National Health Service (NHS) budget set at £98.7 billion in the same year (Treasury, 2013), the direct costs of OA appear to account for approximately 1% of the health expenditure in the UK as well.

Of the indirect costs of OA, almost two thirds are attributed to productivity losses, mainly reduction in employment rates (Hunter et al., 2014). Arthritis is the commonest reason for receiving the Disability Living Allowance (DLA) in the UK, with more OA sufferers claiming DLA than the total of heart disease, stroke, respiratory disease, and cancer combined (Chen et al., 2012). The overall cost of OA, when collating healthcare and non-healthcare-related costs, is thought to be between 0.25-0.5% of a country's gross domestic product (Puig-Junoy and Zamora, 2015).

1.1.2.3 Risk factors

OA is a complex condition for which many risk factors have been identified. These risk factors likely interact to increase the susceptibility of an individual to OA (Felson et al., 2000). The risk factors for OA can be described according to their nature: local versus systemic, modifiable versus non-modifiable, or joint-specific versus other factors. They can be further sub grouped by their associated endpoint: development versus progression of OA or radiographic versus symptomatic OA. For simplicity, the following section will describe the risk factors as systemic factors that increase the risk of OA at any site and local factors that predispose certain joints to OA. The predominant risk factors are described in Figure 1-2.

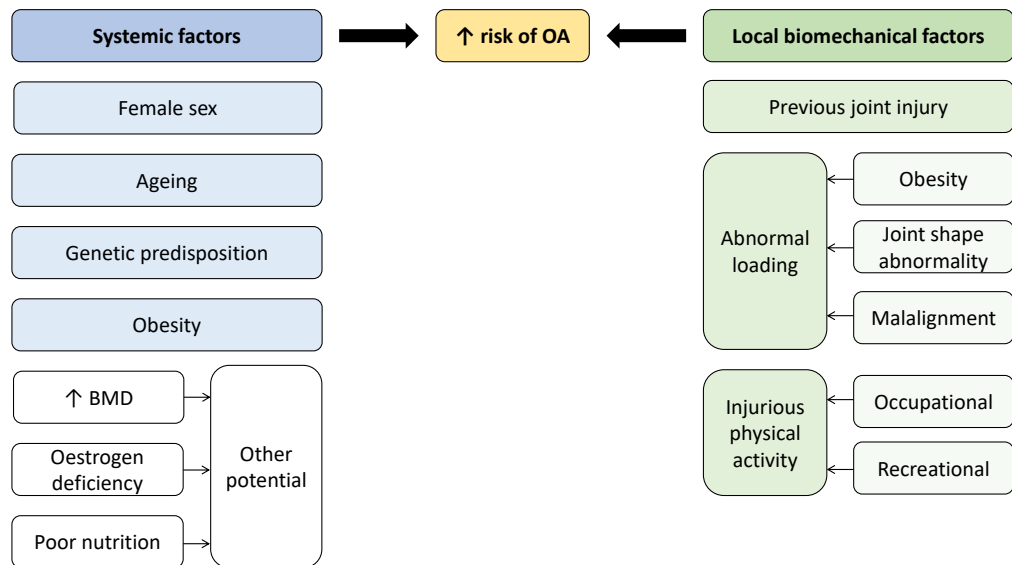


Figure 1-2 - Systemic and local biomechanical risk factors for OA
BMD; bone mineral density; OA, osteoarthritis

1.1.2.3.1 Systemic risk factors

Ageing and female sex are commonly recognised systemic risk factors for OA (Felson et al., 2000). Across all joints, OA occurs more commonly with increasing age (Felson et al., 2000). Sex-specific differences in the prevalence of OA have been identified, and these vary by age group and joint affected. For example, a MA of 34 studies found that, compared to women, men had 37% lower risk of developing knee OA (relative risk (RR) 0.63, 95%CI 0.53 to 0.75) and 19% lower risk of developing hand OA (RR 0.81, 95%CI 0.73 to 0.90) (Srikanth et al., 2005). In contrast, there was no significant difference in hip OA between men and women (RR 1.18, 95%CI 0.91 to 1.52). The sex difference was greater in older populations and when OA was defined by non-radiographic methods (Srikanth et al., 2005).

The development of OA appears to be largely genetically determined. The heritability of OA is estimated to lie between 30% and 65%, and it is recognised that having a relative affected by OA increases the risk of developing OA yourself (Warner and Valdes, 2016). In addition, several genes have been found to associate with large joint OA and pain sensitivity (Warner and Valdes, 2016). However, the gap between these findings and their clinical relevance remains broad.

Other potential systemic risk factors, for which the evidence base is mixed, include oestrogen deficiency, high bone mineral density (BMD), and poor nutrition (Felson and Zhang, 2015, Felson et al., 2000, Silverwood et al., 2015).

Finally, although often regarded as an extrinsic biomechanical factor, obesity may also be a systemic risk factor for OA. Increasing body mass index (BMI) is significantly associated with a higher risk of knee OA, and this risk is higher for those that are obese relative to those that are overweight (Silverwood et al., 2015). Similarly, high BMI is associated with an increased risk of hip OA (Jiang et al., 2011). Its association with knee and hip OA may be attributed to abnormal loading on the joint. There is, however, some evidence that obesity may also increase the risk of hand OA (Jiang et al., 2016). This is likely not explained by excessive loading of the joints, but by central obesity-related metabolic syndrome and deleterious effects of higher circulating levels of adipokines (Robinson et al., 2016).

1.1.2.3.2 Local risk factors

Local factors that increase the risk of OA are predominately joint injury, abnormal loading, and injurious physical activities (recreational or occupational) (Felson and Zhang, 2015). For example, a MA of 27,326 participants found that the pooled odds ratio (OR) for developing knee OA was significantly higher for those with a previous knee injury, compared to those without (pooled OR 2.83, 95%CI 1.91 to 4.19) (Silverwood et al., 2015). In addition, major joint injury is believed to be the primary cause of OA at sites not commonly affected, such as the ankle and wrist (Felson and Zhang, 2015).

Abnormal loading may come in the form of obesity (excessive loading), joint shape abnormalities, or malalignment. Acquired or developmental abnormalities of the hip joint, such as non-spherical femoral head, are recognised risk factors for hip OA as they cause increased focal stress on the joint (Felson and Zhang, 2015). Similarly, knee malalignment is a risk factor for knee OA (Tanamas et al., 2009).

Lastly, excessive or repetitive movements may contribute to the development of OA. Occupational factors, such as heavy/manual work, lifting/carrying, excessive kneeling, and squatting are associated with an increased risk of knee OA (McWilliams et al., 2011, Blagojevic et al., 2010, Silverwood et al., 2015). Similarly, participation in sports may also increase the risk of developing knee OA (Tran et al., 2016).

1.1.3 Diagnosis

1.1.3.1 Clinical presentation

The presentation of OA is varied, but is characterised by a collection of signs and symptoms in the affected joint. On presentation, clinical symptoms are usually limited to one joint, but these often evolve slowly to become bilateral as the disease progresses (Altman, 2015). In a subset of people, finger interphalangeal joint involvement is later accompanied by OA in various other joints such as the knee, hip, and first metatarsophalangeal joints (O'Reilly and Doherty, 2003). This is termed nodal generalised OA (O'Reilly and Doherty, 2003).

Pain is often the most troubling issue for individuals with OA and is the primary reason for visits to primary care, non-steroidal anti-inflammatory drug (NSAID) use, and TJR (Hawker et al., 2008a). The pain in OA fluctuates significantly during the day and from day to day (Hutchings et al., 2007, Allen et al., 2009). In addition, the nature of the pain may change if the condition slowly progresses.

Pain in the early stages of hip and knee OA is usage-related and often described as a sharp, predictable pain that is brought on by certain activities (Hawker et al., 2008a). As the condition progresses, the pain becomes more constant and may associate with unpredictable symptoms, such as locking or giving way (Hawker et al., 2008a). In later stages, the pain is described as dull or aching pain that is punctuated by unpredictable, sharp pain that is emotionally draining (Hawker et al., 2008a). Many OA sufferers also complain of neuropathic-like pain, often described as burning, shooting, or pins and needles (Hawker et al., 2008a). A MA of 3,032 participants across eight observational studies found that the overall prevalence of neuropathic-like pain in knee or hip OA was 0.23 (95%CI 0.10 to 0.39) (French et al., 2017). High levels of heterogeneity were associated with the estimate, which varied from 0.06 to 0.52 in the individual studies (French et al., 2017). Within the Nottingham community, the prevalence of neuropathic-like pain in those with chronic knee pain was estimated at 14% (Fernandes et al., 2018). The prevalence was greater in women (16.6%) than men (11.6%) (Fernandes et al., 2018).

In contrast, the pain experience in hand OA has not been fully characterised. For many people, hand OA is associated with pain in the initial years, after which it resolves. The nature of the pain itself and how it changes through the years has not been described. However, terms such as “pain like a knife cutting”, “pain like fever”, and “ache” have been used by participants to characterise their hand OA pain (Stamm et al., 2009). Most likely, the pain experience and presentation of pain in hand OA is as varied as that for knee and hip OA.

Pain is often accompanied by various other signs and symptoms which vary by joint. In general, these are functional impairment, stiffness, tenderness, crepitus, restriction of movement, deformity, and instability (O'Reilly and Doherty, 2003, Altman, 2015). Features indicative of hand OA are illustrated in Figure 1-3. These include bony enlargements (Heberden's and Bouchard's nodes) and deformities of the interphalangeal joints or the base of the thumb (Altman, 2015). These may be associated with pain, loss of range of motion, stiffness, and subluxation (Altman, 2015, O'Reilly and Doherty, 2003).



Figure 1-3 - Clinical features of hand OA

Bony enlargements (Heberden's and Bouchard's nodes) and deformities of the interphalangeal joints seen in hand OA. Image provided courtesy of Professor Michael Doherty

Patients with knee OA will often complain of pain, stiffness, gelling, and “giving way” (O'Reilly and Doherty, 2003). On examination, coarse crepitus, tenderness, varus/valgus deformities, an antalgic gait, and quadriceps weakness may be evident (O'Reilly and Doherty, 2003, Altman, 2015). The clinical presentation of knee OA is illustrated in Figure 1-4



Figure 1-4 - Clinical features of knee OA

Varus deformity and bony swelling seen in OA of the right knee. Image provided courtesy of Professor Michael Doherty

Hip OA presents with pain, which is usually most intense deep in the anterior groin and may radiate to the buttock, thigh, or knee (O'Reilly and Doherty, 2003, Altman, 2015). In addition, hip OA may be associated with an antalgic gait, stiffness, loss of range of motion, and muscle wasting (O'Reilly and Doherty, 2003, Altman, 2015).

1.1.3.2 Criteria for diagnosis

The diagnosis of OA is often based on the history and clinical examination findings, although radiographic criteria may also be used. Joint-specific criteria for the diagnosis or classification of OA have been established by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR). These are described in Table 1-1. In general, a diagnosis or classification of OA can be made when usage-related chronic joint pain and joint-specific signs and symptoms are identified in adults over 40 years. Further criteria excluding other causes of joint pain, such as inflammatory arthritides, are often included in the diagnostic or classification criteria (Zhang et al., 2010a, Zhang et al., 2009, Altman et al., 1990, Altman et al., 1986, Altman et al., 1991).

Table 1-1 - ACR classification criteria and EULAR diagnostic criteria for OA

Joint	ACR criteria	EULAR criteria
Hand	Hand pain, aching, or stiffness and 3 or 4 of the following features: • Hard tissue enlargement of 2 or more of 10 selected joints* • Hard tissue enlargement of 2 or more DIP joints • Fewer than 3 swollen MCP joints • Deformity of at least 1 of 10 selected joints*	>40 years of age plus typical symptoms: • Pain on usage • Mild morning or inactivity stiffness affecting just one or a few joints at any one time • Intermittent symptoms • Symptoms target characteristic sites (DIP, PIP, thumb base, index, and middle MCP joints)
Hip	Hip pain and at least two of the following features: • ESR < 20 mm/hour • Radiographic femoral or acetabular osteophytes • Radiographic JSN (superior, axial, and/or medial)	No diagnostic criteria

* Selected joints: second and third DIP, second and third PIP, or the first CMC joints of both hands

Continued overleaf

Joint	ACR criteria	EULAR criteria
Knee	1. Clinical and laboratory Knee pain, plus at least five of: • Age > 50 years • Stiffness < 30 minutes • Crepitus • Bony tenderness • Bony enlargement • No palpable warmth • ESR < 40 mm/hour • RF < 1:40 • Synovial fluid signs of OA 2. Clinical and radiographic Knee pain, plus osteophytes and at least one of: • Age > 50 years • Stiffness < 30 minutes • Crepitus 3. Clinical Knee pain, plus at least three of: • Age > 50 years • Stiffness <30 minutes • Crepitus • Bony tenderness • Bony enlargement • No palpable warmth	Typical features and >40 years of age Typical features: • Usage-related knee pain • Short-lived morning stiffness • Functional limitation • Typical examination findings (crepitus, restricted movement, bony enlargement) Clinical diagnosis can be confidently made according to the above features without radiographic examination

Criteria summarised as per ACR and EULAR guidelines (Zhang et al., 2010a, Zhang et al., 2009, Altman et al., 1990, Altman et al., 1986, Altman et al., 1991)
 ACR, American College of Rheumatology, CMC, carpometacarpal; DIP, distal interphalangeal; ESR, erythrocyte sedimentation rate; EULAR, European League Against Rheumatology; JSN, joint space narrowing; MCP, metacarpophalangeal; PIP, proximal interphalangeal; OA, osteoarthritis; RF, rheumatoid factor

EULAR diagnostic criteria were developed for use in clinical practice, whilst most clinical trials use the ACR classification criteria. In trials, the ACR criteria are generally combined with assessment of plain film radiographs, using the Kellgren-Lawrence (KL) grading scale (Kellgren and Lawrence, 1957) (Figure 1-5), to confirm the diagnosis of OA and determine eligibility. This is discordant with clinical practice where radiographic assessment is only required for a diagnosis of hip OA, not for hand or knee OA.

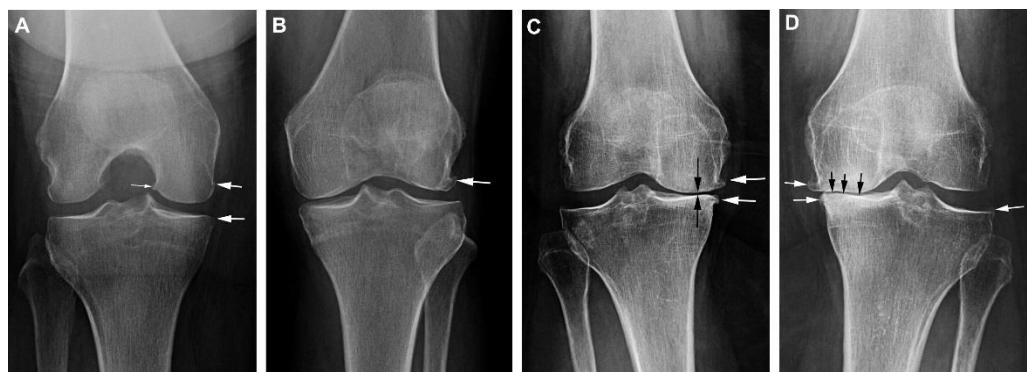


Figure 1-5 - Kellgren-Lawrence classification of radiographic knee OA

Severity of radiographic OA determined using KL grading of knee OA using anterior-posterior plain film radiographs. A. KL grade 1 showing minimal, equivocal osteophytes at the medial joint margins (large arrows). Notch osteophytes (small arrow) are not considered in the KL grading. B. KL grade 2 showing a definite marginal osteophyte (arrow) without evidence of JSN. C. KL grade 3 showing definite JSN (black arrows) and marginal osteophytes (white arrows). D. KL grade 4 showing bone-to-bone contact, complete obliteration of the joint space (black arrow), and definite marginal osteophytes (white arrows). Image reproduced with permission. Reference: (Hayashi et al., 2016)

JSN, joint space narrowing; KL, Kellgren-Lawrence; OA, osteoarthritis

1.1.4 Pathology

OA predates the evolution of man. As evidenced by osteoarthritic changes seen in a Comanchean dinosaur fossil, it appears to have remained pathologically unchanged for 100 million years (Dequeker and Luyten, 2008). The skeletal changes of OA can be observed in human skeletons dating back to the Ancient Egyptian and Neolithic times (Dequeker and Luyten, 2008). The changes in osteoarthritic joints are relatively well recognised, although the causes for these changes are not. A variety of hypotheses aimed at explaining the changes have been voiced, including chronic mechanical overloading, matrix proteolysis, pro-inflammatory cytokine production, activation of cellular inflammatory signalling pathways, premature ageing of chondrocytes and cartilage matrix, and damage to the chondrocyte's deoxyribonucleic acid (DNA) (Aigner et al., 2015). The most robust hypothesis is the role of adverse biomechanical factors in the pathogenesis of OA (Aigner et al., 2015), however, it is by no means likely to be the only driving factor for OA. Most likely, a large variety of factors contribute to the pathogenesis of OA and are responsible for the pathological changes witnessed.

1.1.4.1 Pathophysiology

Although previously thought to be a consequence of normal ageing, described as a disease of “wear and tear”, this theory is now thought to be inaccurate. Although much remains to be known regarding the causes of OA, it is recognised as being separate from the normal ageing process (Wyatt and Doherty, 2016). It encompasses changes in all components of the joint, and it is the biochemical changes of the cartilage, bone, and synovium that we now recognise as OA (Aigner et al., 2015). OA is a metabolically dynamic process and both attrition and synthesis of tissues are involved in its pathogenesis (Wyatt and Doherty, 2016). Following injury to the synovial joint from diverse insults, increases in the metabolic activity of joint tissues and new tissue production are evident. These are thought to reflect a regenerative process (i.e., a condition of “wear, tear, and repair”). In many cases this slow repair process may be successful, thus leaving a

structurally abnormal joint without pain. This may explain the evolutionary preservation of OA in species with synovial joints. However, with overwhelming insult and/or a poor repair process, the joint may continue remodelling and progress towards joint failure.

A schematic summary of the pathological changes evident in an osteoarthritic knee joint are illustrated in Figure 1-6.

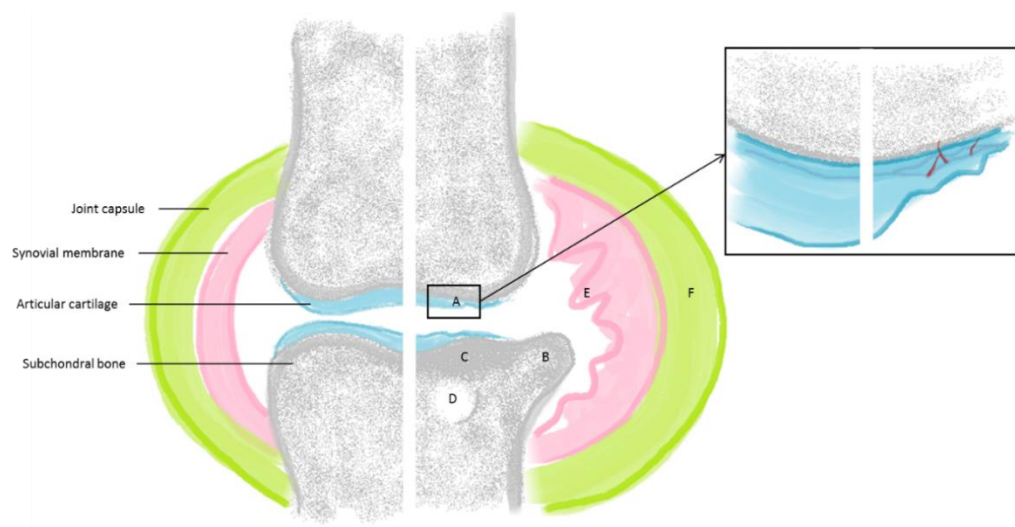


Figure 1-6 - Pathophysiology of osteoarthritic knee joint

Normal knee seen on left and OA changes seen on right. OA changes in the articular cartilage involve cartilage loss, duplication of the tidemark, and the growth of blood vessels and nerves through the tidemark (A). Within the bone, osteophyte formation (B), sclerosis (C), and cysts (D) are apparent. Changes to surrounding structures also occur, such as synovitis (E) and capsular fibrosis (F). Adapted from (Aigner et al., 2015)

Normal articular cartilage forms a smooth covering for the joint surfaces and is responsible for the biomechanical properties of joints (Aigner et al., 2015). Articular cartilage undergoes a variety of changes in OA. Initially, disruption to the type II collagen scaffolding causes the cartilage volume to increase as the water content increases and proteoglycan swelling occurs (Aigner et al., 2015). This is later followed by cartilage loss secondary to the formation of superficial cracks (fibrillation) and continued protein-degrading activity (Aigner et al., 2015). Deep to these changes in the non-calcified cartilage, the calcified cartilage zone becomes thicker. In addition, blood vessels and

nerves break through the tidemark separating the calcified and non-calcified cartilage, leading to neovascularisation and neo-innervation of the cartilage (Aigner et al., 2015).

Deeper within the joint, at the junction between cartilage and bone, periosteal and synovial mesenchymal stem cells are induced to proliferate and form osteophytes (Wyatt and Doherty, 2016).

Osteophytes are bony spurs that occur on the peripheral aspects of joints. Changes also occur directly within the subchondral bone in OA, although it is unclear whether these cause or result from the articular changes (Aigner et al., 2015). Morphologically, the subchondral bone of osteoarthritic joints may display subchondral sclerosis, cysts, and bone marrow lesions (BMLs) (Aigner et al., 2015). The subchondral bone plate becomes thickened and architectural changes occur to the bone trabeculae causing subchondral sclerosis to be visible radiographically (Aigner et al., 2015). As well as the radio-opaque areas of sclerosis, radiolucent areas of subchondral cysts are also evident deep to where articular cartilage is lost (Aigner et al., 2015). Furthermore, localised areas of bone marrow necrosis, fibrosis, and reparative changes occur, termed BMLs (Aigner et al., 2015). These are seen as hyper-intensities on magnetic resonance imaging (MRI) (Figure 1-7) and are associated with pain (Aigner et al., 2015).

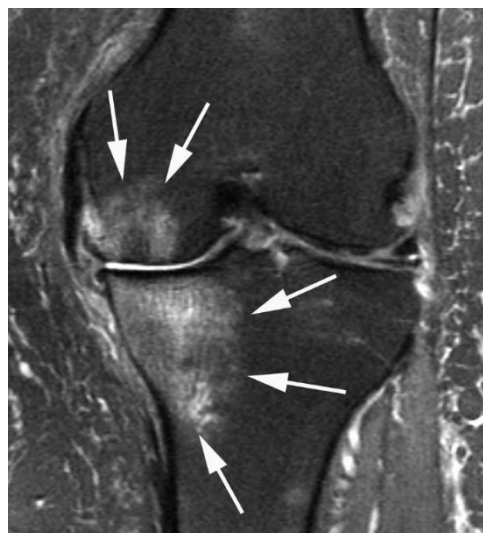


Figure 1-7 - Subchondral BMLs in knee OA

BMLs (white arrows) visualised as areas of ill-defined hyperintensity in an MRI of a knee with OA. Image reproduced with permission. Reference: (Hayashi et al., 2016)

BML, bone marrow lesion; MRI, magnetic resonance imaging; OA, osteoarthritis

Evidence of low-grade inflammation of the synovial membrane (synovitis) may occur in OA (Wyatt and Doherty, 2016). Cartilage debris is released from the damaged cartilage, causing pathological synovial changes and the release of cytokines, growth factors, and enzymes that can further disrupt articular cartilage homeostasis (Aigner et al., 2015). Finally, the pathological changes in synovium, cartilage, and bone are often accompanied by changes in the menisci, joint capsule, intra-articular ligaments, extra-articular connective tissues, and peri-articular muscle (Wyatt and Doherty, 2016).

Most research of OA pathology is centred on large-joint OA, such as that of the hip and knee. The morphological changes described are common across joint sites. They are most easily seen macroscopically in the femoral head (Figure 1-8).

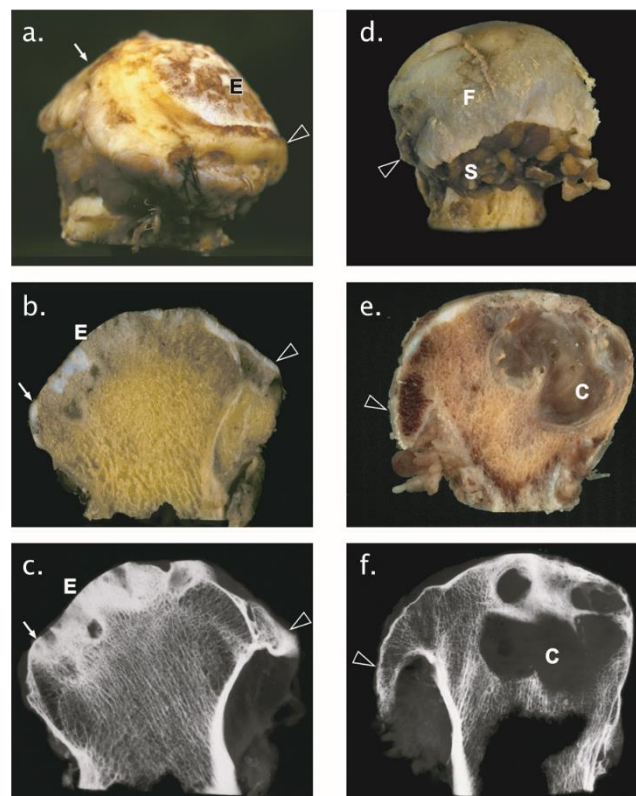


Figure 1-8 - Macroscopic pathology of OA of the femoral head

The surface (a and d), cut surface (b and e), and specimen x-ray (c and f) of two femoral heads with OA that have been removed surgically are shown. Prominent morphological features of OA shown are cartilage erosion (arrow), bone eburnation/sclerosis (E), osteophyte (arrowhead), cartilage fibrillation (F), cyst formation (C), and synovial hypertrophy (S). Image reproduced with permission. Reference: (Wyatt and Doherty, 2016). OA, osteoarthritis

1.1.4.1.1 The role of inflammation

The role of inflammation in the pathogenesis of OA is widely debated. The presence of inflammation in OA is generally accepted, as evidenced by the presence of synovitis, effusion, and stiffness (Robinson et al., 2016). Changes indicative of inflammation, such as effusion, synovial hypertrophy (SH), and Power Doppler signals, can be visualised using ultrasound (US) examination (Figure 1-9).

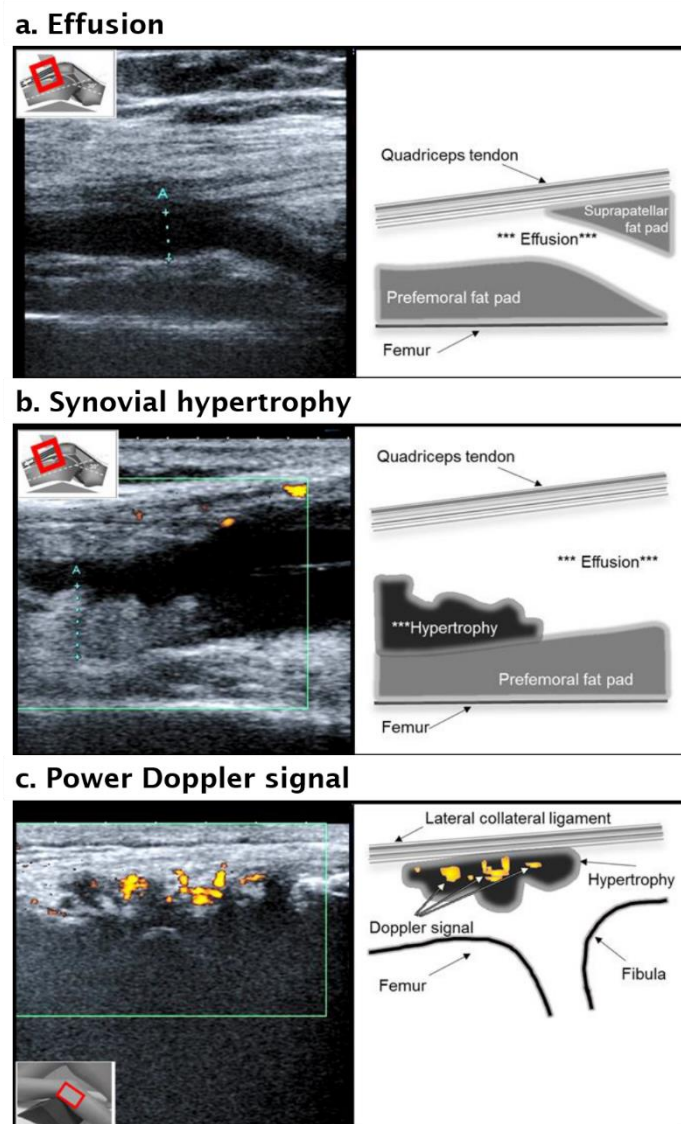


Figure 1-9 - US features of inflammation in OA

Grey-scale US image of an effusion (a), SH (b), and Power Doppler signal (c). US images of pathology seen on the left were taken of the knee area and position depicted in the inset small boxes. Schematic drawings on the right illustrate the synovial pathology in relation to other joint structures. Image reproduced with permission. Reference: (Sarmanova, 2017), contains adapted image from (Martinoli and Bianchi, 2007) OA, osteoarthritis; SH, synovial hypertrophy; US, ultrasound

US markers of inflammation are raised in those with knee pain compared to those without knee pain (Sarmanova et al., 2017). In addition, the intensity of US inflammation increases with worsening structural changes (Sarmanova et al., 2017). Furthermore, a MA found that high sensitivity C reactive protein (hsCRP), a marker of systemic inflammation, was marginally raised in those with OA compared to controls (Jin et al., 2015). However, whether inflammation drives OA or whether it is secondary to a largely mechanical disease process is unclear.

On one side, it is argued that OA occurs secondary to adverse mechanics within the joint (Felson, 2013). Abnormal mechanical loading may result from various insults such as malalignment, congenital dysplasia, meniscal tears, or chronic excessive loading (Felson, 2013). The affected cartilage is damaged and the underlying bone may also undergo remodelling, further driving the abnormal loading (Felson, 2013). Consequently, cartilage debris may cause a secondary inflammatory response in the synovium, evidenced by synovitis and excess fluid secretion (Felson, 2013). The basis for the hypothesis is grounded primarily in studies that have found associations between risk factors that cause abnormal mechanical forces in the joint, such as obesity, injury, and occupational overuse, and the development of OA (Felson, 2013). Additionally, correcting abnormal mechanics may alleviate the problem (Felson, 2013). However, this hypothesis does not wholly explain all cases of OA. For instance, the increased risk of hand OA in obese people may not be explained by excessive loading as the increased weight is unlikely to affect their non-weight bearing joints (Berenbaum, 2013).

Others believe inflammation to be a primary driver of OA (Berenbaum, 2013, Robinson et al., 2016). Low-grade local and systemic inflammation are thought to cause OA through a multitude of mechanisms, including innate and adaptive immune mechanisms and inflammatory mediators (Robinson et al., 2016). Post-traumatic OA is thought to cause OA through local inflammation which causes synovitis and activates mechanoreceptors (Berenbaum, 2013). Similarly, OA is thought to be driven by low-grade systemic

inflammation in metabolic syndrome, by a secretory inflammatory phenotype in increased age, and by innate immunity in crystal OA (Berenbaum, 2013). Much of the evidence for this hypothesis stems from animal model and pre-clinical work that has shown an increase in cytokines, receptors, and other inflammatory mediators locally and systemically in OA (Berenbaum, 2013). However, this has remained largely unsubstantiated in human studies. At their most effective, disease modifying anti-rheumatic drugs (DMARDs) licensed for rheumatoid arthritis (RA) failed to provide a clinically significant improvement in pain in OA (Persson et al., 2018). Furthermore, although hsCRP levels were raised in OA, they did not significantly associate with radiographic OA or disease progression (Jin et al., 2015). In addition, the association between US-detected effusion and knee pain progression was confounded by radiographic severity (Sarmanova et al., 2017).

Indeed, it appears that neither hypothesis is able to aptly describe all cases of OA. Instead, the development and progression of OA is likely multifactorial and results from a complex interaction between adverse mechanics and inflammation. Adverse mechanical loading of the joint may be the preliminary and primary driver of OA, resulting in secondary inflammation which contributes to a lesser extent to the pathogenesis. Some individuals will certainly exhibit an OA phenotype with greater inflammatory features, such as in erosive hand OA, and these individuals may represent the extreme ends of a spectrum in terms of the role of inflammation. OA pathogenesis is unlikely to be driven exclusively by inflammation.

1.1.4.2 Pain mechanisms

Pain can be defined as an unpleasant sensory and emotional experience that results from actual or potential noxious stimuli or tissue injury (IASP, 2017, Schaible, 2015). Importantly, pain is more than just a nociceptive process, it is a subjective experience that influences and is influenced by psychological and social factors.

To better understand pain, the relevant processes can be broadly classed into peripheral (nociceptive or neuropathic) and central

mechanisms. Nociceptive pain arises when noxious stimuli activate the sensory endings of pain-sensing nerves (nociceptors) (Schaible, 2015). These stimuli may be mechanical, thermal, or chemical, and nociceptors will often respond to more than one type of stimulus (Schaible, 2015). Neuropathic pain (NP) arises from injury or damage to the nociceptive neurons themselves, which then generate pathologic ectopic discharges (Schaible, 2015). When faced with persistent or severe stimulation, nociceptors become sensitised to the stimuli (Schaible, 2015). The excitation threshold is lowered and increased input is conveyed to the spinal cord (Schaible, 2015). This, in turn, produces spinal cord hyperexcitability, termed central sensitisation (Schaible, 2015).

Peripheral sensitisation may cause primary hyperalgesia (increased pain from a stimulus that normally provokes pain (IASP, 2017)) and allodynia (pain due to a stimulus that normally does not cause pain (IASP, 2017)) at the site of injury (Schaible, 2015). In addition, central sensitisation may cause secondary hyperalgesia in neighbouring and more remote sites (Schaible, 2015).

1.1.4.2.1 Peripheral

There is no single source for the pain in OA, nor is there a single mechanism through which it is generated. The pain results from a change in the biochemical and mechanical milieu within a joint, caused by changes to the organ as a whole. In the periphery, pain is likely generated through two broad mechanisms: the direct activation of nociceptors in joint tissues and the indirect activation or sensitisation of nociceptors by mediators. These likely work in parallel, and sometimes in unison, to contribute to the pain sensation. For simplicity, the source of pain in OA will be described using the knee joint, although other joints are likely to share similar mechanisms.

Firstly, the relationship between structural changes and pain in OA must be understood. Traditionally, the association between plain film radiographic findings and pain has been regarded as poor, and this is reflected in much of the literature around pain mechanisms in OA (Hunter et al., 2013, Malfait A, 2013, Neogi, 2017). A discordance

between structure and symptoms is often described as an individual with radiographically normal knees may experience pain, whilst another may have severe radiographic changes without any pain. However, there is evidence to indicate that there are correlates between radiographic changes and OA pain when adequate adjustment for psychosocial determinants of pain are taken into account and when radiographs of the whole joint are taken and scored optimally (Neogi, 2017, Neogi et al., 2009, Szebenyi et al., 2006). In addition, it is now recognised that some MRI-detectable structural features, such as BMLs, subarticular bone attrition, and synovitis, associate with pain (Malfait A, 2013, Hunter et al., 2013, Barr et al., 2015, Neogi, 2017). BMLs and synovitis, in particular, associate not only with pain presence and severity, but also with pain fluctuations (Neogi, 2017). A link between structural changes and pain is acknowledged, however, not all structural changes necessarily cause pain and those that do are likely influenced by psychosocial factors.

Having established that there is some link between structure and pain, one can then try to understand *how* the changes cause pain. Knee joints are densely innervated, with nociceptors concentrated in the joint capsule, ligaments, periosteum, menisci, subchondral bone, and synovium (Malfait A, 2013). In addition, changes in the mechanical and biochemical milieu may cause the new growth of blood vessels and nerves in previously aneural and vulnerable tissues. The previously mentioned growth of nerves and vessels through the tidemark and into cartilage is one such example (Mapp PI, 2012). The presence of nerves where there previously were none could lead to aberrant pain signalling as they become exposed to chemical or mechanical stimulation (Mapp PI, 2012). Similarly, pre-existing nociceptors may become exposed to adverse chemical and mechanical stimuli as changes occur within the joint. For example, cartilage loss may cause exposure of subchondral nociceptors (Hunter et al., 2013). Furthermore, nociceptor activation within the bone may be a source of pain. Subchondral microfractures, bone angina secondary to reduced blood flow, and raised intraosseous pressure may explain the association between BMLs and pain (Hunter et al., 2013).

Finally, tissue injury or remodelling stimulates the release of mediators into the joint which activate and sensitise peripheral nociceptors (Miller RE, 2015). The mediators involved include cytokines, chemokines, neuropeptides, prostaglandins, damage-associated molecular products, nerve growth factor, and nitric oxide (Malfait A, 2013, Miller RE, 2015). Many of these molecules have been shown to directly excite nociceptors, so may directly contribute to pain in this manner (Miller RE, 2015). In addition, these mediators may alter the milieu within the joint such that nociceptor activation is facilitated. The mediators may reduce the activation threshold of nociceptors or allow spontaneous nociceptive firing, thereby driving peripheral sensitisation (Malfait A, 2013). Normal joint movements may therefore become painful (mechanical allodynia) as the normally innocuous movements activate the abnormally sensitive nociceptors (Malfait A, 2013).

Most likely, the pain experienced in OA arises from various sources and through various mechanisms. However, it is important to remember that the pain in OA is more complex than can be attributed to peripheral nociceptive mechanisms alone. There is a relationship between structural changes and knee pain, however, the genesis of pain in OA is ultimately not limited to the joint. Compared with people with isolated knee OA pain, a weaker association between structural changes and pain is evident in individuals that also have pain elsewhere (Pan et al., 2017). These findings are largely attributable to central mechanisms.

1.1.4.2.2 Central

Pain in OA is not generated purely through peripheral processes, but is modified and enhanced or suppressed to varying extents by the central nervous system (CNS). The pain may be altered or amplified to the extent that it no longer reflects the presence of a noxious stimulus in the periphery (Woolf, 2011). Pain hypersensitivity secondary to the amplification of neuronal signals within the CNS is termed central sensitisation (Woolf, 2011).

Intense or prolonged input from peripheral nociceptors causes spinal cord neurons to become more excitable and renders them more

sensitive to further nociceptive signalling (Schaible, 2015). Activation of spinal cord neurons by primary afferents that would not normally reach suprathreshold levels becomes possible (Schaible, 2015). In addition, synaptic processing may be further modified by the release of mediators (Schaible, 2015). Finally, higher centres within the CNS can either dampen (descending inhibitory pathways) or enhance (descending facilitatory pathways) the pain experience. In chronic pain states, like OA, the balance between these pathways may become weighted towards facilitation, thereby reducing nociceptor activation thresholds and increasing postsynaptic transmissions (Hassan H, 2014).

In practice, sensitisation may be assessed using the point at which increasing levels of pressure transition into a painful sensation, named pressure pain thresholds (PPT) (Arendt-Nielsen, 2017). In OA, PPTs are lowered at the knee and at remote sites, such as the contralateral forearm (Fingleton et al., 2015). Other changes indicative of sensitisation have also been reported in OA, including increased temporal summation (TS) at remote sites and reduced flexor withdrawal responses (Fingleton et al., 2015). The presence of sensitisation suggests that the pain experience in OA may have an important central component which is able to moderate and maintain chronic pain (Woolf, 2011).

Central sensitisation may be reversible and measures of sensitisation have been shown to normalise following removal of peripheral drivers of pain, such as through TJR in OA (Fingleton et al., 2015). This justifies the belief that central sensitisation is maintained by aberrant peripheral signalling from the joint (Malfait A, 2013). However, approximately one-in-five people continue to report knee pain after TJR and measures of central sensitisation are linked to the presence of post-operative pain (Arendt-Nielsen, 2017, Beswick et al., 2012). As such, peripheral drivers of OA pain may trigger central sensitisation, which modulates and enhances pain. In the majority of individuals, central hypersensitivity normalises following the removal of peripheral signals. However, in a subset of individuals, nociceptive input may cause long-term potentiation of central pain pathways, resulting in

central sensitisation which persists long after the peripheral inputs are removed (Schaible, 2015).

1.1.4.2.3 Neuropathic-like pain

The pain experience in OA varies between individuals and some report the presence of neuropathic-like pain, which is described below.

The current view is that neuropathic-like pain results from the neuronal plasticity that occurs secondary to OA. It is unclear whether the growth of nerves into previously aneural tissues, the effect of changes to the biochemical and mechanical milieu on pre-existing nerves, and abnormalities of the central processing of pain can constitute lesions to the nervous system (Thakur et al., 2014). The pain is therefore most often referred to as *neuropathic-like* pain rather than NP, as the changes in OA may not meet the criteria for NP nor result primarily from nerve damage. True NP is described in section 1.1.4.2.3.1.

Regardless of the nomenclature of the pain experienced, it is hypothesised that the spontaneous firing of damaged nociceptors from affected joints drives neuropathic-like pain (Thakur et al., 2014). As previously discussed, such increases in nociceptor firing may also contribute to central sensitisation. This may explain the association between neuropathic-like pain and quantitative sensory testing (QST) measures indicative of central sensitisation and central markers related to abnormal pain processing (e.g., multiple regional pain, pain catastrophising, anxiety, depression) (Hochman et al., 2013, Kurien et al., 2018, Fernandes et al., 2018).

1.1.4.2.3.1 Neuropathic pain (NP)

NP is a clinical descriptor, rather than a diagnosis, which must be associated with a lesion or disease that meets diagnostic criteria for a neurological condition (IASP, 2017). Examples of lesions and diseases associated with NP include post-herpetic neuralgia (PHN), diabetes mellitus, human immunodeficiency virus (HIV), nerve entrapment syndromes, and post-traumatic neuralgia (Freynhagen and Bennett, 2009). CNS lesions, such as spinal cord injuries and infarctions, may also cause NP (Freynhagen and Bennett, 2009). Although pathologically

distinct from OA, examination of NP may serve to better our understanding of neuropathic-like pain in OA and will be summarised below.

1.1.4.2.3.1.1 Clinical presentation of NP

NP can be elicited by a multitude of nociceptor-damaging stimuli, however, the clinical presentation of the condition is similar regardless of the differences in aetiology (Baron et al., 2010). Irrespective of its cause, the nerve lesion results in changes to the nervous system which causes the signs and symptoms associated with NP (Baron et al., 2010).

NP is characterised by spontaneous pain, coupled with sensory deficits in the territory innervated by the damaged nerve. These deficits include loss of mechanical, vibratory, noxious, or thermal sensation (Baron et al., 2010, Freynhagen and Bennett, 2009, Cruccu et al., 2004). Other sensory abnormalities in the affected territory are also commonly reported, such as paraesthesia, dysaesthesias, abnormal thermal sensations, and paroxysmal pain (such as electric shock-like sensations) (Baron et al., 2010, Freynhagen and Bennett, 2009). Furthermore, evidence of allodynia, hyperalgesia, and summation (progressive worsening pain felt in response to repeated mildly noxious stimuli) is often present (Baron et al., 2010).

1.1.4.2.3.1.2 Criteria for diagnosis of NP

NP is best diagnosed clinically through careful history-taking and examination (Freynhagen and Bennett, 2009, Cruccu et al., 2004). Screening tools, such as the neuropathic pain questionnaire, ID-Pain, DN4, painDETECT questionnaire (PDQ), and LANSS can be used, but may miss 10-20% of cases of NP when compared to clinical examination by an experienced clinician (Freynhagen and Bennett, 2009, Baron et al., 2010). However, these tools are particularly useful for identifying participants with likely NP in research settings as they are quick and easy to implement (Cruccu et al., 2004). Investigations such as QST and nerve conduction studies may provide further insight into NP, but as these studies may depend on expensive equipment and

are often time consuming they are better suited for research purposes rather than clinical practice (Cruccu et al., 2004).

1.1.4.2.3.1.3 Pain mechanisms in NP

The mechanisms of pain production in neuropathic conditions can be broadly divided into peripheral and central mechanisms.

Peripherally, various changes occur which are responsible for ectopic discharges, primary allodynia, and primary hyperalgesia (Baron et al., 2010). Increased expression of voltage-gated sodium channels and receptor proteins, such as transient receptor potential vanilloid 1 (TRPV1), in injured and non-injured neighbouring nociceptors may cause spontaneous nociceptive activity (Baron et al., 2010). These changes, coupled with inflammation and other biochemical changes in the environment of nerve endings, may also associate with primary allodynia and hyperalgesia (Baron et al., 2010).

Centrally, ongoing discharges from peripheral nociceptors cause a release of excitatory signals in the dorsal horn of the spinal cord, thereby contributing to neuronal hyper-excitability (Baron et al., 2010). These changes may even allow non-nociceptive afferent nerves to activate second order nociceptive neurons, which manifests as allodynia (Baron et al., 2010). Finally, impaired central modulation of pain signals, in the form of reduced inhibition, may contribute to central sensitisation (Baron et al., 2010).

1.1.4.2.3.1.4 Management of NP

Management of NP follows a stepped approach that takes into account the patient as a whole. An important element of management involves understanding the underlying cause of the pain and managing this condition as well as the pain (Dworkin et al., 2010). For example, in painful diabetic neuropathy (PDN), this involves improving glycaemic control.

Pain management is undertaken using various approaches, and it is often most efficacious to use a combination of approaches (Dworkin et al., 2010). The National Institute for Health and Care Excellence (NICE) guidelines recommend amitriptyline, duloxetine, gabapentin, or

pregabalin for the initial pharmacological management (NICE, 2018). This is mirrored in the recommendations by the International Association for the Study of Pain (IASP) and European Federation of Neuropathic Societies (EFNS) (Finnerup et al., 2015, Attal et al., 2010). If adequate pain relief cannot be achieved, switching to another first-line drug or using combinations of first-line drugs is recommended (Dworkin et al., 2010, NICE, 2018). IASP classed topical treatments, such as the 8% capsaicin patch and lidocaine patches, as second-line treatments that may be appropriate in peripheral NP with presumed local generation of pain (Finnerup et al., 2015). The EFNS recommendations also advise using topical lidocaine patches in peripheral NP, particularly in the elderly, but could not provide recommendations for the high concentration capsaicin patch as not enough evidence was available (Attal et al., 2010). Other medications that can be trialled, only in certain situations and generally only after first-line treatments have failed, are tramadol, strong opioids, topical capsaicin cream, botulinum toxin A, and other antidepressant and antiepileptic medications (Dworkin et al., 2010, NICE, 2018, Finnerup et al., 2015, Attal et al., 2010).

1.1.4.2.3.1.5 Topical NSAIDs and capsaicin in NP

Topical NSAIDs are generally not recommended for NP and literature on their use and efficacy is scarce.

Topical capsaicin is widely recommended for NP. It is believed to act through defunctionalisation of cutaneous nociceptors, particularly those that are spontaneously active, hypersensitive, or maintain chronic pain syndromes by driving neuronal hyperexcitability (Anand P, 2011).

1.1.5 Management of OA

OA care should adopt a holistic approach that targets the many ways that OA can affect a person. Management should be patient-centred and pain should be regarded through a biopsychosocial framework (Geenen et al., 2018). Management decisions should be made through shared decision making, whereby current best available evidence is considered along with the patient's wishes and expectations (Kloppenburg et al., 2018). Individual management plans will vary according to the site and number of joints affected, but the general management plan is similar for all people affected by OA.

As illustrated in NICE guidelines (Figure 1-10), pain assessment is only a portion of the management plan in OA (NICE, 2014). Although pain is often the most troubling symptom for individuals with OA (Hawker et al., 2008a), it is by no means the only aspect of OA that management should address. Focus should also be on the effect of OA on other areas of life, such as activities of daily living (ADLs), occupation, mood, and quality of sleep (NICE, 2014, Fernandes et al., 2013). Management should aim to achieve the best possible activity participation, activity performance, and QoL for the patient (Kloppenburg et al., 2018).

Several bodies provide recommendations for the treatment of OA and the recommendations described below are based on those from NICE, EULAR, OARSI, and ACR. In the UK, clinical practice is most likely to be guided by NICE recommendations, however, understanding other recommendations provides a better insight into the management of OA.

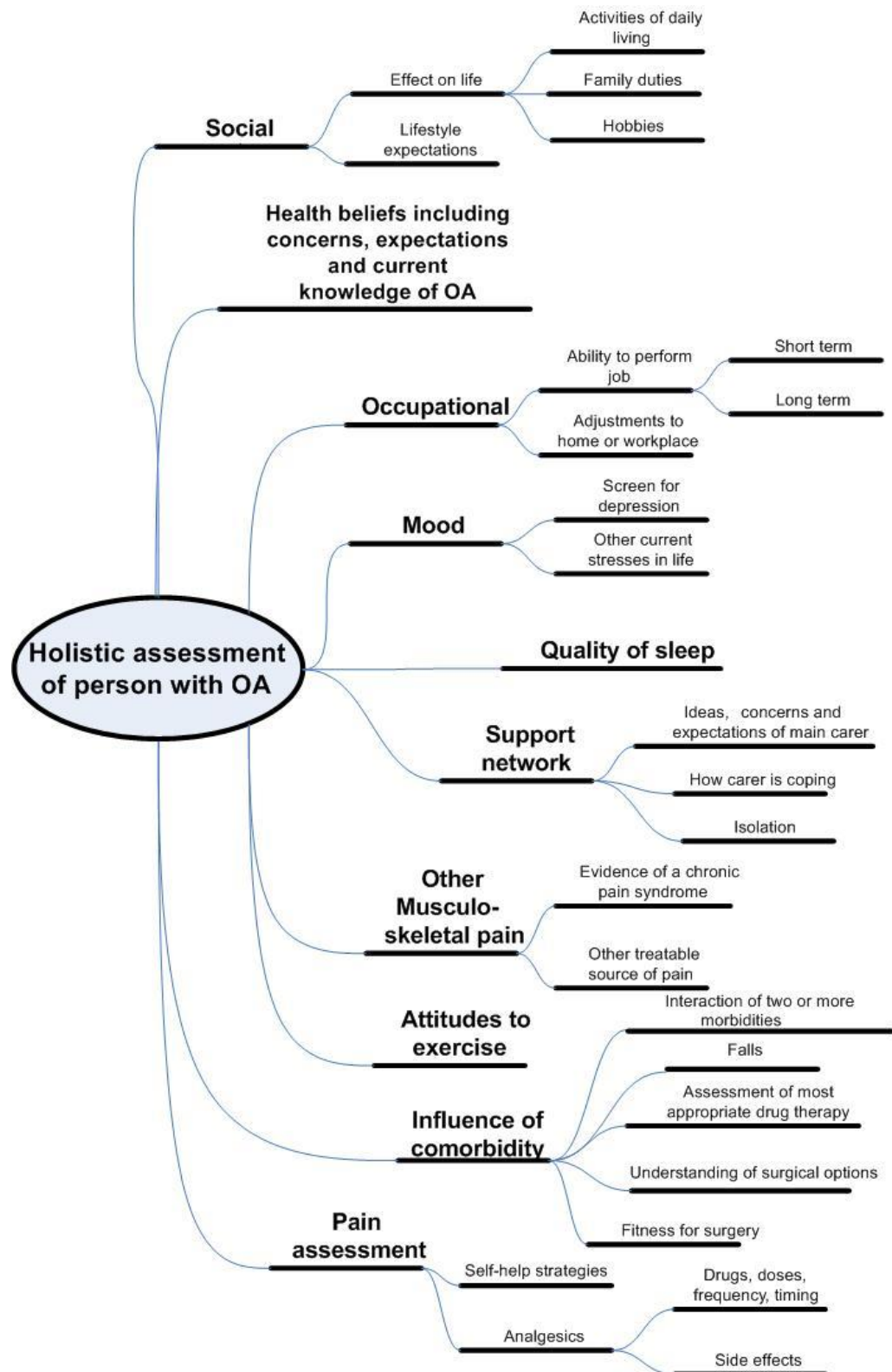


Figure 1-10 - NICE recommendations for a holistic assessment of a person with OA
 Recommendations as per NICE guidelines for OA. Image reproduced with permission.
 Reference: (NICE, 2014)

1.1.5.1 Non-pharmacological

All patients with OA should be offered a variety of non-pharmacological treatment options regardless of the joint affected. The core non-pharmacological treatment options are patient education, exercise (muscle strengthening, general aerobic, aquatic, range of motion exercises), and weight loss (if overweight/obese) (NICE, 2014, Hochberg et al., 2012, Fernandes et al., 2013, McAlindon et al., 2014, Zhang et al., 2008a, Zhang et al., 2005, Geenen et al., 2018, Kloppenburg et al., 2018). Other commonly recommended non-pharmacological OA management options include pacing of activities, thermotherapy, biomechanical aids such as insoles for lower limb OA or thumb base orthoses, and assistive devices for ADLs (Fernandes et al., 2013, McAlindon et al., 2014, Hochberg et al., 2012, NICE, 2014, Zhang et al., 2008a, Zhang et al., 2005, Geenen et al., 2018, Kloppenburg et al., 2018). More debated non-pharmacological options, which are recommended in some guidelines but not in others, are Tai Chi, acupuncture, balneotherapy, and transcutaneous electrical nerve stimulation (TENS) (McAlindon et al., 2014, Hochberg et al., 2012, NICE, 2014, Zhang et al., 2008a). Finally, EULAR recommendations for the health professional's approach to pain management in OA also include psychological, social, and sleep interventions (Geenen et al., 2018).

1.1.5.2 Pharmacological

Several pharmacological treatments are recommended for OA, and individuals will often rely on these analgesics for years. The general principles for the pharmacological management of OA are to use local treatments where possible, prioritise analgesics with favourable long-term safety profiles, and to use a step-wise approach to move to more potent analgesics only when required. Attempts to use medications with favourable safety profiles as the mainstay of treatment are reflected in the recommendations.

The most commonly recommended local treatment are topical NSAIDs. These are widely recommended for hand and/or knee OA (NICE, 2014, Hochberg et al., 2012, McAlindon et al., 2014, Zhang et al., 2008a,

Jordan et al., 2003, Kloppenburg et al., 2018). Topical capsaicin is commonly recommended alongside topical NSAIDs for knee and/or hand OA (NICE, 2014, McAlindon et al., 2014, Zhang et al., 2008a, Jordan et al., 2003), although it is not consistent across all guidelines. For example, ACR recommend topical capsaicin only for hand, and not knee, OA (Hochberg et al., 2012). In contrast, EULAR mention capsaicin as a topical treatment available for hand OA, but do not provide recommendations for or against its use in hand OA (Kloppenburg et al., 2018). The final commonly recommended local treatment is intra-articular corticosteroid injections. These are recommended for individuals with moderate to severe pain, particularly flare pain, in the knee or hip (NICE, 2014, Hochberg et al., 2012, McAlindon et al., 2014, Jordan et al., 2003, Zhang et al., 2008a, Zhang et al., 2005). IA corticosteroids are not generally recommended for hand OA (Kloppenburg et al., 2018, Hochberg et al., 2012).

Of the oral systemic treatments, paracetamol has traditionally been the first to be recommended (NICE, 2014, Hochberg et al., 2012, Jordan et al., 2003, Zhang et al., 2008a, Zhang et al., 2005). Its recommendation is commonly justified by its perceived safety. However, this belief is being questioned as a recent SR of observational studies raised concerns that paracetamol may not be as safe as previously thought (Roberts et al., 2016). Although the safety of paracetamol is still debated (Battaggia et al., 2016), it may be that OA guidelines are changed in the future with regards to paracetamol. The most recently developed recommendations for hand OA acknowledge the current safety concerns for paracetamol, but conclude that “for now there is no reason to refrain from prescribing paracetamol, preferably for a limited duration, in selected patients...” (Kloppenburg et al., 2018). Other widely recommended systemic pharmacological treatments are oral NSAIDs and cyclo-oxygenase (COX)-2 inhibitors, although these are generally reserved for patients that have not responded to local treatments or paracetamol (NICE, 2014, Hochberg et al., 2012, McAlindon et al., 2014, Jordan et al., 2003, Zhang et al., 2008a, Zhang et al., 2005, Kloppenburg et al., 2018). The use of oral NSAIDs should

ideally be for a limited duration and be accompanied by co-prescription of a proton pump inhibitor (NICE, 2014, Kloppenburg et al., 2018).

Finally, opioids, duloxetine, and symptomatic slow acting drugs for osteoarthritis (SYSADOA) are inconsistently recommended for OA. Opioids are generally recommended for severe knee and/or hip OA, except in the most recent OARSI update for knee OA where their appropriateness was deemed “uncertain” (NICE, 2014, Hochberg et al., 2012, McAlindon et al., 2014, Zhang et al., 2008a, Zhang et al., 2005, Jordan et al., 2003). Similarly, this most recent set of recommendations from OARSI is the only to recommend duloxetine, a serotonin norepinephrine reuptake inhibitor traditionally used for depression and anxiety, for knee-only OA without co-morbidities and multi-joint OA (McAlindon et al., 2014). Recommendations advocating the use of SYSADOA, such as glucosamine sulfate, chondroitin sulfate, and diacerein, are generally limited to older publications: the EULAR recommendations for knee and hip OA and the 2008 OARSI recommendations for knee and hip OA (Zhang et al., 2008a, Jordan et al., 2003, Zhang et al., 2005). In these, the use of SYSADOA is recommended as they *may* provide some relief without being associated with serious adverse events (AEs). In contrast, the use of these treatments is discouraged in newer recommendations (NICE, 2014, Hochberg et al., 2012, McAlindon et al., 2014). The exception is the most recent recommendation for hand OA which *suggests*, rather than recommends, the use of chondroitin sulfate (Kloppenburg et al., 2018).

1.1.5.3 Surgical

Surgical management of OA is generally reserved for patients with marked symptoms, radiographic evidence of OA, and who are refractory to conservative therapy. Surgical management for OA varies according to the joint involved, but can be categorised into three main groups. Firstly, there are surgical options that are aimed at improving the current symptoms, such as lavage and joint debridement (Altman and Hochberg, 2015). Secondly, there are surgical options that aim to prevent the progression of OA, such as osteotomy (Altman and

Hochberg, 2015). Finally, the last group of surgical procedures include joint arthroplasty or replacement, and aim to improve the symptoms related to advanced disease (Altman and Hochberg, 2015). Joint lavage and debridement does not appear to be effective for pain relief (Zhang et al., 2010b) and is not recommended by NICE (NICE, 2014). Instead, guidelines recommend trapeziectomy, arthroplasty, or arthrodesis for hand OA; osteotomy, unicompartmental or total knee replacement for knee OA; and osteotomy, joint preserving surgical procedures, or total hip replacement for hip OA (Zhang et al., 2005, Jordan et al., 2003, Zhang et al., 2008a, Kloppenburg et al., 2018).

1.1.6 Outcome measures for OA

The efficacy of a treatment in a clinical trial of OA can be assessed using several outcome measures. In 1996, a task force of the Osteoarthritis Research Society declared that pain, physical function, and patient's global assessment (PGA) should be the core outcome measures for OA clinical trials (Altman et al., 1996). In addition, they recommended that trials of >1 year duration also use imaging outcome measures (Altman et al., 1996). Other recommended outcomes were health-related quality of life (HR-QoL) and physician's global assessment (Altman et al., 1996).

These recommendations were recently updated by OARSI, who described joint-specific outcome measures for use in symptom-modifying clinical trials (McAlindon et al., 2015, Lane et al., 2015, Kloppenburg et al., 2015). These are listed in Table 1-2.

Table 1-2 - OARSI clinical trial recommendations for outcome measures in symptom-modifying OA trials

Hand OA	Knee OA	Hip OA
Pain	Pain	Pain
Physical function	Physical function	Physical function
PGA	PGA	PGA
HR-QoL	HR-QoL	Physician's assessment
Joint activity	Physical performance	of global status
Hand strength	measures	HR-QoL
Physician's assessment		Physical performance
of global status		measures
Stiffness		
Rescue medication		

Outcome measures in bold indicate core outcome measures. Hip OA recommendations did not specify which were core outcomes
Summarised from clinical trial recommendations (Kloppenburg et al., 2015, McAlindon et al., 2015, Lane et al., 2015)
HR-QoL, health-related quality of life; OA, osteoarthritis; PGA, patient global assessment

Pain, physical function, and PGA remained foremost on the list of recommended outcomes. Additional outcome measures were recommended for structure modifying drug trials, including hand mobility and assessment of structural damage through imaging (McAlindon et al., 2015, Lane et al., 2015, Kloppenburg et al., 2015). Finally, although no biochemical biomarker is yet available for OA, OARSI recommended that biological fluids be collected to assess the metabolic effect of a treatment on joint tissues or for further investigation of soluble biomarkers (McAlindon et al., 2015, Kloppenburg et al., 2015).

The instruments used to record these outcome measures should be valid, reliable, and responsive to change (Altman et al., 1996). In addition, Kloppenburg et al. (2015) recommended that the instruments be feasible and publicly available without associated mandatory costs for their use. Numerous instruments are available for each outcome and the choice of instrument should be made according to the joint investigated, the aim of the study, the responsiveness of the instrument, and the practical aspects of applying the instrument (Bischoff et al., 2003). OA as an illness is a subjective experience and is therefore most appropriately assessed using patient-reported outcome measures (Bischoff et al., 2003).

1.1.6.1 Pain

The European Medicines Agency (EMA) recommend that pain be the primary endpoint for symptom-modifying clinical trials in OA (EMA, 2010). In general, pain may either be assessed using a generic unidimensional pain question, a generic multidimensional pain questionnaire, or a disease- and joint- specific pain questionnaire (Hawker et al., 2011).

Generic unidimensional pain instruments, such as the visual analogue scale (VAS) and the numeric rating scale (NRS), generate either continuous (VAS) or integer (NRS) values that describe the intensity of pain during a specified activity (Hawker et al., 2011). Although praised for their simplicity, adaptability, and ease of use, they are only able to

capture one component of the pain experience: pain intensity (Hawker et al., 2011).

To better capture the complexity of pain, generic multidimensional instruments, such as the McGill pain questionnaire or the chronic pain grade scale, may be used for OA (Hawker et al., 2011). The McGill pain questionnaire is designed to capture the sensory, affective, and evaluative nature of pain, as well as its intensity (Hawker et al., 2011, Melzack, 1975).

OA- and joint-specific instruments used to assess pain include the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), the Australian/Canadian Osteoarthritis Hand Index (AUSCAN), the Knee Injury and Osteoarthritis Outcome Score (KOOS), and the Lequesne algofunctional index pain subscales (Bischoff et al., 2003, Bellamy, 2014). WOMAC is a knee and hip OA-specific tool, comprised of 24 questions in the domains of pain, physical function, and stiffness (Bellamy et al., 1988). The WOMAC pain subscale is the most commonly presented pain outcome in knee OA randomised controlled trials (RCTs) published in high impact journals (Juhl et al., 2012). However, there is correlation between the pain and disability subsections, which may indicate that WOMAC pain is a measure of pain during physical activity, rather than a tool for measuring the overall pain experience (Gandek, 2015). Furthermore, it does not assess pain during vigorous activities, so may not be appropriate for more physically able people (Gandek, 2015). AUSCAN is comparable to WOMAC, only that it is targeted at hand OA (Bellamy et al., 2002). Its pain subscale is commonly used to assess pain in hand OA studies (Visser et al., 2015). KOOS is a knee-specific tool that covers five domains: pain, other symptoms, function in daily living, function in sport and recreation, and knee related QoL (Roos et al., 1998). It is an extension of the WOMAC questionnaire as it contains the full WOMAC questions within it, but also assesses vigorous functional outcomes and HR-QoL (Roos et al., 1998). Finally, the Lequesne algofunctional index is an 11-item questionnaire for knee and hip OA that assesses pain, distance walked, and joint-related ADLs (Lequesne et al., 1987).

Unfortunately these tools ask only about recent pain, thereby limiting their ability to capture overall trends in pain.

Additional tools have been developed that aim to capture aspects of the OA pain experience that are not assessed by the previously listed tools. For example, Hawker et al. (2008a) demonstrated that pain in OA is characterised by the quality of pain, its impact on function, its effect on sleep, and its psychological consequences. The aforementioned tools are not able to fully capture this complexity and an 11-item questionnaire was therefore developed by OARSI-OMERACT (Hawker et al., 2008b). The Measure of Intermittent and Constant Osteoarthritis Pain questionnaire assesses the intensity of constant and intermittent pain and its association with sleep impairment, anxiety, frustration, and QoL (Hawker et al., 2008b). Another aspect of OA pain that is not adequately captured by commonly used tools is the presence of neuropathic-like pain. For this, a modified knee-specific PDQ can be used to identify and quantify neuropathic-like pain in people with OA (Hochman et al., 2011).

Focusing only on the *intensity* of pain in OA has been likened to describing the visual world only in terms of the influx of light, whilst ignoring colour, texture, and the countless other dimensions that make up the visual experience (Melzack, 1975). This statement aptly captures the difficulty faced in OA research. It is recognised that pain is of utmost importance to people with OA, however, it is by its very nature difficult to quantify. As a researcher one becomes accustomed to visualising OA pain as a numerical value; this is necessary for the examination and comparison of treatments. But what does this number mean? A single number cannot capture a person's whole experience of pain, and although multidimensional pain tools may capture a greater part of the pain experience, it remains a limitation that underlies efficacy analyses in OA.

1.1.6.2 Disability

Disability can be assessed using the function subscale of the OA-specific tools previously described or using joint-specific tools that are disease generic. In knee OA RCTs, the WOMAC function subscale is the

most frequently used patient-reported outcome used to describe disability (Juhl et al., 2012). It measures a person's difficulty in performing ADLs, however, its use is limited as it does not assess more physically demanding tasks (Collins et al., 2011). The KOOS questionnaire overcomes this limitation and may be used as an alternative measure of disability in knee OA trials (Collins et al., 2011). Finally, the Lequesne algofunctional index or the Oxford Knee Score (OKS), which is a post-TJR tool, may also be used (Collins et al., 2011).

In hand OA trials, the most frequently used instruments to assess functional status are AUSCAN and the Functional Index for Hand OA (FIHOA) (Visser et al., 2015). Unlike the three subsets of AUSCAN (pain, stiffness, function), FIHOA focuses only on hand function using ten questions (Poole, 2011). Similarly, the Cochin Hand Function Scale may be used (Poole, 2011). It assesses the difficulty in performing tasks related to the kitchen, dressing, hygiene, office, and other activities (Poole, 2011).

1.1.6.3 Global assessment

PGA is a core outcome for trials in OA, however, there is not a well-established method for how to measure this (Altman et al., 1996, Bellamy, 2014). PGA can address a variety of aspects of OA, such as overall health status, symptom severity, or disease severity (Bellamy, 2014). It can be used to assess a specific joint, OA in general, or the patient as a whole (Bellamy, 2014). Furthermore, it can be used to assess the current status or a change in status (Bellamy, 2014). It is the careful phrasing of the question that allows researchers to capture the outcome they are interested in. There is no internationally-preferred wording or scaling format (Bellamy, 2014), but it can be measured using a VAS or Likert scale in response to a question such as: "Considering all the ways your OA [joint site] affects you, how are you doing [time frame]?" (Altman et al., 1996).

Although not a core outcome measure, physician's assessment of global status can be assessed similarly to the PGA (Altman et al., 1996). The physician's global assessment may add further insight into the status of OA, as they will have access to radiographic, laboratory,

and clinical examination findings (Bellamy, 2014). Interestingly, mean patient's and physician's global assessment scores have been shown to be closely matched in clinical trials of musculoskeletal (MSK) conditions (Evangelou et al., 2008).

1.1.6.4 Quality of life

Traditionally, HR-QoL has been assessed in OA trials using generic measures such as the short form (SF)-36, the Sickness Impact Profile, the Euroqol, or the Nottingham Health Profile (Bischoff et al., 2003, Altman et al., 1996, Lane et al., 2015). Another approach to assessing HR-QoL is to use OA-specific composite tools that contain HR-QoL domains, such as KOOS (Roos et al., 1998). Alternatively, the Osteoarthritis Knee and Hip Quality of Life questionnaire may be used (Rat et al., 2006). It is a 43 item questionnaire that assesses pain, physical functioning, mental health, social functioning, and social support (Rat et al., 2006).

1.1.6.5 Structural

Medical imaging is used to screen participants, establish the diagnosis, and monitor therapy in clinical trials (Eckstein et al., 2014). As an outcome measure, it is generally reserved for trials of structure-modifying interventions (Kloppenburger et al., 2015, Lane et al., 2015, McAlindon et al., 2015). Plain film radiographs are ordinarily used in OA trials and are most often classified according to the KL grading system (Eckstein et al., 2014). There is, however, an ongoing debate on the adequacy of radiographs as an outcome measure for trials of structure modifying drugs (Guermazi et al., 2013). Radiographs lack the sensitivity to detect structural changes over time and lack specificity as joint space narrowing (JSN) may be due to a variety of non-OA related factors, such as meniscal extrusion (Guermazi et al., 2013). Furthermore, many of the structural changes associated with pain in OA, such as BMLs and synovitis, are not apparent on radiographs (Guermazi et al., 2013). These markers can, however, be visualised using MRI (Figure 1-11) (Guermazi et al., 2013).

MRI is more sensitive to change than plain radiographs, and it has therefore been proposed that MRI be used as the imaging method of choice in structure modifying trials (Eckstein et al., 2014).

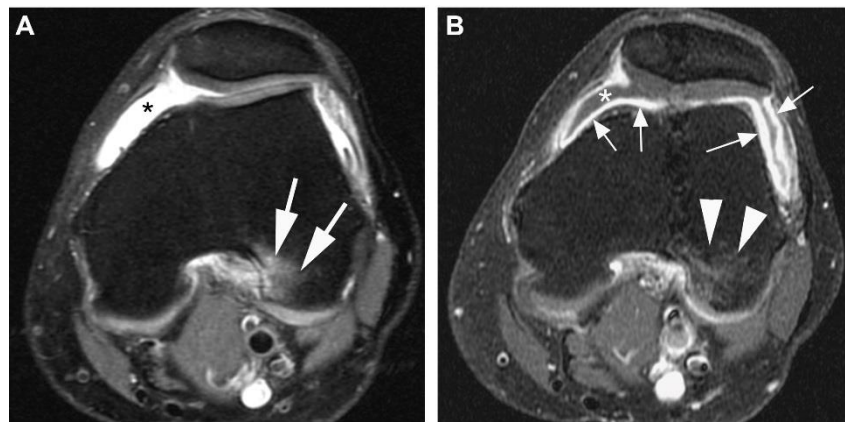


Figure 1-11 - BMLs and signs of inflammation seen in an MRI of a knee with OA
A. A non-enhanced MRI showing effusion-synovitis (*) and BML (arrow). B. A contrast-enhanced MRI of the same individual which differentiates between synovial thickening (arrows) and effusion (*). BMLs (arrowheads) seen also in the contrast-enhanced MRI.
Image reproduced with permission. Reference: (Hayashi et al., 2016)
BML, bone marrow lesion; MRI, magnetic resonance imaging; OA, osteoarthritis

1.1.6.6 Others

A variety of other symptoms, semi-objective, and objective outcomes can be assessed in OA trials. Early morning stiffness or stiffness after inactivity is commonly reported in clinical trials, and can be measured using a VAS or verbal rating scale (Bischoff et al., 2003). Semi-objective measures reported in OA trials include analgesic consumption, physical examination, and performance tests such as timed walk and muscle strength (Bischoff et al., 2003). Objective clinical measures used include joint count, joint circumference, and range of motion (Bischoff et al., 2003).

1.2 Introduction to topical NSAIDs

1.2.1 History and Definitions

NSAIDs are chemically heterogeneous compounds that provide pain relief and reduce inflammation in a variety of conditions (Patrono, 2015). Their use dates back to 400 B.C. when Hippocrates prescribed willow extract to treat fever and inflammation (Rao and Knaus, 2008). The active ingredient, salicin, was identified in the 17th century and was developed into acetylsalicylic acid (aspirin) in 1899 (Rao and Knaus, 2008). Years later, non-salicylic NSAIDs, such as phenylbutazone and indomethacin, were developed (Altman et al., 2015). However, it was not until the 1970's that the mechanism of action of NSAIDs was discovered, thereby driving the development of new NSAIDs (Rao and Knaus, 2008). Although found to be effective analgesics, the dose-dependent gastrointestinal (GI), cardiovascular (CV), and renal AEs of oral NSAIDs limited their use (Altman et al., 2015). This encouraged work into the modification of their pharmacological properties, the provision of novel modes of delivery, and the combination of the drugs with gastro-protective agents (Altman et al., 2015). One approach taken to limit systemic exposure and minimise dose-dependent side effects was through the development of topical NSAIDs (Altman et al., 2015).

A variety of topical NSAIDs are available, including diclofenac, felbinac, ibuprofen, indomethacin, ketoprofen, and piroxicam. These come in a variety of formulations, including gels, creams, solutions, and patches. They are available as over-the-counter or prescription-only medications, and are becoming increasingly commonplace in the management of a variety of acute and chronic pain conditions. In fact, older adults appear to prefer topical over oral NSAIDs for mild, transient, localised knee pain (Carnes et al., 2008).

1.2.2 Possible mechanism of action

NSAIDs are primarily thought to act by inhibiting the COX enzymes 1 and 2, thereby preventing the production of prostanoids such as PGE₂, PGI₂, and thromboxane A₂ (Patrono, 2015). COX-1 is constitutively expressed, whilst COX-2 is induced by inflammatory stimuli (Patrono, 2015). Traditionally, COX-1 is viewed as physiological whilst COX-2 is pathological, however, this is an oversimplification (Abramson, 2003). COX-2 has been found to be a physiological modulator of renal, brain, ovarian, uterine, and CV function (Rouzer and Marnett, 2009). In addition, work by Gilroy et al. (1999) in carrageenin-induced pleurisy in rats showed that COX-2 may be involved in inflammation resolution in the latter stages of the inflammatory process. Despite its physiological roles, COX-2 is still the predominant COX isoform expressed in injured tissue and is responsible for the majority of prostanoid production in inflammation (Burian M, 2005). In OA, COX-2 is prominently expressed in the synovium, fibrocartilage of osteophytes, and knee vasculature (Koki et al., 2002).

The COX enzymes are responsible for the production of prostaglandins which contribute to the inflammatory reaction and to peripheral sensitisation (Burian M, 2005). Prostaglandins are hyperalgesic agents that, by themselves, do not cause spontaneous pain (Burian M, 2005). They are potent sensitising agents that enhance nociceptive signal transduction (peripheral sensitisation) and transmission (central sensitisation) (Burian M, 2005). They exert their mechanism of action both peripherally and within the CNS. NSAIDs are believed to act by normalising the hyperalgesic effects of inflammation, rather than by directly elevating normal pain thresholds (Burian M, 2005). As such, it may be more appropriate to view NSAIDs as anti-hyperalgesic agents rather than analgesic agents (Burian M, 2005).

COX-inhibition is the most widely accepted mechanism of action of NSAIDs, however, a multitude of other mechanisms have also been proposed and are listed in Table 1-3 (Hagen and Baker, 2017).

Table 1-3 - Putative mechanisms of action of NSAIDs

Putative mechanisms of action of NSAIDs
COX inhibition (primary)
Leukotriene synthesis inhibition
Phospholipase A2 inhibition
Arachidonic acid level modulation
N-methyl-D-aspartate pathway inhibition
Increase in β -endorphin levels
Peroxisome proliferator-activator receptor gamma inhibition
Reduction in plasma and synovial substance P and interleukin-6 levels
Thromboxane-prostanoid receptor inhibition
Acid-sensing ion channel inhibition

Putative mechanisms of action as per (Hagen and Baker, 2017)
COX, cyclo-oxygenase; NSAIDs, non-steroidal anti-inflammatory drugs

1.2.2.1 Topical mechanism of action in OA

Topical NSAIDs are thought to act by deep penetration to periarticular structures and delivery via local blood supply to intra-articular structures (Hagen and Baker, 2017). Applied to the skin overlying the affected joint, they are believed to result in effective concentrations at the target joint whilst limiting excessive systemic exposure (Haroutiunian et al., 2010). To exert their effect, the NSAIDs must therefore first and foremost penetrate the skin barrier. The extent of their penetration will vary according to the drug, different salt of the same molecule, amount applied, frequency of application, formulation used, and presence of any penetration enhancers (Haroutiunian et al., 2010). In addition, marked differences in absorption between individuals are likely to exist secondary to variations in age, skin disease or damage, hydration of epidermis, and local blood flow (Hagen and Baker, 2017).

Following application, topical NSAIDs penetrate through the skin and permeate underlying tissues. They may accumulate for prolonged periods of time in the skin or subcutaneous tissue, resulting in sustained release of the drug over time (Hagen and Baker, 2017). Upon reaching the dermal level, the drug may be absorbed into local blood

vessels for distribution to deeper tissues or may be transported from the dermal microcirculation to the systemic circulation (Hagen and Baker, 2017). Alternatively, the drug may continue to diffuse through the tissues to reach deeper structures, preferentially distributing to areas of inflammation (Hagen and Baker, 2017). The exact mechanisms through which topical NSAIDs reach the target tissues remains unclear.

Following topical application, NSAIDs are distributed to local and deep joint tissues and the systemic circulation. Locally, there is evidence that topical NSAIDs are distributed to the subcutaneous tissue, muscle, synovium, and cartilage (Hagen and Baker, 2017, Haroutiunian et al., 2010). In addition, NSAIDs are also present at lower concentrations in the synovial fluid (Haroutiunian et al., 2010, Hagen and Baker, 2017). However, research into the levels of NSAIDs in joint tissues after topical administration is sparse and inconsistent (Hagen and Baker, 2017). In addition, it is unclear what concentrations of NSAIDs must be achieved within target tissues to exert a therapeutic effect (Hagen and Baker, 2017). As such, it is difficult to understand their mechanism of action in the periphery. Their clinical efficacy suggests that they do achieve therapeutic concentrations in and around the joint. As established in section 1.1.4.2.1, the environment in the osteoarthritic joint drives hyperalgesia through the release of mediators, such as prostaglandins. Topical NSAIDs may therefore reduce pain in OA by reducing prostaglandin levels in and around the joint, thereby normalising nociceptor hypersensitivity.

Systemically, topical application achieves maximum NSAID levels that are approximately 0.2 – 8% of those achieved with oral exposure (Hagen and Baker, 2017). Their presence in the plasma may allow topical NSAIDs to exert some anti-hyperalgesic action in the CNS, however, their effects are likely predominantly exerted in the periphery where they reach higher concentrations.

1.2.3 Recommendations and efficacy

1.2.3.1 Recommendations

Topical NSAIDs are recommended for the pharmacological management of hand and knee OA by ACR (Hochberg et al., 2012), EULAR (Jordan et al., 2003, Kloppenburg et al., 2018), OARSI (McAlindon et al., 2014), and NICE (NICE, 2014). Further details of the recommendations are described in Table 1-4.

Table 1-4 - Details of topical NSAID recommendations in OA

Group	Joint	Details of topical NSAID recommendation
ACR	Hand	Conditionally recommend topical NSAID and/or topical capsaicin, oral NSAID, tramadol
	Knee	Topical NSAID (or oral NSAID or intra-articular steroid) should be used if full-dose paracetamol has not provided sufficient pain relief Topical NSAIDs should be used ahead of oral NSAIDs for people ≥ 75 years of age (strongly recommended)
EULAR	Hand	Topical treatments are preferred over systemic treatments. Topical NSAIDs are the first choice pharmacological topical treatment
	Knee	Topical treatments (topical NSAID or capsaicin) are effective and safe
NICE	Hand, knee	Topical NSAIDs (or paracetamol) are first line pharmacological treatments
OARSI	Knee	Topical NSAIDs are appropriate for individuals with knee-only OA Uncertain recommendation for individuals with multiple-joint OA

OA recommendations for topical NSAIDs as per (Hochberg et al., 2012, Jordan et al., 2003, NICE, 2014, McAlindon et al., 2014, Kloppenburg et al., 2018)
ACR, American College of Rheumatology; EULAR, European League Against Rheumatism; NICE, National Institute for Health and Care Excellence; OARSI, Osteoarthritis Research Society International

1.2.3.2 Efficacy

Topical NSAIDs are effective for pain in OA and have been shown to have a small-to-moderate effect size (ES) compared to placebo. Several MAs of topical NSAIDs in OA have been published in the last two decades, all of which support their superiority over placebo (see Table 1-5). Although differences in inclusion/exclusion criteria, time points, and statistical methods hinder direct comparison of the MAs, they allow an insight into the evidence base for the efficacy of topical NSAIDs.

For this thesis, topical NSAIDs are examined as a single treatment; different formulations and drugs are pooled together as one entity. This is the commonest approach, and the MAs using this approach are listed in Table 1-5. Early analyses predominantly pooled salicylates with topical NSAIDs, however, salicylates are now accepted as being distinct entities as their action is limited to that of a rubefacient (Derry et al., 2014). Grouping all topical NSAIDs together reflects clinical guidelines, which do not differentiate between different drugs or formulations in their recommendations. However, different topical NSAIDs can also be examined separately. This approach was adopted by Cochrane in their most recent review (Derry et al., 2016). They presented the efficacy of topical diclofenac and ketoprofen individually, and found that both were superior to placebo. In addition, MAs have been conducted that are focused on only one topical NSAID type (Towheed, 2006, Deng et al., 2016). Finally, the relative efficacy of different topical NSAIDs was recently examined using a network meta-analysis (NMA) (Zeng et al., 2018). Of all available topical NSAIDs, diclofenac patch had the highest probability of being the best treatment (95.7%), followed by topical ibuprofen (88.81%) and piroxicam (75.99%) (Zeng et al., 2018). Not all topical NSAIDs were superior to placebo (e.g., nimesulide, ketoprofen, eltenac, indomethacin, etoricoxib), suggesting important differences in the efficacy of topical NSAIDs or their associated carriers (Zeng et al., 2018).

Table 1-5 - MAs of topical NSAID efficacy in OA or chronic MSK pain

Reference	Population	Time (week)	n	Measure	Effect (95%CI)
Moore et al. (1998)	Chronic pain*	2	1,097	RR NNT	2.0 (1.5 to 2.7) 3.1 (2.7 to 3.8)
Mason et al. (2004)	Chronic MSK pain	2	1,502 (567 with knee OA)	RR Knee OA only	1.92 (1.68 to 2.18) 2.02 (1.57 to 2.60)
Lin et al. (2004)	Clinical or radiographic OA	1-4	Week 1: 1,000 Week 2: 893 Week 3: 442 Week 4: 558	ES	Week 1: 0.41 (0.16 to 0.66) Week 2: 0.40 (0.15 to 0.65) Week 3: 0.05 (-0.11 to 0.22) Week 4: 0.04 (-0.11 to 0.19)
Biswal et al. (2006)	Radiographic knee OA	>4	709	ES	0.28 (0.14 to 0.42)
Bjordal et al. (2007)	ACR or radiographic knee OA	Peak (mean: 1.6)	749	WMD (VAS)	11.6 (7.4 to 15.7)
Zhang et al. (2010b)	OA			ES	0.44 (0.27 to 0.62)
Derry S (2012)	Chronic MSK pain	2-12	Week 2-3: 917 Week 4-6: 810 Week 8-12: 2,440	RR NNT	Week 2-3: 1.9 (1.6 to 2.4) Week 4-6: 1.7 (1.4 to 2.1) Week 8-12: 1.2 (1.1 to 1.3) Week 2-3: 5.5 (4.2 to 8.1) Week 4-6: 5.8 (4.2 to 9.1) Week 8-12: 10 (7.3 to 17)
Zeng et al. (2018)	OA	1-12	6,969	ES	0.30 (0.20 to 0.40)

Effect presented for topical NSAID versus placebo. RR>1 and ES>0 favours topical NSAID over placebo for clinical success or pain relief

*single joint arthritis and rheumatological disorders

ACR, American College of Rheumatology; CI, confidence interval; ES, effect size; MA, meta-analysis; MSK, musculoskeletal; NNT, number needed to treat; OA, osteoarthritis; RR, relative risk; VAS, 0-100 mm visual analogue scale; WMD, weighted mean difference

Overall, topical NSAIDs appear to be most effective in the first two weeks of treatment, with one MA indicating that they are at their most effective at 1.6 weeks of treatment (Bjordal et al., 2007). The first MA focusing exclusively on participants with OA found that topical NSAIDs and salicylates were only superior to placebo at one and two weeks, and not at three or four weeks of follow-up (Lin et al., 2004). However, the analyses at three and four weeks were limited to salicylates and eltenac, a topical NSAID that is now limited only to veterinary use (likely due to its limited efficacy). Subsequent MAs have shown that topical NSAIDs do continue to exert an effect greater than that of placebo at four weeks and beyond (Biswal et al., 2006, Derry S, 2012). Indeed, the 2012 Cochrane review found that topical NSAIDs were superior to placebo at all duration of follow-up, however, longer duration studies had a higher (worse) number needed to treat (NNT) (Derry S, 2012). A similar trend was reported for topical diclofenac in the most recent Cochrane review (Derry et al., 2016).

The efficacy of topical NSAIDs has predominantly been examined in knee OA, however, different populations have been included in the listed MAs. For example, a subset of MAs, including the Cochrane reviews, are for chronic MSK pain (Derry et al., 2016, Derry S, 2012, Moore et al., 1998, Mason et al., 2004). In these MAs, the majority of participants had primary OA and the RR of clinical success was similar in knee OA compared to other MSK disorders (Mason et al., 2004). Focusing only on OA, Lin et al. (2004) conducted subgroup analyses by joint site and found that topical NSAIDs had a greater ES for hand than knee OA (Lin et al., 2004). However, the applicability of this analysis to current practice is questionable as the hand data were limited to one study that examined the efficacy of a salicylate only two hours after application (Lin et al., 2004). Subgroup analyses by joint site have not been conducted in more recent MAs.

Finally, the assertion of topical NSAIDs' efficacy has been based on establishing their superiority over placebo. However, the efficacy of topical NSAIDs can also be assessed with regards to their effects compared with oral NSAIDs or their overall treatment effect (change from baseline in treatment arm). All MA estimates of topical versus oral

NSAIDs indicate that there are no differences between the two formulations in terms of efficacy (Mason et al., 2004, Lin et al., 2004, Derry S, 2012, Derry et al., 2016). The most recent estimate, generated through a Cochrane review, found that the RR between topical and oral NSAIDs was 1.03 (95%CI 0.95 to 1.12) (Derry et al., 2016). With regards to the overall treatment effect of topical NSAIDs in OA, Zou et al. (2016) estimated it at 1.37 (95%CI 1.19 to 1.55). This is equivalent to an average decrease in pain score (0-10 cm VAS) of 3.4 cm. 85% of the overall treatment effect of topical NSAIDs was attributed to contextual effects (0.85, 95%CI 0.77-0.93) (Zou et al., 2016).

1.2.4 Safety

Overall, topical NSAIDs appear to be a safe long-term option for individuals with OA. A MA of observational studies found that those on topical NSAIDs did not report significant increases in AEs compared to non-users (OR 1.19, 95%CI 0.92 to 1.55) (Zeng et al., 2018).

The AEs reported for topical NSAIDs can be divided into local (application site) reactions and systemic AEs. Local AEs are the most commonly occurring AE associated with topical NSAID use and involve reactions such as dry skin, erythema, irritation, paraesthesias, and pruritus (Makris et al., 2010). On the other hand, the most commonly reported systemic AEs associated with topical NSAID use are GI complaints and headaches (Makris et al., 2010). The major area of concern and focus of systemic AE examination is upper GI events.

1.2.4.1 Local AEs

The presence of local AEs in topical NSAID users varies considerably between RCTs, with dry skin reported in 0.79 to 39.3% of participants and rashes reported in 0.8 to 13% of participants (Makris et al., 2010). The incidence of local AEs is likely to vary between drugs, populations, and as a result of variability in the recording and reporting of AEs. Nevertheless, it is evident that a proportion of people using a topical NSAID will experience an application site reaction. However, whether or not topical NSAIDs associate with significantly higher rates of local AEs than placebo/vehicle is unclear.

Most MAs point to a similar risk of local AEs in topical NSAID and placebo/vehicle groups (Moore et al., 1998, Mason et al., 2004, Lin et al., 2004, Zeng et al., 2018). These MAs have examined topical NSAIDs as a single entity, and it is the evidence for individual topical NSAIDs that is most conflicting. Given that topical NSAIDs differ by drug, vehicle components, formulations, and penetration enhancers (Makris et al., 2010), all of which may affect the incidence of local AEs, one would expect the incidence of local AEs to differ between drugs. As expected, the 2012 Cochrane review found that topical diclofenac users had an 84% higher risk of local AEs than those on vehicle (RR

1.84, 95%CI 1.54 to 2.21), driving an overall increased risk for all topical NSAIDs use compared to vehicle (RR 1.69, 95%CI 1.45 to 1.98) (Derry S, 2012). Similarly, the most recent Cochrane review found that topical diclofenac, but not ketoprofen, had a significantly increased risk of local AEs compared to vehicle (RR 1.8, 95%CI 1.5 to 2.2) (Derry et al., 2016). No conclusions were drawn regarding other topical NSAIDs. In addition, MAs focusing on topical diclofenac support the finding that it is associated with a greater risk of local AEs than vehicle (Towheed, 2006, Deng et al., 2016). In contrast, a recent NMA did not find any significant differences in local AEs when comparing across all topical NSAID groups, including diclofenac (Zeng et al., 2018). However, the MAs vary by inclusion criteria and statistical methods and these differences may explain the discrepancy in findings. Despite its conflicting nature, the evidence suggests that topical NSAIDs in general do not appear to increase the risk of local AEs significantly more than placebo/vehicle, although topical diclofenac may do.

Overall, choosing a topical NSAID is associated with a greater risk of local AEs than an oral NSAID. Topical application infers an approximately three- to five-fold increased risk of local AEs compared to oral NSAIDs (Mason et al., 2004, Lin et al., 2004, Derry S, 2012, Derry et al., 2016). The application of possible irritants directly to the skin may explain the increased risk, although the aforementioned placebo/vehicle controlled evidence suggests that it may not be the NSAID *per se* that increases the risk. Although not formally examined in a MA, one SR indicated that the incidence of local AEs was lower in placebo compared to vehicle comparator groups (Makris et al., 2010). With many vehicles containing substances such as dimethyl sulfoxide (DMSO), pluronic lecithin organogel base, or isopropyl alcohol, it may be that these substances contribute to the increased incidence of local AEs (Makris et al., 2010).

1.2.4.2 Systemic AEs

The major concern with NSAID use is their association with upper GI AEs, and establishing their relationship has been widely investigated. MAs of RCTs indicate that topical NSAIDs do not increase the risk of

systemic or GI AEs compared to vehicle/placebo (Moore et al., 1998, Mason et al., 2004, Lin et al., 2004, Derry S, 2012, Derry et al., 2016, Zeng et al., 2018). However, RCTs may be limited in their ability to detect rare AEs that occur with longer term NSAID exposure.

Observational studies are more appropriate for this purpose due to the longer duration of follow-up, greater power, and heterogeneity in participants. MA results from observational studies indicate that even compared to non-users, topical NSAIDs do not increase the risk of GI AEs (Zeng et al., 2018).

The systemic AEs associated with NSAID use are generally regarded as being secondary to their systemic inhibition of COX enzymes. Topical, rather than oral, administration results in significantly reduced concentrations of NSAIDs in the blood (Hagen and Baker, 2017). This may explain why, compared to oral administration, topical NSAID use has a 26-34% lower risk of GI AEs (Lin et al., 2004, Derry S, 2012, Derry et al., 2016). Only one MA has compared the overall risk of systemic AEs in topical versus oral NSAIDs, and it did not find any statistically significant differences (Mason et al., 2004).

1.3 Introduction to capsaicin

1.3.1 History and definitions

Capsaicin is the principal pungent component of the fruits from the genus *capsicum*; it is what makes chilli peppers hot (Szolcsanyi, 2014). Chilli plants are indigenous to the Americas and have been consumed since 7000 BC (Szolcsanyi, 2014). Aside from its culinary uses, capsicum has long been used also for its analgesic properties. The Native Americans would rub pepper pods on their gums to soothe toothache, and extracts prepared from the plants were used for analgesia in Europe in the 19th century (Szolcsanyi, 2014, Szallasi and Blumberg, 1999). Capsaicin was first purified in 1876 and its chemical structure was determined in 1919 (Szolcsanyi, 2014).

Capsaicin has the distinctive property that it has the ability to both produce and alleviate pain. Initially, capsaicin was predominantly used to understand pain processing and modulation (O'Neill et al., 2012). Even now, it is used to reversibly invoke central sensitisation for research, producing symptoms such as allodynia and secondary hyperalgesia (O'Neill et al., 2012). However, with repeated or high concentration applications, capsaicin has the opposing effect: it acts as an analgesic. This characteristic of capsaicin, which was capitalised on centuries ago, has led to the development of topical capsaicin as an analgesic agent.

In the modern age, topical capsaicin was first used therapeutically as a low-concentration (0.025-0.075%) cream, applied several times per day for multiple weeks (Haanpää and Treede, 2012). Its unpleasant irritant effect on mucous membranes drew researchers to look for alternative solutions that would cause pain relief without the side effects associated with repeated, extended use. Attempts to develop structural analogues with the anti-nociceptive action of capsaicin, but without the irritant effect, failed (Haanpää and Treede, 2012). In addition, the production of high-concentration creams (10%) were halted due to extreme pain on application and high risks of contamination to the patient and healthcare professional (Haanpää and Treede, 2012). In the

last decade, a high concentration (8%) capsaicin patch has been developed and approved for certain chronic NP conditions.

Topical capsaicin creams are available in low concentrations of 0.025% or 0.075%, although 0.25% creams are available (Derry and Moore, 2012). In addition, lower concentrations of capsaicin can also be found in topical rubefacients (Derry and Moore, 2012). High-concentration capsaicin is found as an 8% patch which, unlike the low-concentration capsaicin, is applied by a healthcare practitioner using pre-treatment with local anaesthesia (Derry et al., 2017). The patch is left on for 30-60 minutes, followed by careful cleansing of the area (Derry et al., 2017). This procedure can be repeated after 12 weeks, if required (Derry et al., 2017).

The contextual factors related to the use of high-concentration capsaicin therefore differ markedly from those of topical NSAIDs or low-concentration capsaicin cream. Whilst topical NSAIDs and capsaicin cream involve multiple applications per day, done by the patient themselves, application of high-concentration capsaicin is more invasive. It requires patients to attend clinic for several hours, is applied by specialist healthcare professionals, and is only done every three months.

1.3.2 Possible mechanism of action

Capsaicin is a potent, highly selective, exogenous agonist of the transient receptor protein channel V, otherwise known as the TRPV1 (O'Neill et al., 2012, Anand P, 2011). The TRPV1 is a non-selective ligand-gated channel that is activated by capsaicin, temperatures above 43°C, pH lower than 6.0, and various lipids, proteins, and inflammatory mediators (Anand P, 2011, O'Neill et al., 2012). In the periphery, TRPV1 is mainly found in nociceptive fibres (C fibres and some Aδ fibres), where capsaicin exerts its effect following topical administration (Anand P, 2011, O'Neill et al., 2012). Capsaicin binding to the TRPV1 channel causes ion channel activation, Na⁺ and Ca²⁺ influx, depolarisation, and transduction of the nociceptive signal, which is experienced as a stinging, burning, or itching sensation (O'Neill et al., 2012, Anand P, 2011).

As previously mentioned, capsaicin is used in humans both to elicit pain in experimental pain models and to relieve pain in a variety of pain states (O'Neill et al., 2012). Its analgesic action is brought about through a cascade of events that occur after a single high-concentration application or repeated low-dose applications (Anand P, 2011). Capsaicin was initially thought to act by depleting skin substance P, a substance which was believed to play a pivotal role in pain neurotransmission (Anand P, 2011). This hypothesis has now been replaced with that of defunctionalisation and denervation, which are elegantly able to describe the apparent reductions in substance P following capsaicin administration (Anand P, 2011).

Initially, the binding and continued presence of capsaicin causes desensitisation of nociceptors as neuropeptides become depleted and ion channels become inactivated (O'Neill et al., 2012). As a result, nociceptive neurons become less responsive to capsaicin. These effects occur immediately following application and are only maintained in the presence of capsaicin so do not explain the prolonged analgesic effect of capsaicin (Milner and Doherty, 2015). The prolonged analgesic effect is instead explained by two main processes: defunctionalisation and denervation.

Defunctionalisation is the impairment of nociceptive function that is maintained after capsaicin is no longer present in the tissues. In contrast to desensitisation, defunctionalisation encompasses a reduction in nociceptor responsiveness to a variety of stimuli, and not just capsaicin. Defunctionalisation occurs through multiple mechanisms. Firstly, the raised intra-cellular Ca^{2+} levels activate calcium-dependent enzymes which cause depolymerisation of cytoskeletal components (Anand P, 2011, Milner and Doherty, 2015). Secondly, the Cl^- influx that inevitably accompanies cation influx causes osmotic swelling of nociceptors (Anand P, 2011, Milner and Doherty, 2015). Finally, unrelated to TRPV1, capsaicin directly inhibits mitochondrial respiration (Anand P, 2011, Milner and Doherty, 2015). These events all result in reversible, but prolonged, impairment of nociceptive activity.

Denervation of the epidermis and dermis following capsaicin application is thought to occur secondary to the previously described processes, which render the nerve endings unable to maintain plasma integrity and therefore collapse (Anand P, 2011).

1.3.2.1 Mechanism of action in OA

Topical application of capsaicin is believed to act locally and not through transdermal delivery to deeper tissues (Anand P, 2011). Its action in conditions, such as OA, where the source of the pain is not in the superficial skin layers, is not fully understood. It is postulated that capsaicin acts through a CNS modulatory mechanism (Anand P, 2011). Defunctionalisation of cutaneous nociceptors that spontaneously fire, are hypersensitive, or maintain central sensitisation through the provision of trophic excitatory factors is thought to be the main mechanism whereby capsaicin causes pain relief (Anand P, 2011).

1.3.3 Recommendations and efficacy

1.3.3.1 Recommendations

In the UK, low concentration capsaicin is licensed for pain relief in PHN, PDN, and OA (BNF, 2018). The recommendations for topical capsaicin use in OA are listed in Table 1-6.

Table 1-6 - Details of topical capsaicin recommendations in OA

Group	Joint	Details of topical capsaicin recommendation
ACR	Hand	Conditionally recommend topical capsaicin and/or topical NSAID, oral NSAID, tramadol
	Knee	Conditionally recommend that people with knee OA should <i>not</i> use topical capsaicin
EULAR	Hand	Topical treatments are preferred over systemic treatments. Topical capsaicin is mentioned as an option, but no recommendations for or against its use are given
	Knee	Topical treatments (topical NSAID or capsaicin) are effective and safe
NICE	Hand, knee	Topical capsaicin should be considered as an adjunct to core treatments
OARSI	Knee	Topical capsaicin is appropriate for individuals with knee-only OA without relevant co-morbidities Uncertain recommendation for individuals with multiple-joint OA or relevant co-morbidities

Recommendations as per (Hochberg et al., 2012, Jordan et al., 2003, NICE, 2014, McAlindon et al., 2014, Kloppenburg et al., 2018)
ACR, American College of Rheumatology; EULAR, European League Against Rheumatism; NICE, National Institute for Health and Care Excellence; OARSI, Osteoarthritis Research Society International

The recommendations for topical capsaicin use in OA vary across guidelines, which provide a spectrum in terms of the strength of recommendations. On one hand, EULAR recommends topical capsaicin for knee OA, placing equal importance on topical capsaicin as for topical NSAIDs (Jordan et al., 2003). NICE guidelines limit the role of topical capsaicin somewhat, recommending it as an *adjunct* to core treatments (topical NSAID or paracetamol) (NICE, 2014). ACR

conditionally recommend topical capsaicin for hand OA, but is the only guideline to conditionally advise *against* its use in knee OA (Hochberg et al., 2012). In contrast, OARSI recommend topical capsaicin for knee-only OA without relevant comorbidities, although its use in multi-joint OA or knee-only OA with relevant comorbidities remains uncertain (McAlindon et al., 2014). Similarly, EULAR recommendations for hand OA mention capsaicin, but do not provide any recommendations regarding its use (Kloppenburger et al., 2018).

Although the recommendations vary, the general consensus appears to be that topical capsaicin may be considered in knee and hand OA. There is greater confidence in the recommendation of topical NSAIDs, and capsaicin is therefore generally considered as a second-line topical treatment.

1.3.3.2 Efficacy

Despite decades of capsaicin use, there is limited evidence for its efficacy in OA. The number of MAs conducted in OA are few and far between, and are limited to small numbers of trials/participants (Table 1-7).

The first MA of capsaicin in OA was conducted in 1994 and found that it was effective for arthritis (OA and RA), with an OR for improvement of 4.36 (95%CI 2.77 to 6.88) (Zhang and Li Wan Po, 1994). Rather than examining pain, improvement was assessed with regards to articular tenderness and physician's global assessment. A decade later, the relative benefit of capsaicin for reducing MSK pain by 50% was found to be 1.5 (95%CI 1.1 to 2.0) (Mason et al., 2004). Although conducted at a later date, the MA did not build on the findings of the previous review *per se*. Differences in selection criteria resulted in the latter MA still including only three RCTs, only two of which had also been included in the 1994 review (one was only available as an abstract before). Five years later, another MA found that the mean difference (MD) between capsaicin and placebo on 0-100 mm VAS for pain at three or four weeks was -7.65 (95%CI -12.69 to -2.61) (Cameron et al., 2009). Again, only three RCTs were analysed for this estimate and there was some overlap between the RCTs included in this and previous MAs. The most

recent MA, conducted two decades after the first, included five RCTs and found a small-to-moderate effect for capsaicin in OA (Laslett and Jones, 2014). The standardised mean difference (SMD) for capsaicin at four weeks was 0.44 (95%CI 0.25 to 0.62) (Laslett and Jones, 2014).

Table 1-7 - MAs of capsaicin efficacy in OA or chronic MSK pain

Reference	Population	Week	n	Measure	Effect (95%CI)
Zhang and Li	Arthritis	4	382	OR	4.36 (2.77 to 6.88)
Wan Po (1994)				RD	0.29 (0.20 to 0.37)
Mason et al. (2004)	MSK pain	4	368	RB	1.5 (1.1 to 2.0)
				NNT	8.1 (4.6 to 34)
Cameron et al. (2009)	OA	3	122	MD (VAS)	-6.81 (-14.58 to 0.96)
		4	179	MD (VAS)	-8.26 (-14.88 to -1.65)
		3 or 4		MD (VAS)	-7.65 (-12.69 to -2.61)
Laslett and Jones (2014)	OA	4	376*	SMD	0.44 (0.25 to 0.62)

Effect presented for capsaicin versus placebo. RB>1, OR>1, MD>0, and SMD>0 favours capsaicin over placebo for clinical success or pain relief

*99 participants are from a cross-over trial and contribute data to capsaicin and placebo group

CI, confidence interval; MA, meta-analysis; MD, mean difference; MSK, musculoskeletal; NNT, number needed to treat; OR, odds ratio; RB, relative benefit; RD, response rate difference; SMD, standardised mean difference; VAS, 0-100 mm visual analogue scale

The perceived efficacy of capsaicin in OA has not changed substantially in the two decades it has been examined. Despite differences in inclusion criteria, outcomes, and methodology, all MA authors concluded that capsaicin was superior to placebo. However, it remains difficult to draw definitive conclusions as to the extent of capsaicin's effectiveness from the selection of MAs as the pooled data have been presented differently for each. In addition, discrepancies in the number of RCTs and participants analysed throughout bring into question the interpretability of each result to current practice. Finally, the adequacy of blinding and consequent validity of the placebo comparator was questioned in one MA due to the warming sensation experienced on capsaicin application (Cameron et al., 2009). Although disregarded in the other MAs, this concern further highlights another source of uncertainty in the efficacy estimate of capsaicin. Inability to successfully blind capsaicin from placebo complicates the examination

of the difference between them (specific treatment effect). However, no evidence is available for the overall treatment effect of capsaicin in OA.

1.3.4 Safety

Topical capsaicin is widely regarded as a safe treatment, with no reports of systemic toxicity (Laslett and Jones, 2014). Due to its lipophilic nature, very low levels of capsaicin are reported systemically for very short durations of time (Anand P, 2011). The effect of topical capsaicin is thought to be limited to the skin (Anand P, 2011), and this is reflected in the nature of the AEs reported with its use.

Topical application of capsaicin causes skin irritation and a burning or stinging sensation. In the MA by Mason et al. (2004), participants on the low concentration cream had a five times greater risk of experiencing a local AE (risk benefit 5.0, 95%CI 2.6 to 9.6) than those on placebo, with about half of the participants on capsaicin reporting a local AE. This finding is reflected in a more recent MA, which reported a risk ratio of 4.22 (95%CI 3.25 to 5.48) for capsaicin cream versus placebo (Laslett and Jones, 2014). The peak incidence for application site AEs is in the first week of application, and their incidence diminishes with continued use (Laslett and Jones, 2014). Contact with mucous membranes in the eyes, mouth, or nose results in irritation, however, the prevalence of this has not been reported. Finally, inhalation of dried capsaicin cream particles can cause a cough, and this is reported by a small proportion of those using capsaicin cream (Mason et al., 2004). Less than one in ten report coughing after using the 0.075% cream, whilst substantially less experience a cough when using the 0.025% cream (Mason et al., 2004).

Although they are focused on NP, the most recent Cochrane reviews on topical capsaicin echo the perceived favourable safety profile of capsaicin. The reviews conclude that using either low or high concentration capsaicin is associated with short-lived localised application site reactions, but that serious or systemic AEs were rare and equally distributed among capsaicin and placebo groups (Derry and Moore, 2012, Derry et al., 2017).

Topical capsaicin appears to be consistently associated with skin irritation following application, but it is an otherwise safe option for patients due to the limited systemic exposure.

1.4 Introduction to precision medicine

1.4.1 Background

Various terms have been used to describe the movement away from a “one size fits all” approach to patient management and towards tailoring treatment to the person. Stratified medicine, individualised medicine, and personalised medicine have all been used, although they have recently been overtaken by the term Precision Medicine, which was popularised by Barack Obama (Collins and Varmus, 2015). The idea that treatment responses should be examined at an individual level was raised in the 1950’s by Hogben and Sim (1953). They argued that given two treatments (A and B, where B has better efficacy *on average*) giving an individual with a certain characteristic treatment A under the assumption that the characteristic is associated with an improved response to A may be more advantageous than giving them B, as A is better for them specifically (Hogben and Sim, 1953). Indeed, whether *averages* can adequately answer a clinician’s question regarding an *individual* was questioned (Hogben and Sim, 1953), and current belief is that it cannot. Precision medicine aims to bridge the gap between population averages and individual responses (Schork, 2015).

Precision medicine aims to increase the benefit gained from a treatment, to reduce harm, and to improve health-care efficiency. A greater understanding of the pathophysiology of diseases and mechanism of action of drugs is allowing the matching of treatments to responsive subgroups (Trusheim et al., 2007). This approach is commonplace in some fields of medicine, for example prescribing antibiotics based on the sensitivity of the bacterial infection or trastuzumab (Herceptin) only to breast cancer sufferers that are HER2/*neu* positive, but is yet to emerge in other areas (Trusheim et al., 2007). For a precision medicine approach to occur, several conditions need to be met. Firstly, the disease must be variable due to a multifactorial aetiology (Trusheim et al., 2007). Secondly, there must exist several treatment options for which there are heterogeneous responses (Trusheim et al., 2007). Finally, there is a need for the

discovery of a clinical biomarker, which is defined as “any diagnostic test or clinical observation that indicates a preferred treatment for a patient subpopulation” (Trusheim et al., 2007). As such, precision medicine invites us to ask the question eloquently posed by Hogben and Sim (1953): “what peculiarities [clinical biomarkers] are common to individuals who respectively respond or fail to respond to one or other treatment?”.

Clinical biomarkers can be anything from histological findings to self-reported patient survey results, and are distinguished by the fact that they are reliably and predictably correlated with different subgroup responses (Trusheim et al., 2007). These subgroup factors are effect modifiers of the treatment response. In other words, they are the defining characteristic of a subgroup that shows a different treatment effect for the same drug compared to another subgroup (Bierma-Zeinstra and Verhagen, 2011). These markers, which guide the precision medicine approach, need to be easily identified by clinicians and should not be more expensive, invasive, or time consuming to ascertain than the treatment itself (Bierma-Zeinstra and Verhagen, 2011). If they cannot be implemented in practice, the subgrouping serves only to understand the mechanism of action of the treatment (Bierma-Zeinstra and Verhagen, 2011).

1.4.2 Methods for identifying treatment responders

Various approaches can be used to identify subgroups and these all provide different quality of evidence. One approach is to identify subgroups with different responses to treatment within RCTs. This may take the form of post-hoc analyses or pre-defined analyses, although these generally provide a weak level of evidence (Bierma-Zeinstra and Verhagen, 2011). Pre-defined analyses are preferable to post-hoc analyses, although both may be subject to a high risk of false positive results (type I error) (Bierma-Zeinstra and Verhagen, 2011). In addition, within-trial evaluations are most commonly underpowered for this purpose, and so are subject to the potential of false negative findings (type II error) (Hingorani et al., 2013).

At a MA level, three main approaches can be taken. Two methods rely on aggregate data extracted from published reports whilst the third uses individual patient data (IPD). The two aggregate data MA approaches are meta-regression (relating the average treatment effect in trials to the characteristics of the trial) and MA of subgroup effects (Bierma-Zeinstra and Verhagen, 2011). Meta-regression is subject to study level confounding and may result in spurious results affected by ecological bias (Hingorani et al., 2013). MAs of subgroup effects are heavily reliant on separate subgroup effects being reported in the publications, which is rarely done and therefore not generally feasible (Koopman et al., 2008). The more robust method for identifying treatment responders is IPD MA (Bierma-Zeinstra and Verhagen, 2011, Hingorani et al., 2013).

An IPD MA has many advantages. Firstly, it overcomes the lack of power associated with many of the other methods (Bierma-Zeinstra and Verhagen, 2011, Hingorani et al., 2013). Secondly, whilst subgroup factors are rarely reported in published reports, this information is commonly measured at an individual level, so is readily available for analysis (Koopman et al., 2008). In fact, a comparison of aggregate data and IPD MA approaches to identifying subgroups showed that IPD MA provided a more precise and accurate estimate than the aggregate method (Koopman et al., 2008).

Using traditional RCTs to identify treatment responders and characteristics associated with response is not without caveats. Traditional clinical trials are generally aimed at quantifying the efficacy of a drug and not at identifying treatment responders. Relevant factors that can be used to predict response to therapy are often not measured and participants undergo strict inclusion criteria that ensure homogeneity at baseline, thereby making it very difficult to examine differences between individuals (Schork, 2015).

An approach that may overcome this difficulty is the use of individual patient trials, also known as n-of-1 trials or single patient randomised trials (Lillie et al., 2011, Schork, 2015). N-of-1 trials involve multiple crossovers between treatments, in a random order, within a single individual (Vohra et al., 2015, Guyatt et al., 1986). Multiple n-of-1 trials can be combined and treatment responders can be analysed for shared characteristics (Lillie et al., 2011). Often, n-of-1 trials are more flexible in their inclusion criteria than traditional RCTs and reflect clinical practice more accurately. They may, therefore, provide more information on potential predictors than RCTs or MAs of RCTs. Despite being heralded as a promising way forward for precision medicine, few n-of-1 trials have been implemented for this purpose.

1.4.3 Precision medicine in OA

Identifying subgroups in OA and the treatments they are most responsive to has been considered “one of the promising ways forward in clinical research” (Bierma-Zeinstra and Verhagen, 2011). OA is a variable condition with heterogeneous responses to therapy, but the implementation of precision medicine is dependent on the identification of subgroups with distinct responses to treatment.

The first step is to determine how to subgroup individuals with OA. With a condition as varied as OA, many different subgrouping factors can be used. Bierma-Zeinstra and Verhagen (2011) suggested several subgroups that may be relevant to clinical research (Table 1-8).

Table 1-8 - Suggested main subgroups of OA in clinical research

OA phenotypes	Structural or symptomatic stage	Co-morbidity	Personal factors
Joint site/joint compartment	Pain severity	Obesity	Gender
Localised or generalised OA	Restricted motion	CV disease	Age
Structural OA subtypes	Radiographic severity	Chronic obstructive pulmonary disease	Treatment preference
Pain phenotypes	Effusion/synovitis	Depression	Psychosocial factors
BML			

Suggested subgroups as per (Bierma-Zeinstra and Verhagen 2011)
BMLs, bone marrow lesions; CV, cardiovascular; OA, osteoarthritis

Many of the subgroups described by Bierma-Zeinstra and Verhagen (2011) have since been shown to significantly associate with clinical progression in observational studies (Deveza et al., 2017). Differences in pain sensitisation, psychological distress, radiographic severity, BML, muscle strength, inflammation, and comorbidities (especially metabolic syndrome) have been shown to characterise OA phenotypes with clinically distinct outcomes (Deveza et al., 2017). It may be that these same factors also associate with different treatment responses.

Work on identifying predictors of response to different treatments in OA are mainly limited to small trials with considerable methodological shortcomings, predominantly lack of power (Bruyère et al., 2015). In addition, data dredging may underlie these spurious significant findings, as the analyses are often conducted *post hoc*.

SRs unfortunately do little to further our knowledge in this area. For example, one such review found that baseline OA severity appeared to predict response to treatment, with individuals with less severe structural changes showing a better response to various treatments (Bruyère et al., 2015). Another review found inconclusive evidence for a relationship between measures of sensitisation and treatment response (O'Leary et al., 2017). SRs are unable to adequately identify predictors and a more methodologically robust approach has been developed by the OA Trial Bank. The OA Trial Bank is a collaborative group that aims to bring together IPD for MAs with the purpose of identifying responsive subgroups. Two IPD MAs have been completed by the OA Trial Bank so far. One found that those with severe baseline pain showed a better response to intra-articular glucocorticoids in the short term than those with less baseline pain, although no differences were observed for medium or long term pain outcomes (Van Middelkoop et al., 2016). The other did not find any predictors of response to oral glucosamine, although this may be because the effects of glucosamine were equal to placebo (Runhaar et al., 2017). The present work is conducted in collaboration with the OA Trial Bank and will contribute to the evidence base for predictors of treatment response in OA.

1.4.3.1 Responders to topical NSAIDs and capsaicin in OA

Little is known regarding predictors of response to topical NSAIDs or capsaicin in OA despite being highlighted as an area requiring further research in OA guidelines (Jordan et al., 2003, Zhang et al., 2007, NCCCC, 2008).

One small, uncontrolled study of radiographic knee OA examined for QST-assessed indices that would predict response to 1% diclofenac sodium gel (Edwards et al., 2016). QST measures, including PPTs, TS,

cuff pressure algometry, cold pressor task, and conditioned pain modulation (CPM) were assessed at baseline. The correlation between baseline QST measures and changes in pain scores after four weeks of topical NSAID were assessed. Of these, only baseline CPM scores predicted end-of-treatment pain scores (β adjusted for baseline pain intensity, age, and sex = -0.30, p-value = 0.03). Those with higher CPM at baseline had the greatest pain reduction in response to a topical NSAID. CPM is a measure of the endogenous pain inhibitory pathways which, in healthy individuals, dampen down the pain experienced from a painful stimulus when a second painful (conditioning) stimulus is applied (Lewis et al., 2012). This inhibitory system is often impaired in individuals with chronic pain (Lewis et al., 2012). The results of Edwards et al. (2016) indicate that participants with less impairment may respond better to topical NSAIDs. However, the study was small (n=35) and did not account for multiple testing, thereby questioning the validity of the conclusions. The results should therefore be viewed as exploratory and require validation through other means. In addition, the lack of control group means that CPM may be a prognostic factor for improvement in those treated with topical NSAID, rather than a predictor of treatment response *per se*.

No further evidence is available for predictors of response to topical NSAIDs in OA, however, attempts have been made to identify predictors of response to oral NSAIDs and COX-2 inhibitors. Walker et al. (1997) were unable to identify any predictors of treatment response. DeTora et al. (2001) and Hmamouchi et al. (2012) also failed to identify characteristics that consistently predicted response to oral NSAIDs across multiple response criteria. Characteristics that associated with improvements in some, but not all, outcome criteria were increasing age, higher level of education, absence of joint swelling, greater number of joints involved, no previous joint injury, and knee height (Detora et al., 2001, Hmamouchi et al., 2012). Bingham et al. (2011) found that poor function at baseline was associated with lower odds of responding to COX-2 inhibitors. However, a similar pattern was observed within the placebo group, indicating that poor function may be a predictor for improvement or

placebo response, rather than response to treatment. In addition, the study investigated 16 potential predictors within three different treatment groups using only 1,207 participants thereby significantly increasing the risk of type I error. Moreover, the RCTs used a flare design and it was only the pre-flare baseline function that associated with NSAID response. Over 80% of participants used a coxib or NSAID prior to the study and the results may reflect that those that were already showing a poor response to oral NSAID before the study (poor physical function) were the ones that continued to have a poor response on resuming treatment in the study. No predictors of response were identified using post-flare baseline data and no variable distinguished a response in the active compared to the placebo group. Most recently, a preliminary analysis of 80 participants indicated that higher baseline pain and TS significantly associated with improved responses to oral ibuprofen (Petersen et al., 2017). Interestingly, previous work by the same group also found that higher TS associated with improved response to etoricoxib, but only in *non-responders* (pain reduction <30% or <50%) (Arendt-Nielsen et al., 2016). Inversely, the negative correlation between pain relief and TS may suggest that lower TS (more normal) is associated with less pain relief in non-responders. However, no predictors for *response* to oral NSAID were identified thus bringing into question the relevance of the TS findings. Overall, the evidence for predictors of response to NSAIDs is limited and of low quality. The studies have generally been small (underpowered to identify responders) and tested multiple potential predictors, thus subject to type I and II error. In addition, regardless of the level of evidence, it may not be possible to extrapolate the results of such trials to topical NSAID formulations. Evidence suggests that oral NSAIDs modulate central sensitisation (Arendt-Nielsen et al., 2016), and this may explain the finding that some QST measures may predict response to treatment. As topical NSAIDs have comparatively low systemic exposures, the relevance of centrally-acting NSAID mechanisms to treatment response is likely limited.

There is only very scarce evidence for predictors of response to topical capsaicin in OA. A statistically significant treatment-by-baseline

severity interaction was detected in an RCT of 0.075% versus 0.01% civamide (cis-isomer of capsaicin) in knee OA (Schnitzer et al., 2012). A greater efficacy was observed for the 0.075% cream in participants with higher baseline scores compared to those with lower scores. No further details are available as the finding was mentioned only in the methods, and not formally presented in the results.

There has, however, been more interest in identifying predictors of response to topical capsaicin (predominantly 8% capsaicin patch) in NP. Small uncontrolled trials have identified a subset of characteristics that may predict an improved response to topical capsaicin (Mainka et al., 2016, Serrano et al., 2017, Gustorff et al., 2013), listed in Table 1-9. Evidence from pooled analyses of RCTs further contribute to the list of potential predictors (Katz et al., 2015a, Martini et al., 2013). Each study pooled the IPD from four RCTs, but did not use appropriate multilevel or two-stage MA methods to account for study clustering. Regardless of the methodology, the relevance of these findings to an OA population is questionable. Although both conditions may have substantial contributions from central pain mechanisms, the factors driving the hypersensitivity may be different and therefore require different treatments. In addition, the 8% capsaicin patch differs fundamentally from the lower concentration creams and it may not be possible to extrapolate the findings across the formulations. In fact, the only trial examining for predictors of response to 0.025% capsaicin cream in NP failed to find any characteristics that were significantly associated with response in PHN (Peikert et al., 1991).

Table 1-9 - Predictors of increased response to topical capsaicin in NP

Predictor for improved response	Level of evidence	n
Person-related		
Younger age*	RCTs pooled	1,248
Female sex*	RCTs pooled	2,262
Disease-related		
Lower baseline pain scores*	RCTs pooled	1,248
Shorter disease duration	Uncontrolled study & RCTs pooled	1,268
Shorter history of paraesthesia	Uncontrolled study	20
Not using opioids	RCTs pooled	1,248
Increased variability in pain scores pre-treatment*	RCTs pooled	1,248
Absence of allodynia (examination and self-reported)	RCTs pooled	1,014
Hypoesthesia	RCTs pooled	1,014
Lower McGill Pain Questionnaire score	RCTs pooled	1,014
QST-related		
Lower cold pain threshold	Uncontrolled study	20
Lower mechanical pain threshold	Uncontrolled study	20
Homogeneity in warm sensation threshold and heat pain threshold between affected and non-affected limb	Retrospective, uncontrolled analysis	55
Brush and cotton wool allodynia	Uncontrolled study	57

*Predictor for 0.04% and 8% capsaicin patch. All others are for 8% patch only (Mainka et al., 2016, Martini et al., 2013, Serrano et al., 2017, Katz et al., 2015a, Gustorff et al., 2013)

n, number of participants analysed; NP, neuropathic pain; QST, quantitative sensory testing; RCT, randomised controlled trial

1.5 Justifications for thesis

Treatment guidelines have emphasised the need to identify predictors of response and tailor the treatment of OA to the individual (NCCCC, 2008, Jordan et al., 2003, Zhang et al., 2007). However, little is known about the predictors of response to topical NSAIDs or capsaicin in OA.

Furthermore, it is established that a subset of individuals with OA may exhibit neuropathic-like pain symptoms whereas others may present symptoms that are inflammatory in nature. However, whether these patient subgroups respond differently to topical NSAIDs and capsaicin, which have different mechanisms of action, remains unknown.

The aims and objectives of this thesis are:

- **Aim 1:** To determine the relative efficacy of topical NSAIDs and capsaicin in OA.
 - **Objective 1A:** To estimate the ES of topical NSAIDs and capsaicin over placebo (conventional MA (CMA))
 - **Objective 1B:** To compare the efficacy of topical NSAIDs to capsaicin (NMA)
 - **Objective 1C:** To compare the absolute change from baseline scores (overall treatment effect) between topical NSAIDs and capsaicin
- **Aim 2:** To identify potential predictors of the specific and overall treatment effects of topical NSAIDs and capsaicin in OA
 - **Objective 2A:** To examine existing RCT data for predictors of response (IPD MA)
 - **Objective 2B:** To determine if there are differences in the response to topical NSAIDs and capsaicin within the individual (n-of-1)

- **Objective 2C:** To explore potential participant-level characteristics that associate with an improved response to topical NSAIDs or capsaicin in OA (n-of-1)
- **Objective 2D:** To evaluate n-of-1 trials for the purpose of identifying treatment responders in OA
- **Aim 3:** To determine the relative efficacy of topical NSAIDs and capsaicin in individuals with NP, in order to examine the possible effect of these two agents in neuropathic-like pain in OA
 - **Objective 3A:** To estimate the ES of topical NSAIDs and capsaicin over placebo (CMA)
 - **Objective 3B:** To compare the efficacy of topical NSAIDs to capsaicin (NMA)
 - **Objective 3C:** To compare the absolute change from baseline scores (overall treatment effect) of topical NSAIDs and capsaicin
 - **Objective 3D:** To examine for predictors of response to topical NSAIDs and capsaicin (IPD MA)

2 CHAPTER 2: General methods

The following chapter provides an overview of the methods used for this thesis. It initially describes the methods that are common to the CMA, NMA, and IPD MAs. These include the methods for identifying studies and extracting data from published reports. The chapter then introduces the methods utilised for the CMA, NMA, IPD MA, and pilot n-of-1 trial series. Detailed methods are then provided within each results chapter.

Results for all projects were described as per the relevant PRISMA guidelines or CONSORT statement (Moher et al., 2009, Hutton et al., 2015, Stewart et al., 2015, Vohra et al., 2015).

Unless otherwise stated, all analyses were conducted using Stata software (*Stata Statistical Software: Release 15*. College Station, TX: StataCorp LP).

2.1 General meta-analysis methods

The identification of studies, through systematic literature searches and application of the inclusion and exclusion criteria, are common to all MAs conducted. These methods will be described below. As the nature of the data and the statistical analysis methods used vary between the MA types these are described separately within each results chapter.

2.1.1 Identification of studies

RCTs of topical NSAIDs or capsaicin in OA were identified using systematic literature searches, hand searches, and direct enquiries to pharmaceutical companies regarding unpublished data. All identified studies, regardless of the comparator and publication status, were eligible.

2.1.1.1 Systematic literature search

2.1.1.1.1 Search terms and strategies

Two search strategies were developed to identify studies investigating topical NSAIDs or capsaicin in OA.

The search strategies were created using four main search terms combined using “and”: (1) terms for RCTs; (2) terms for topical administration; (3) terms for OA; and (4) terms for the treatment (NSAID or capsaicin). MeSH headers were used where possible and were combined with specific terms using truncations to capture the full scope of the field.

The search strategies were based on those from the Cochrane reviews of topical NSAIDs (Derry S, 2012) and topical capsaicin (Derry S, 2013, Derry and Moore, 2012). They were expanded upon through the addition of terms found by searching the internet. The search strategies used can be found in Appendix 1 – Systematic literature search strategies.

2.1.1.1.2 Databases

Table 2-1 lists the databases searched. The search strategy was conducted up until 16/11/2015.

Table 2-1 - List of databases used for systematic literature search

Database	Interface	Dates
Medline	OVID	1946 onwards
Embase	OVID	1974 onwards
Allied and Complementary Medicine Database (AMED)	OVID	1985 onwards
Cumulative Index to Nursing and Allied Health Literature (CINAHL)	EBSCOhost	All years
Web of Science	Core Collection	1900 onwards
Cochrane Library	Wiley	Current issue

2.1.1.1.3 Analysing the literature search output

The results of the literature search were exported from each database to EndNote X7.7.1. Duplicates were identified and removed. The titles of the remaining publications were read and the abstracts for those that were potentially eligible were sought. The abstracts were screened for eligibility and the full texts were identified only for those that potentially met the eligibility criteria. The full texts were then scrutinised for eligibility. The reasons for excluding studies at the full text stage were recorded.

2.1.1.2 Identification of other studies

Various approaches were taken to identify eligible studies not captured by the systematic literature search. These involved searching the reference lists of included publications, reviews in the area, and hard copies of potentially relevant trials kept in the department. Reviews in the area were identified through literature searches in Medline and Cochrane Library using MeSH headers for “osteoarthritis” and “administration, topical”. The results were then limited to “meta-analysis or systematic review”.

Pharmaceutical suppliers of topical NSAIDs and topical capsaicin, identified via the British National Formulary (BNF), the electronic Medicines Compendium (eMC), and Clinicaltrials.gov, were contacted to ask for any unpublished trial data. Finally, collaborating authors were asked if they were aware of any eligible studies not yet included in the MA.

2.1.2 Inclusion and exclusion criteria

Table 2-2 describes the inclusion and exclusion criteria that were applied to the studies to determine eligibility. The criteria were based on five domains, specifically the types of: (1) studies; (2) participants; (3) interventions used; (4) comparators used; and (5) outcome assessments.

Table 2-2 - Inclusion and exclusion criteria used for MAs of topical NSAIDs and capsaicin in OA

Domain	Inclusion criteria	Exclusion criteria
Study design	RCT Minimum 1 week duration No language or geographic restrictions Open-label and blinded studies	Trials with the same drug and concentration in both arms
Participants	Clinically or radiographically diagnosed OA Chronic joint pain, not due to inflammatory or other arthropathies	Back or cervical pain Other causes of arthritis (e.g. rheumatoid, psoriatic)
Interventions	Any topical NSAID or capsaicin Any formulation, dosage, or drug within group	Salicylates Interventions using US or magnetophoresis for application
Comparators	Controls: placebo, active comparators, continuing usual treatment, no treatment	
Outcomes	Adequate assessment of pain	

MA, meta-analysis; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis; US, ultrasound

2.1.2.1 Study design

Published and unpublished RCTs that evaluated the efficacy of either topical NSAIDs or topical capsaicin in participants with OA were included. Both open label and blinded studies were included. Duration of treatment was a minimum of one week. No language or geographical restrictions were applied to the studies.

2.1.2.2 Participants

Men and women with clinically or radiographically diagnosed OA at any site (excluding the spine) were included. Spinal pain was excluded as it is often difficult to differentiate osteoarthritic back pain from back pain secondary to other aetiologies. Participants with chronic joint pain, not

attributable to other chronic pain conditions such as RA or fibromyalgia, were also included. OA is known by many names, and participants with conditions such as “activated gonarthrosis” and “Heberden’s and Bouchard’s nodes” were included under the assumption that these terms were referring to OA. Studies that included only a subgroup of eligible participants were included as long as data could be collected separately for the eligible participants.

2.1.2.3 Interventions

Topical NSAIDs and topical capsaicin, in any formulation, dosage, drug, or salt, were eligible. Salicylates, although pharmacologically related to NSAIDs, were excluded as their principal action when applied topically is different from that of topical NSAIDs (Derry et al., 2014).

Interventions involving US or magnetophoresis of the topical treatment were excluded.

2.1.2.4 Comparators

Eligible controls included placebo, active comparators, continuing usual treatment, no treatment, and head-to-head comparisons between topical NSAIDs and capsaicin. Studies comparing the same drug at different concentrations were included, whilst studies using the same drug and concentration in both treatment arms were excluded.

2.1.2.5 Outcomes

Trials were required to have an adequate assessment of pain for inclusion. Pain at or nearest to four weeks was the primary outcome. Pain at other durations of treatment was a secondary outcome.

2.1.3 Data collection

The following section describes the methods for data collection and preparation of the data from published reports. It initially describes the methods for data extraction, including the risk of bias assessment, and data validation. It then describes the process of managing the data, as well as the checks and calculations carried out to prepare the data for

analysis. Finally, it describes the data extracted from the database for analysis.

The data from published reports were used for the analysis of the CMAs and NMAs and to establish the completeness of the IPD MA. Descriptions of the acquisition of the data for analysis in the IPD MA are described in section 4.2.1.

2.1.3.1 Data extraction

Study-level variables were extracted from the published reports using predefined data extraction forms. Section 2.1.3.1.1 provides a description of the data extraction form and its design, whilst section 2.1.3.1.2 describes the outcomes extracted.

2.1.3.1.1 Data extraction form

The data extraction form was developed in Microsoft Word. It was designed to capture all potential study-level variables, including general trial details, participant details, intervention details, quality assessment, and outcome details. As the studies varied in various aspects, including the number of treatment arms and the number of outcomes, the form allowed the addition or deletion of columns as required. In order to standardise the data extraction procedure, detailed descriptions of how to extract data using the form were provided. The form and instructions for its use are presented in Appendix 2 – Data extraction form and instructions.

2.1.3.1.2 Outcomes extracted

All clinically relevant outcome data, at baseline and at all durations of follow-up (≥ 1 week), were extracted. In order to examine the effect at different *treatment* durations (rather than durations of *follow-up*) any outcomes presented after cessation of the treatment were not extracted.

Intention-to-treat (ITT) data were extracted where possible. If only per protocol (PP) data were presented, they were extracted instead. Data were not extracted from figures unless the numerical value was

presented. Partially presented data, such as mean change with no associated variance, were extracted. Where a study included non-eligible participants, such as those with RA, demographic and outcome data were only extracted for eligible participants. If the data for eligible and non-eligible participants were combined, they were not extracted.

For cross-over trials, the first phase was extracted and used for analysis. If this was not presented, the combined data from the first and second phases were extracted.

2.1.3.1.3 Non-English published reports

Data from non-English publications were extracted using various methods depending on the publication language. Where the publication was in a language spoken by a member of the review team, the data were extracted by that member. If no individuals in the department spoke the language, efforts were made to locate native speakers by contacting the School of Medicine, the International Office, and personal contacts. Where no native speakers volunteered to extract the data, the extraction was carried out using Google Translate (online version and smartphone app that translates photographs of text).

2.1.3.2 Data validation

A shortened version of the data extraction form was created for data validation purposes. A member of the review team (Dr Joanne Stocks, JS) independently extracted data for all English-language published reports using the validation form. For convenience, only the primary outcome data were extracted.

The validation forms were compared to the data previously extracted. Where any discrepancies were found, the publication was reviewed. Where an error had been made in the extraction, this was amended accordingly. Where the differences arose from different interpretations of the report, this was discussed with JS. The guidelines for data extraction were referred to in order to make a decision, and if a

consensus could still not be reached, it was discussed with a third reviewer (WZ).

2.1.3.3 Study quality assessment

The study quality assessment was carried out during data extraction using a modified version of the Risk of Bias tool recommended by the Cochrane Collaboration (Higgins JPT, 2011) (Table 2-3 and Appendix 2 – Data extraction form and instructions). It is comprised of nine questions that assess each of the domains set out in the Risk of Bias tool developed by Cochrane.

Table 2-3 - Modified Risk of Bias assessment

Source of bias	Yes	No	Unclear	Comments
1. Was the randomisation procedure adequate?	☐	☐	☐	
2. Was the treatment allocation adequately concealed?	☐	☐	☐	
3. Were participants adequately blinded to the intervention?	☐	☐	☐	
4. Were physicians adequately blinded to the intervention?	☐	☐	☐	
5. Were outcome assessors adequately blinded to the intervention?	☐	☐	☐	
6. Incomplete outcome data: Is the attrition rate <15%?	☐	☐	☐	
7. Are all pre-specified outcomes of interest reported in the pre-specified way?	☐	☐	☐	
8. Was intention-to-treat analysis used?	☐	☐	☐	
9. Were the treatment and control group similar at baseline?	☐	☐	☐	

Modified from Cochrane's Risk of Bias tool (Higgins JPT, 2011). See Appendix 2 – Data extraction form and instructions for criteria for a judgment of “yes”, “no”, and “unclear”

With the exception of non-English publications, the quality assessments were also undertaken during data validation by JS. Where the study was in a language not spoken by both authors, the risk of bias assessment was done only by the author or colleague that spoke the language.

The risk of bias for each domain was presented for all included studies. Furthermore, the risk of bias score for allocation concealment was used for sensitivity analyses.

2.1.3.3.1 Modified risk of bias tool

In contrast with the narrative nature of the Cochrane tool, the risk of bias tool was modified to use questions that were scored as yes, no, or unclear. Each question evaluates the risk of bias in an objective manner so as to minimise the inter-rater disagreement that is associated with the more subjective domains in the Cochrane Risk of Bias tool (Hartling et al., 2009). The modifications were aimed at minimising variability in the quality scoring as it was planned that assessments of non-English publications would be conducted by native speakers (i.e., the same person would not be able to assess all included trials).

The first five questions of the modified tool assess selection bias, performance bias, and detection bias. Each of these questions were scored using the criteria outlined by Cochrane (Higgins JPT, 2011). The sixth question is a measure of attrition bias, and similarly to the PEDro and van Tulder quality scales (Olivo et al., 2008), a cut-off for acceptable drop-out rates was used to determine the risk of this bias. After discussion between review authors, a consensus of 15% cut-off for attrition bias was agreed upon (PEDro, 1999). The seventh question addresses selective outcome reporting by comparing the outcomes pre-specified in the methods with those presented in the results. Within the domain “Other potential threats to validity”, two questions cover the use of intention-to-treat analysis and whether the treatment and control group were similar at baseline. These two sources of bias are assessed in a variety of RCT quality scales (Olivo et al., 2008).

2.1.3.4 Data management

2.1.3.4.1 Microsoft Access database

A Microsoft Access database was created for this project. The database was created in a long format to allow the addition of a varying number of outcomes and time points per study.

Several tables, each corresponding to the data extraction domains, were used. A study ID was given to each included study, and this was used to link the tables together. Forms were created for data entry and queries were run to extract data for analysis.

2.1.3.4.2 Data entry, checks, and calculations

Data were entered from the data extraction forms into the Access database. All study data were then checked for missing values. If any variables were missing, the publication was checked to determine whether the information was available. Any missed variables were entered. Any calculations required, such as calculating change from baseline scores using endpoint and baseline scores or combining treatment arms, were done at this point using Excel. The formulae used for these analyses are presented in Table 2-4. Any calculated variables were marked as such in the database. Any missing variables that were not presented in the publication and could not be calculated were noted.

Table 2-4 - Formulae used to prepare data for analysis

Calculation and formula

To calculate mean from median (Hozo et al., 2005)

$mean = median$, where $n > 25$

$mean = \frac{minimum + 2median + maximum}{4}$, where $n < 25$

To calculate SD from standard error of mean (SE) (Higgins JPT, 2011)

$$SD = SE \times \sqrt{n}$$

To calculate SD from interquartile range (IQR) (Wan et al., 2014)

$$SD = \frac{(q3 - q1)}{\eta(n)}$$

where $\eta(n)$ is derived from a table in the reference paper, using "n"

To calculate SD from range (Hozo et al., 2005)

$SD = \frac{range}{4}$, where $n = 15-70$

$SD = \frac{range}{6}$, where $n > 70$

To calculate SD from 95% CI (Higgins JPT, 2011)

$SD = \sqrt{n} \times \frac{upperlimit - lower limit}{3.92}$, where $n > 60$

$SD = \sqrt{n} \times \frac{upperlimit - lower limit}{2t}$, where $n < 60$

To combine two groups (Higgins JPT, 2011)

	Group 1	Group 2	Combined groups
n	n_1	n_2	$n_1 + n_2$
mean	x_1	x_2	$\frac{n_1x_1 + n_2x_2}{n_1 + n_2}$
SD	SD_1	SD_2	$\sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2 + \frac{n_1n_2}{n_1 + n_2}(x_1^2 + x_2^2 - 2x_1x_2)}{n_1 + n_2 - 1}}$

To calculate change scores using baseline and endpoint scores (Borenstein et al., 2011)

Mean $Mean_{change} = mean_{endpoint} - mean_{baseline}$

SD $SD_{change} = \sqrt{SD_{endpoint}^2 + SD_{baseline}^2 - 2 \times r \times SD_{endpoint} \times SD_{baseline}}$

Assuming $r=0.5$ (Fu et al., 2013), this simplifies to:

$$SD_{change} = \sqrt{SD_{endpoint}^2 + SD_{baseline}^2 - SD_{endpoint} \times SD_{baseline}}$$

IQR, interquartile range; n, number of participants; r, correlation between baseline and endpoint scores; SD, standard deviation; t, t distribution with degrees of freedom equal to the group sample size minus 1

2.1.3.5 Data for analysis

Data were extracted from the Access database into Excel using queries. Each query linked the outcome table with the study level information required for sensitivity analyses or meta-regression, such as the allocation concealment status, the drug type, and the overall study size.

The change from baseline scores (as published or calculated) were extracted for analysis. Where it was not possible to calculate the change from baseline score, the endpoint score was used. This was done under the assumption that the baseline pain scores were homogeneous across the treatment groups secondary to the randomisation process. Where the outcome scale differed in the direction of severity, the studies using a different direction were multiplied by -1 (Higgins JPT, 2011).

Within Excel, the data were browsed to ensure that the correct data were captured in the query. The outcome highest on the hierarchy, as detailed in the following section (2.1.3.5.1) was used for analysis

2.1.3.5.1 Outcome hierarchy for analysis

All MAs in this thesis used pain as the outcome of interest. However, many studies present more than one measure for pain and ordering the instruments prior to extraction may minimise bias by preventing the most favourable outcome from always being extracted (Juhl et al., 2012). The hierarchical list of outcomes is presented in Table 2-5 and the justifications for this list are outlined below.

Prioritising outcomes with the highest responsiveness, and downgrading single measures that lacked evidence for their reliability and validity, Juhl et al. (2012) deemed the WOMAC pain subset to be the first-choice pain outcome measure for use in knee OA MAs. This is in contrast with the hierarchy implemented by Jüni et al. (2006) and McAlindon et al. (2000) that listed the global pain score in index joint and pain on walking in index joint higher than the WOMAC pain subset. This hierarchy was developed from the Altman et al. (1996)

recommendations for clinical trials in OA, however, neither paper provided a rationale for the hierarchical ordering of the pain outcomes.

Table 2-5 – Pain outcome hierarchy list used for analysis

Pain outcome
1. VAS global pain score
2. Categorical global pain score
3. Pain during activity, for lower limb this is likely to be walking
4. WOMAC pain subscale or pain subscale of other disease-specific composite tools (e.g., KOOS, AUSCAN, Lequesne)
5. SF-36 bodily pain subscale
6. Health Assessment Questionnaire (pain subscale), McGill Pain Questionnaire (pain intensity)
7. Tenderness
8. Physician's assessment of pain

Modified from hierarchy of outcomes recommended for MAs of OA (Jüni et al., 2006, Juhl et al., 2012, McAlindon et al., 2000)
AUSCAN, Australian/Canadian Hand Osteoarthritis Index; KOOS, Knee Injury and Osteoarthritis Outcome Score; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index

When choosing the hierarchical ordering of pain outcomes, it was felt that prioritising outcomes that were used in both OA and NP would facilitate comparison between the conditions (Chapter 6). Ideally, tools that assessed the multidimensional experience of pain would be used, however, these are not commonly used to measure pain in RCTs. To ensure that the pain scales used for analysis were as comparable as possible, a hierarchy was determined that prioritised the most commonly used tools and was based on the recommendations by Juhl et al. (2012), Jüni et al. (2006), and McAlindon et al. (2000). VAS was prioritised as it is commonly used in OA, but is also the recommended measure for assessing treatment efficacy in NP (Cruccu et al., 2004).

The outcome highest in the hierarchical list was chosen for analysis. Pain severity, measured using VAS or a categorical scale, was prioritised. Following from this, disease-specific composite tools, such as WOMAC were chosen. If the chosen outcome was not presented fully (missing the number of participants (n), mean, or standard deviation (SD)), the missing data were calculated from the data available using

the formulae described in Table 2-4. When it was not possible to calculate the missing data, the next highest fully presented outcome in the hierarchy was used. If no outcomes were fully presented, the study was excluded from the MA of published reports.

2.2 Meta-analyses of published reports

2.2.1 Conventional meta-analysis

2.2.1.1 Concept

The term “meta-analysis” was coined by Gene Glass in 1976 who described MA as:

“...the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussion of research studies...” (Glass, 1976)

This early definition highlights one of the main overriding principles of MAs which is the synthesis of statistical summary using all similar, but separate, studies. Good quality MAs are based not only on the statistical aggregation of study results but also on the systematic search of the literature.

A systematic search of the literature involves identifying all relevant studies that answer the research question of interest and critically appraising the quality of each study (Higgins JPT, 2011). For this, a search strategy needs to be developed and applied to databases as well as searching the reference list of relevant publications (Higgins JPT, 2011). Once the search has been conducted, pre-specified criteria are used to determine which of the identified studies are eligible for inclusion. Transparency is essential throughout the process and this is facilitated by presenting a PRISMA flow diagram (Figure 2-1) of the number of studies identified at each stage (Moher et al., 2009).

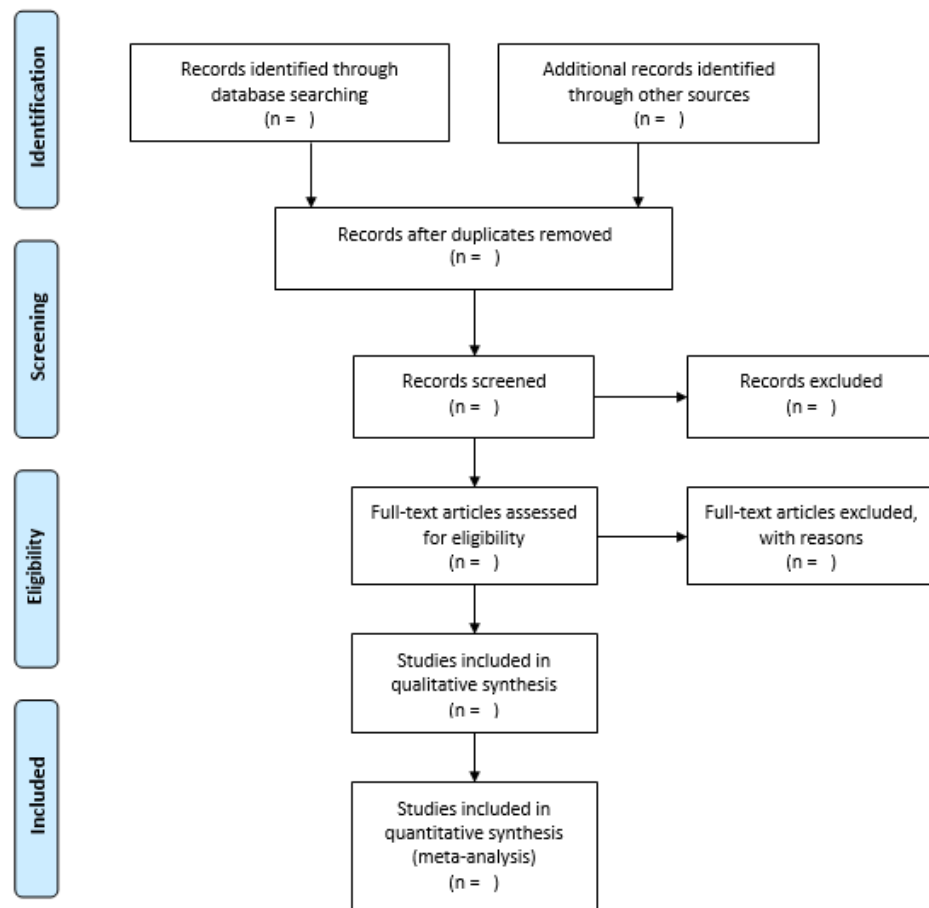


Figure 2-1 - PRISMA 2009 Flow Diagram

Image reproduced with permission. Reference: (Moher et al., 2009)

What differentiates MAs from narrative SRs is the quantitative nature of the summary results. Statistical aggregation of results from individual studies provides a more powerful and precise estimate of the treatment effect as it has greater power than any individual trial involved (Higgins JPT, 2011). A summary effect is calculated by combining the findings from individual studies using the appropriate statistical method and weighting the estimate from each study according to the precision of the estimate (Higgins JPT, 2011). This summary effect (i.e., ES) indicates the magnitude and direction of the association between the treatment and outcome.

The summary results from a MA are most commonly represented using a forest plot. The plot presents the ES of all included studies and the summary finding, i.e., pooled ES. Each estimate is represented by a

point estimate and a horizontal line that indicates the 95% CI of the estimate.

2.2.1.2 Application

A random-effects MA, pooled using inverse variance methods, of pain relief at or nearest to four weeks was conducted for the specific and overall effects of topical NSAIDs and capsaicin (primary). Further analyses were conducted, including cumulative MAs, examination of small study effects/publication bias, sensitivity analyses by trial quality, subgroup analyses by drug/dose, and examination of the specific and overall treatment effects at all durations of follow-up.

2.2.2 Network meta-analysis

2.2.2.1 Concept

A NMA is a natural extension of the CMA that allows comparisons to be made between two treatments using direct and indirect evidence (Higgins and Welton, 2015). The term NMA was first used in 2002 by Lumley (2002) and aptly describes the process of using a “network” of treatment comparisons to indirectly compare treatments (Higgins and Welton, 2015). Where two drugs have not been directly compared in an RCT, a network based on a common comparator (e.g., placebo) can be used to estimate the difference between the two drugs. This method is vital in estimating the comparative effects of drugs that have not been assessed in head-to-head trials, thereby reducing the need for expensive RCTs for all treatment comparisons (Mills et al., 2013).

NMAs allow indirect comparisons to be made between interventions whilst maintaining the randomisation process within each trial (Jansen et al., 2008). This maintains the same relative contribution of contextual effects and balances out differences in baseline risks (Jansen et al., 2008).

For example, if trials exist for the comparison of drug A versus placebo (AP) and for drug B versus placebo (BP) (Figure 2-2), then assuming that the AP and BP trials are alike (i.e., are comparable with regards to effect modifiers), the relative efficacy of the two drugs are mathematically related (Jansen et al., 2008). The estimates are assumed to be internally consistent and the difference between AP and BP therefore allows an estimation of the effect of BA (Jansen et al., 2008). A crucial assumption is that the population across the trial network is similar and that participants in the AP comparison could have partaken in the BP comparison (transitivity).

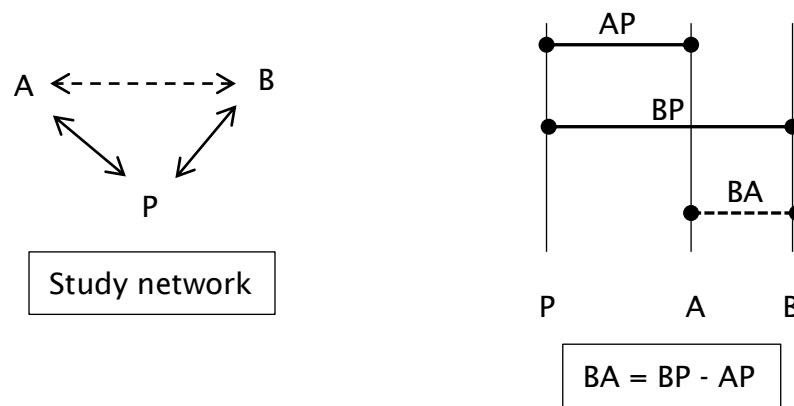


Figure 2-2 - Representation of a network meta-analysis

Direct comparisons: AP and BP. Indirect comparison: BA. Adapted from (Jansen et al., 2008).

NMAs can be conducted using frequentist or Bayesian statistics; the latter of which is the preferred method (Hong et al., 2013). Bayesian methods allow more intuitive interpretation of the results, which are often presented using probabilities and rankings (Hong et al., 2013, Jansen et al., 2008). For example, Bayesian NMAs generate the probability of each intervention being best, second best, etc. which facilitates decision making.

Bayesian methods consider the parameter of interest as a random quantity with a probability distribution which is defined by our prior belief (Sutton and Abrams, 2001). The data collected in the study summarise the likelihood function. The likelihood function and prior information together form the posterior distribution function (Sutton and Abrams, 2001). In order to avoid influencing the results, which is often quoted as a critique for the Bayesian approach, a non-informative prior is recommended (Jansen et al., 2008). In simple terms, Bayesian statistics uses the newly acquired evidence to update previous beliefs about the parameter of interest in order to reach a new estimate (posterior distribution).

Although NMAs provide vital evidence where head-to-head comparisons do not exist, they are not without caveats. Effect estimates obtained are influenced by factors such as the number of trials in the network, the number of trials using more than two comparison arms, heterogeneity, inconsistency, and bias (Li et al.,

2011). Furthermore, the use of Bayesian statistics is computationally complex and does not result in an automatic measure of statistical significance, such as $p < 0.05$.

However, steps can be taken to attempt to minimise or quantify the effect of these factors, such as ensuring that a systematic search of the literature is conducted to ensure all available data are captured, assessing the risk of bias of included studies, and quantifying the inconsistency of the results.

2.2.2.2 Application

A NMA was conducted to compare topical NSAIDs to capsaicin using placebo as a common comparator. Bayesian and frequentist approaches were used to determine the comparative efficacy of these drugs using indirect evidence.

2.3 Individual patient data meta-analysis

2.3.1 Concept

As with other MA methods, IPD MAs aim to answer a research question by combining all sources of evidence that address the research question. However, unlike other approaches, IPD MAs require original trial authors to share the data on all individual participants in the included trials (Higgins JPT, 2011). The raw data is re-analysed centrally whilst preserving the clustering of participants within their original trials (Riley RD, 2010).

Analyses can either be done using a one-stage or two-stage approach. A one-stage approach involves modelling all data simultaneously whilst maintaining the participant clustering within trials (Riley RD, 2010). A hierarchical or multilevel model is used (Debray et al., 2015). For continuous outcomes, the models are often extensions of analysis of covariance (ANCOVA) models (Burke et al., 2017). Although the one-stage approach is more flexible (Debray et al., 2015), it is often criticised for being computationally intensive, prone to convergence issues, and harder to learn (Burke et al., 2017). Furthermore, it requires the within-study and between-study variability to be carefully separated to avoid ecological bias (Riley RD, 2010).

The two-stage approach involves two steps. Firstly, the IPD from each study is analysed separately using appropriate models in order to obtain a study-specific estimate of interest (Debray et al., 2015, Burke et al., 2017). These estimates are then pooled together in the second step using accepted MA methods (Burke et al., 2017, Debray et al., 2015). The two-step approach is the most commonly used method to date (Tierney et al., 2015). It is felt to be conceptually more intuitive, it uses well documented MA methods in the second step, and its study-specific estimates are not influenced by external information (Riley RD, 2010, Debray et al., 2015). However, the two-stage approach has limited power for detecting treatment-covariate interactions and non-linear associations (Debray et al., 2015). Problems also arise when few studies or participants are analysed with the two-stage approach, which has been found to be more prone to bias in pooled effects,

standard errors, heterogeneity, and correlations in these cases (Debray et al., 2015, Tierney et al., 2015).

Although there are currently no recommendations for which IPD MA approach is preferred, the literature favours the one-stage approach for its flexibility and power. Both approaches are generally believed to produce similar results (Debray et al., 2015, Burke et al., 2017). However, this is not always the case and it is important to pre-specify the method prior to analysis or to present the results from both approaches (Tierney et al., 2015).

IPD MAs were traditionally viewed as being substantially more-resource intense than CMAs, but were also thought to be the “gold standard” of SRs. Current literature continues to name IPD MAs as the “gold standard”, however, this is an overly simplistic view of IPD MAs. Although IPD MAs have substantial benefits over CMAs, they are not without disadvantages.

The benefits of IPD MAs over CMAs are related to the data collected and the analyses that are possible. One of the difficulties associated with conducting a CMA arises from extracting and combining aggregate data that are often missing, poorly presented, measured using different scales, or derived differently across the studies (Riley RD, 2010, Stewart LA, 2002). IPD MAs allow inclusion of these studies, regardless of the quality of the reporting. Furthermore, IPD MAs are able to capture a larger proportion of relevant data as they allow the use of unpublished data, the analysis of participants that may have been excluded from analysis in the published report, and the use of outcomes that may have been assessed but not reported (Stewart LA, 2002, Riley RD, 2010).

Perhaps the most important benefit of IPD MAs is that they are the most powerful MA method for detecting predictors of response (Bierma-Zeinsträ and Verhagen, 2011, Riley RD, 2010, Stewart LA, 2002). Rather than extracting data for subgroups from published trials, which are often not presented, IPD MAs allow a more precise estimate of the subgroup effect (Koopman et al., 2008). This examination reduces ecological fallacy which hinders subgroup

analysis in CMAs (Riley RD, 2010). Furthermore, IPD MAs allow examination of the effect modification of multiple characteristics together (Riley RD, 2010).

In reality, IPD MAs are associated with various disadvantages. IPD MAs are often viewed as being more resource intense than CMAs. Although this is often the case, the difference is becoming less noticeable as technology improves (Stewart LA, 2002). With the increasing use of e-mail and electronic databases, the process of data collection is being facilitated (Stewart LA, 2002). This process is expected to be further facilitated by a proposal from the International Committee of Medical Journal Editors to demand that all trialists make plans for sharing IPD as this will be a prerequisite for having studies published in journals (Taichman et al., 2016). Furthermore, websites such as clinicalstudydatarequest.com have been created to allow for requests to be made directly to pharmaceutical companies for access to the IPD from their studies. Often IPD MAs are conducted as large, international, collaborative projects in order to pool resources and overcome these difficulties (Stewart LA, 2002, Riley RD, 2010). There has been a steep increase in the number of IPD MAs conducted over the last two decades (Riley RD, 2010) and as the logistics for data sharing improve this is likely to continue to rise in the future.

IPD MAs are reliant on collecting original trial data from authors and it is often difficult to find the contact details for the authors and convince them to collaborate. Studies that are not included in the MA, whether it is because the data have been destroyed or lost, or the author is not contactable or not willing to collaborate, may introduce bias to the MA, particularly if they are systematically different from those included (Riley RD, 2010).

2.3.2 Application

IPD MAs were conducted to identify participant- and study-level predictors of response to topical NSAIDs and capsaicin. One-stage (primary) and two-stage (secondary) IPD MA approaches were used to examine for predictors of the specific and overall treatment effects.

2.4 Pilot n-of-1 trial series

2.4.1 Concept

N-of-1 trials, or individual patient trials, are within-individual, randomised, multiple cross-over trials (Kravitz and Duan, 2014). The idea of single-participant trials has been present in the literature since the 1950's (Hogben and Sim, 1953), but was only formalised and applied to medicine as an n-of-1 trial in 1986 (Guyatt et al., 1986). Although largely forgotten thereafter, n-of-1 trials have been resurrected in recent years as they meet the demand for precision medicine (Schork, 2015). They border between clinical care and clinical research and provide a rigorous approach for eliciting patient preferences, essential for the application of precision medicine (Vohra et al., 2015). In fact, they are viewed by some as the pinnacle for decision making for the individual as they are the epitome of evidence based medicine (Vohra et al., 2015).

The use of n-of-1 trials is limited to conditions that are chronic, symptomatic, and stable or only slowly progressive (Kravitz and Duan, 2014). The treatments examined shall, if effective, be associated with a rapid improvement on their commencement and return to pre-treatment state on their withdrawal (Guyatt et al., 1986). As such, n-of-1 trials are theoretically appropriate for the investigation of non-disease modifying analgesics in OA (March et al., 1994, Wegman et al., 2003, Yelland et al., 2007). Furthermore, n-of-1 trials have been found to be acceptable to people with OA (Brookes et al., 2007).

A single participant is given the treatments of interest (or placebo) in a randomly determined order. Randomisation serves to balance the effect of time-dependent confounders on treatment estimates and is most often done using paired randomisation (Kravitz and Duan, 2014, Gabler et al., 2011). The interventions are usually blinded, although it is not critical (Gabler et al., 2011, Kravitz and Duan, 2014). The participant is assessed at the beginning and end of each treatment period and their response to the treatment is determined. The course is repeated multiple times to confirm the results and minimise the effect of random error and the confounding effect of lifestyle

modifications (Kravitz and Duan, 2014, Lillie et al., 2011). The participant is at the heart of the n-of-1 trial and the design and conduct of the trial is often viewed as a collaboration between the physician/researcher and the participant. As such, the design of n-of-1 trials may be individualised to the participant (Lillie et al., 2011, Vohra et al., 2015).

Hogben and Sim (1953) eloquently argued for the case of single participant trials to answer the question of why “some individuals respond or fail to respond to one or other treatment”. In order to answer this question, n-of-1 trials must be expanded to n-of-1 trial *series* by the aggregation of multiple n-of-1 trials of the same design. Series of n-of-1 trials may be used to examine shared characteristics between individuals that do or do not respond to a treatment (Lillie et al., 2011, Schork, 2015). Although heralded as a promising way forward for precision medicine, few n-of-1 trial series have been implemented thus far for this purpose.

2.4.2 Application

Few potential predictors of response have been measured in traditional RCTs and the analyses are therefore mainly limited to age, gender, BMI, and baseline pain. A pilot n-of-1 trial series was therefore conducted in order to determine 1) if differences exist in the response to topical NSAIDs and capsaicin in OA; 2) if these differences in response can be explained by baseline characteristics (focusing on neuropathic-like pain and inflammation); and 3) whether n-of-1 trial series can be used for this purpose.

3 CHAPTER 3: Study-level efficacy and relative efficacy of topical NSAIDs and capsaicin in OA

3.1 Overview

The following chapter describes the efficacy and relative efficacy of topical NSAIDs and capsaicin in OA. Study-level data used were extracted from published reports of placebo-controlled trials in OA. CMAs were conducted to examine the specific effects of topical NSAIDs and capsaicin versus placebo. In the absence of direct evidence, a NMA was conducted to compare topical NSAIDs to capsaicin indirectly using placebo as a common comparator. Finally, the overall treatment effect and proportion of contextual effects (PCE) of topical NSAIDs and capsaicin in OA were examined using CMA methods.

3.2 Methods

Eligible studies were identified and data extracted as per section 2.1. Data extracted from the published reports were used for the CMAs and NMAs. The statistical analysis methods relevant for each of the MAs are detailed in the following sections.

3.2.1 Conventional meta-analysis

Random effects CMAs were conducted by pooling studies weighted by the inverse of the sampling variance. A random-effects model was chosen to account for the considerable variation between studies. The variation in studies was evident across many domains, including the drug class used, the concentration and formulation used, and the populations examined.

The SMD, or ES, was calculated for the continuous outcome pain using Hedges unbiased approach (Hedges, 1982). SMD was chosen above the raw (unstandardised) MD as pain was presented using different continuous scales across the included trials. Outcomes not presented as continuous data were converted to Hedges' ES prior to pooling (section 3.2.1.1.2.).

The ES between treatment and placebo (specific treatment effect) and for the change from baseline scores in treatment arm (overall treatment effect) and associated 95% CIs were presented for each MA. ES were interpreted using the following accepted convention: 0.2 is a small difference, 0.5 a moderate, and 0.8 a large. For ease of interpretability, the ES was also back-translated to change scores on 0-10 cm VAS scale (Guyatt et al., 2013). As per Wandel et al. (2010), the ES were multiplied by a median pooled SD of 2.5cm, which was obtained from large scale OA trials.

3.2.1.1 Specific treatment effect

The specific effect of a treatment represents the difference between the treatment and placebo. CMAs of the specific effects of topical NSAIDs and capsaicin were conducted as specified above.

The analyses conducted for each treatment comparison are described below:

1. Primary analysis of pain relief at or nearest to four weeks
 - a. Four weeks was chosen *a priori* as the primary time point as this was expected to be the most commonly presented time point across the treatment comparisons. In addition, previous MAs (Table 1-5 and Table 1-7) indicated that four weeks was sufficient to capture the treatment effects of both topical NSAIDs and capsaicin. Finally, OA requires long-term treatment and four weeks was felt to adequately reflect this
2. Cumulative MA of pain relief at or nearest to four weeks
 - a. Cumulative MA conducts separate MAs, each consecutively containing an additional study, until all relevant studies have been included in the pooled estimate. When studies are sorted in chronological order, it allows an examination of how evidence has accumulated and how the effect estimate has changed over time (Borenstein et al., 2011)
3. Examinations for publication bias and small-study effects

- a. Funnel plots of the ES against its standard error (SE) were plotted and examined visually for signs of asymmetry
 - b. Formal tests for publication bias were conducted if more than ten studies were available (Higgins JPT, 2011)
 - c. Egger's test was used to test for statistically significant evidence of asymmetry (Egger et al., 1997)
 - d. Duval and Tweedie's nonparametric "trim and fill" method was used to further examine for any potential publication bias (Duval and Tweedie, 2000)
 - e. The impact of small-study effects was examined in a meta-regression of specific ES and trial size at four weeks. Previous work has shown that smaller studies often associate with larger ES estimates in OA trials (Nüesch et al., 2010)
4. Exploration of the data: Sensitivity analyses, subgroup analyses, meta-regression, and analyses by treatment duration
- a. Sensitivity analyses by study quality were carried out using two indicators of study quality: (1) adequate allocation concealment and (2) treatment arm size greater than 100 participants. Previous work has indicated that these are indicators of high quality trials (Schulz et al., 1995, Nüesch et al., 2010)
 - i. Further exploration was conducted by applying Duval and Tweedie's nonparametric "trim and fill" method to studies with adequate allocation concealment (Duval and Tweedie, 2000)
 - b. Pain was analysed as a continuous outcome and sensitivity analysis, including studies that had presented pain as a dichotomised outcome, was conducted. If continuous outcomes were not the predominant outcome type presented, all outcomes types were combined for the primary analysis in order to minimise the systematic loss of data. The ES for all data types were calculated as follows:
 - i. The ES for continuous within-arm data were calculated in Stata using the "metan" command

- ii. The ES from dichotomous outcomes was calculated as described in section 3.2.1.1.1.2
- c. Subgroup analyses by weeks of treatment were conducted. Different follow-up time points were presented in the included trials and each time point was analysed separately. Time points were grouped as required using the methods described in Table 3-1.
- d. Where more than 10 trials were available overall, meta-regression for ES, trial size, and duration of follow-up was conducted using data at all follow-up time points
 - i. ES was the dependent variable.
 - ii. Sample size and duration of follow-up were first examined in separate models, each including only sample size or duration as the independent variable
 - iii. Multivariable meta-regression was then conducted including both sample size and duration as independent variables in the model.

3.2.1.1.1 Formulae

3.2.1.1.1.1 Grouping multiple time points per study

To avoid double-counting, each study contributed data only once per primary MA or MA subgroup. Where a trial reported multiple time points at durations of follow-up greater than four weeks, these were grouped as per Table 3-1.

Table 3-1 - Formulae used to group multiple time points per study

Formulae
$ES_{combined} = \frac{1}{2}(ES_{time1} + ES_{time2})$
$Variance_{combined} = \frac{1}{4}(V_{time1} + V_{time2} + \sqrt{V_{time1}}\sqrt{V_{time2}})$
$SE_{combined} = \sqrt{Variance_{combined}}$
Reference: Borenstein et al. (2011) ES, effect size; SE, standard error; V, variance

3.2.1.1.1.2 Calculating ES from dichotomous outcomes

Where pain relief was presented as a dichotomised outcome (e.g., percent of participants with improvement in pain) the methods presented in Table 3-2 were used to estimate the ES. The OR between treatment and placebo for achieving pain relief (according to thresholds set in the publication) were extracted or calculated. The OR were then used to derive Hedges' ES between treatment and placebo for the study (Higgins JPT, 2011, Borenstein et al., 2011).

Table 3-2 - Steps for calculating the ES for dichotomous outcomes presented as number/percent with pain reduction or OR for pain reduction

Step	Associated formula												
1. Derive a 2x2 table of events (pain reduction) and non-events (no pain reduction) according to publication data. Where required, A, B, C, and D can be calculated from the total number of participants and proportion that reached the pre-defined threshold for response (i.e., pain reduction).	<table><tr><td></td><td>Pain reduction</td><td>NOT pain reduction</td><td>N</td></tr><tr><td>Intervention</td><td>A</td><td>B</td><td>n₁</td></tr><tr><td>Control</td><td>C</td><td>D</td><td>n₂</td></tr></table>		Pain reduction	NOT pain reduction	N	Intervention	A	B	n ₁	Control	C	D	n ₂
	Pain reduction	NOT pain reduction	N										
Intervention	A	B	n ₁										
Control	C	D	n ₂										
2. Calculate OR	$2. OR = \frac{AD}{BC}$												
3. Calculate log OR	$3. \log OR = \ln(OR)$												
4. Calculate the variance of log OR	$4. V_{\ln OR} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$												
5. Calculate the SE of log OR	$5. SE_{\ln OR} = \sqrt{V_{\ln OR}}$												
6. Calculate d from the calculated (steps 1-5) or extracted OR	$6. d = \frac{\sqrt{3}}{\pi} \ln OR$												
7. Calculate the SE of d	$7. SE_d = \frac{\sqrt{3}}{\pi} SE_{\ln OR}$												
8. Calculate variance of d	$8. V_d = SE_d^2$												
9. Calculate J (Hedges' conversion factor)	$9. J = 1 - \frac{3}{4(n_1+n_2-2)-1}$												
10. Calculate Hedges' g	$10. g = d \times J$												
11. Calculate variance of Hedges' g	$11. V_g = J^2 \times V_d$												
12. Calculate SE of Hedges' g	$12. SE_g = \sqrt{V_g}$												
13. Combine Hedges' g and its associated SE with other effect estimates													

Steps 1-5 and 8-12 are from Borenstein et al. (2011). Steps 6-7 are from Higgins et al (2011)
ES, effect size; OR, odds ratio; SE, standard error

3.2.1.2 Overall treatment effect and PCE

The overall treatment effect of an intervention represents the mean change from baseline pain score reported within the treatment arm. It is a measure of the pain reduction that would be experienced by a person taking the medication. A negative ES indicates an increase in pain, 0 indicates no change from baseline in pain scores, and a positive ES indicates a reduction in pain compared to baseline.

The overall effect consists of two components: the specific effect and the contextual effect. The contextual effect comprises a number of sub-effects, illustrated in Figure 3-1.

Contextual effect	Doctor-patient communication	Placebo response
	Treatment expectation	
	Conditioning	
	Natural history or fluctuation of symptoms	Other determinants
	Regression to mean	
	Biases and false positives	
	Co-interventions and other non-specific effects	

Figure 3-1 - Sub-effects of the contextual effect

Sub-effects that comprise the contextual effect of medical treatments, including factors contributing to the placebo response (doctor-patient communication, treatment expectation, and associative learning or conditioning processes) and other determinants (Schedlowski et al., 2015)

For the present work, the contextual and specific effects were assumed to combine in an additive manner (Figure 3-2). The assumption that the contextual effects are of equal size in the treatment and placebo arms is central to the analysis and interpretation of RCT data

(Schedlowski et al., 2015). However, the possibility that they combine in an interactive manner, thereby rendering unequal contextual effects in the trial arms, has been suggested (Schedlowski et al., 2015).

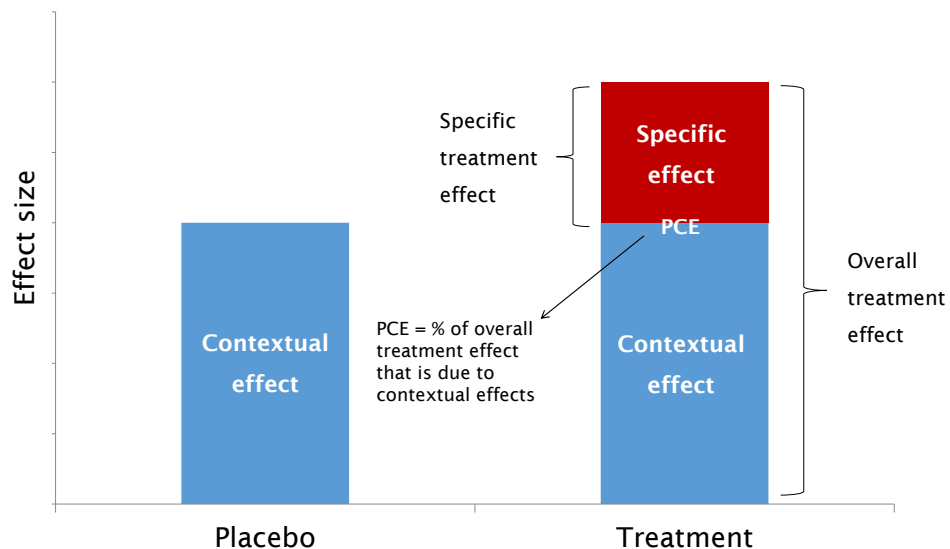


Figure 3-2 - Illustration of the overall treatment effect and PCE
PCE, proportion of contextual effects

The PCE is the percentage of the overall effect that is attributable to contextual effects and can vary from 0 (no contextual effects) to 1 (100% of overall effect is due to contextual effects). It can only be estimated for trials where both treatment and placebo arms improved from baseline as 1) the worsening may be due to nocebo or side effects and 2) the calculation of the PCE does not allow negative values (Zou et al., 2016). Thus outcome data where one or both groups worsened were excluded from the analysis of the PCE and the corresponding overall ES.

The methods for calculating the overall treatment effect and PCE were previously developed within the department and described by Zou et al. (2016). They were further refined for this project. The steps for the calculation of the overall treatment effect and PCE are presented in section 3.2.1.2.1. The subsequent section then describes the analyses conducted for this thesis.

3.2.1.2.1 Formulae

The overall ES and PCE were calculated for each study using change from baseline data. The steps are presented in Table 3-3. Steps 1-11 were carried out for both the treatment and placebo arms separately. The PCE was subsequently based on the treatment and placebo overall ES estimates.

Table 3-3 - Steps for calculating the overall ES and PCE

Steps	Associated formula
1. Calculate SMD (d)	1. $d = \frac{\text{mean change from baseline}}{\text{SD of mean change from baseline}}$
2. Calculate Hedges' correction factor (J)	2. $J = 1 - \frac{3}{4(n-1)-1}$
3. Calculate SMD with Hedges' correction (g)	3. $g = J \times d$
4. Calculate variance of uncorrected SMD (V_d), assuming $r=0.5$ (Fu et al., 2013)	4. $V_d = \frac{1}{n} + \frac{d^2}{2n}$
5. Calculate variance of Hedges' SMD (V_g)	5. $V_g = J^2 \times V_d$
6. Calculate SE of Hedges' SMD (SE_g)	6. $SE_g = \sqrt{V_g}$
7. Calculate pooled overall treatment effect for treatment and placebo arms	7. Random effects MA (Stata)
8. Calculate lnPCE	8. $\ln PCE = \ln(g_P) - \ln(g_T)$ P= placebo, T=Treatment
9. Calculate variance of lnPCE ($V_{\ln PCE}$)	9. $V_{\ln PCE} = \frac{(SE_{gP})^2}{n_P \times (g_P)^2} + \frac{(SE_{gT})^2}{n_T \times (g_T)^2}$
10. Calculate SE of lnPCE ($SE_{\ln PCE}$)	10. $SE_{\ln PCE} = \sqrt{V_{\ln PCE}}$
11. Calculate pooled exponentiated PCE	11. Random effects MA (Stata)

Steps 1-8 are adapted from Borenstein et al. (2011) and steps 8-11 are adapted from Hedges et al. (1999)

MA, meta-analysis; PCE, proportion of contextual effects; SD, standard deviation; SE, standard error; SMD, standardised mean difference

3.2.1.2.2 Meta-analysis methods

The overall treatment effects of the drugs were calculated using the previously described steps (Table 3-3). The subsequent MAs conducted reflect those of the specific treatment effect and are outlined below:

1. Primary analysis of overall treatment effect and PCE at or nearest to four weeks
2. Cumulative MA of overall treatment effect at or nearest to four weeks
3. Exploration of the data: Sensitivity and subgroup analyses
 - a. Sensitivity analyses by study quality using two indicators of study quality: (1) adequate allocation concealment and (2) treatment arm size over 100 participants
 - b. Subgroup analysis of the overall treatment effect by drug class
 - c. Subgroup analyses by weeks of treatment. Time points over four weeks were grouped as described in the analyses of the specific effects (section 3.2.1.1).

3.2.2 Network meta-analysis

A NMA was conducted for topical NSAIDs and capsaicin in OA using placebo as a common comparator. The trial network was plotted to illustrate the interventions and common comparator (placebo). The nodes were linked to illustrate the direct and indirect comparisons available for analysis. Both Bayesian and frequentist approaches were adopted for the NMA.

The ES calculated for the primary specific treatment effect CMAs were used for the NMA. The primary analysis included all trials analysed in the CMA. A subgroup analysis was then conducted for topical NSAIDs and capsaicin used as recommended in the BNF (BNF, 2018). Table 3-4 lists the drugs and associated concentrations recommended by the BNF for topical application in OA. Trials were excluded from the subgroup analysis if the investigated drug concentration was lower than recommended in the BNF. The subgroup analysis was conducted to provide evidence for treatments currently available in practice.

Table 3-4 - Topical NSAID and capsaicin recommendations for topical administration in OA

Drug	Concentration
Diclofenac diethylammonium	1.16%
Diclofenac sodium*	1%
Felbinac	Concentration not listed
Ibuprofen	5% or 10%
Ketoprofen	2.5%
Piroxicam	0.5%
Capsaicin	0.025%

* Equivalent to 1.16% diclofenac diethylammonium (eMC, 2017)

List represents the recommended topical NSAID drugs, topical NSAID

concentrations, and capsaicin concentration for use in OA (BNF, 2018)

BNF, British National Formulary; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis

3.2.2.1 Bayesian analysis

A random-effects Bayesian NMA was conducted using Markov chain Monte Carlo simulations. Non-informative prior distributions were set for parameters and three Markov chains were run simultaneously with different initial values, chosen arbitrarily. A normal likelihood distribution was assumed for the ES. The model fit, assessed using the posterior mean residual deviance, was examined and deemed appropriate. The chains converged within 10,000 iterations and a total of 20,000 simulations comprised the burn-in period. The subsequent 50,000 iterations were examined.

The median of the posterior distribution of the ES was generated. The 2.5th and 97.5th percentiles of the posterior distribution comprised the lower and upper limits of the 95% credible interval (CrI). The CrI represents the boundary where the true value is believed to lie with 95% certainty. Although this interpretation is often used for the frequentist 95% CI, such probability statements can only be attributed to the CrI. Strictly speaking, the interpretation of the 95% CI is that if numerous confidence intervals were drawn from a population, 95% of them would contain the true value.

To facilitate the interpretation of the results, the treatments were ranked using the mean posterior probability of each treatment being the best. The probability of each treatment being the best is generated using the posterior distributions for all relative treatment effects generated within the Bayesian framework. Random samples are “drawn” and the number of times that an intervention ranks first out of the total number of random samples gives the probability of that treatment being the best.

All Bayesian analyses were conducted in WinBUGs software (version 1.4.3, MRS Biostatistics Unit UK, 2007) using methods supplied by the NICE Decision Support Unit (Dias et al., 2011, updated 2016).

3.2.2.2 Frequentist analysis

A random-effects NMA was conducted in a frequentist framework. The pooled ES of each direct and indirect treatment comparison was

calculated and presented with its corresponding 95% CI. The estimated probability of each treatment being the best was also presented. The probability was generated using frequentist resampling methods, by drawing on a parametric bootstrap procedure that is analogous to drawing from a Bayesian posterior distribution (White et al., 2012). The probability of the treatment being the best is estimated by the proportion of samples in which the treatment ranked first. Analyses were conducted in Stata using the “network” command (White, 2015).

3.2.3 GRADE assessment of quality

The quality of the CMA and NMA evidence was determined using GRADE guidelines (Balshem et al., 2011). In the current context, the GRADE rating of quality reflects our confidence that the treatment effect estimate is correct (Balshem et al., 2011). The lower the quality, the greater the chance that further research would change the estimate and our confidence in it (Balshem et al., 2011).

The GRADE approach involves initially rating the quality of the evidence as high when it is based on RCTs and low when based on observational studies. The quality is then downgraded as appropriate based on the examination of the following domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias (Guyatt et al., 2011b, Guyatt et al., 2011c, Guyatt et al., 2011d, Guyatt et al., 2011e, Guyatt et al., 2011g). The rating of the evidence for the specific treatment effect analyses commenced as high quality (RCT data used) and was downgraded as appropriate. The evidence for the overall treatment effect were considered low quality to start. This is because examination of the treatment arm alone effectively removes the randomisation process, thus rendering the data observational. The quality rating was then upgraded as appropriate (Guyatt et al., 2011f).

The quality of the treatment effect estimate from the NMA was also determined using the GRADE approach (Puhan et al., 2014, Brignardello-Petersen et al., 2018). The evidence rating commenced as per the lowest quality rating of the direct estimates that formed the treatment loop. The body of evidence was then examined for

imprecision, intransitivity, and incoherence (where possible) and graded accordingly.

The evidence for the primary analyses (at or nearest to four weeks) were presented in a summary of findings table. Evidence profile tables, including the justifications for the quality rating, were compiled and presented in the appendices.

3.3 Results

The following section reports the results of the conventional and network MAs of topical NSAIDs and capsaicin in OA. It first provides a description of the studies included in the analyses. It then reports the specific treatment effects of topical NSAIDs and capsaicin. Having established the drugs' efficacy relative to placebo, the section then presents the results of the NMA. Finally, the section describes the overall effect and PCE of topical NSAIDs and capsaicin.

Throughout the section, the results of topical NSAIDs are followed by those of capsaicin. The findings at or nearest to four weeks (primary endpoint) are presented first. The remaining analyses are presented as listed in the methods.

3.3.1 Description of studies

The following section describes the results of the literature search, the details of the trials included in the analyses, and the risk of bias summaries.

3.3.1.1 Topical NSAIDs

3.3.1.1.1 Literature search

The results of the systematic literature search are presented (Figure 3-3). Publications for 52 studies that compared a topical NSAID to placebo or active comparator were identified in the systematic literature search, of which 30 studies compared topical NSAID to placebo. Three of these included non-eligible participants and were not included in the analyses. A further two studies did not report data in sufficient detail or in the correct format for extraction and could not be included in the analysis of topical NSAIDs versus placebo (Ergun et al., 2007, Kageyama, 1987). Pain could only be extracted as a dichotomous outcome for four studies and they were only included in the sensitivity analysis (section 3.3.2.1.1.3.2) (Yamamoto et al., 1979, Kageyama, 1984, Rose et al., 1991, Barthel et al., 2009). 21 studies contributed to the primary analyses described in this chapter.

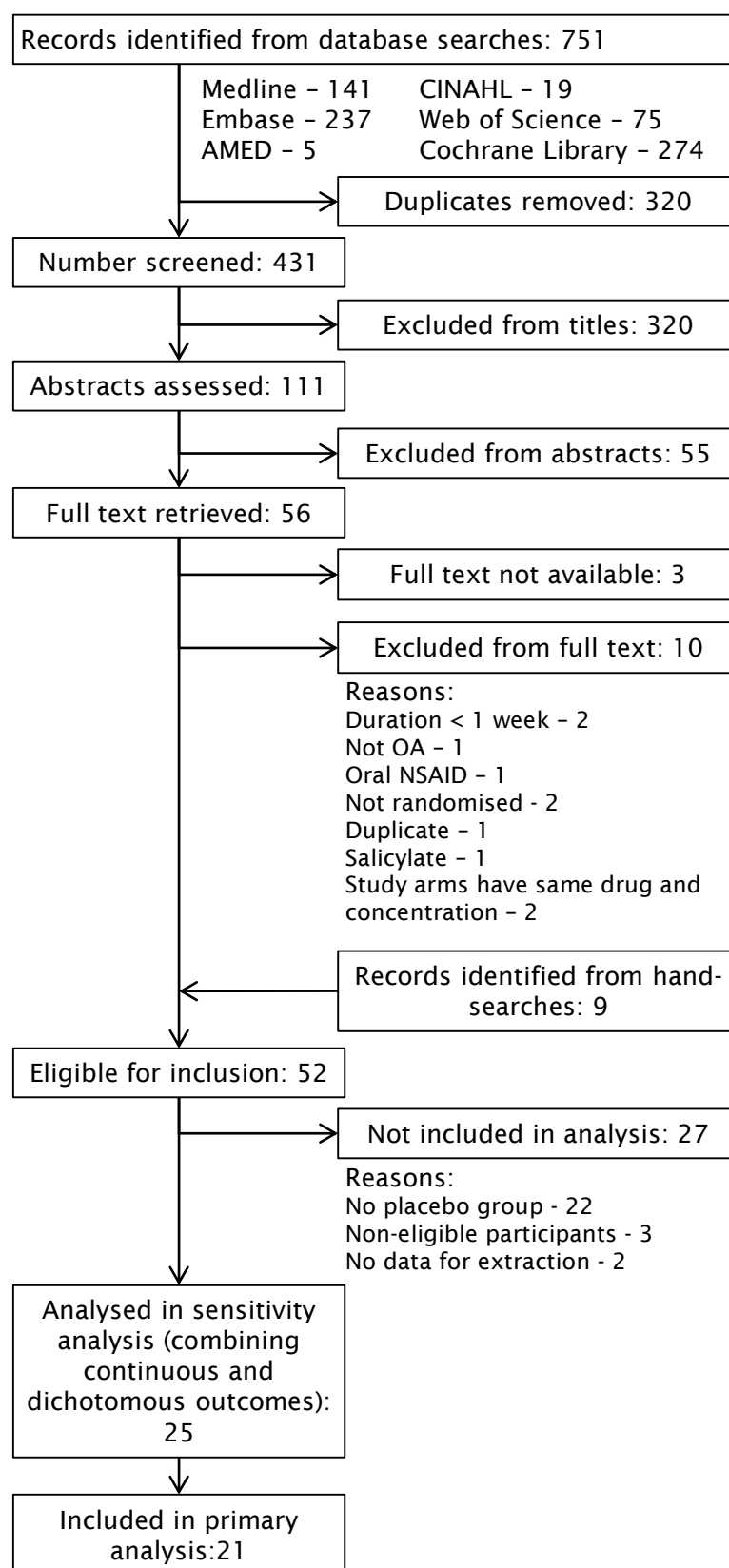


Figure 3-3 - Flow Diagram: Literature search results for topical NSAIDs in OA

3.3.1.1.2 Description of the studies

21 parallel-design studies were included in the CMAs and NMAs of topical NSAIDs in OA. These studies are summarised in Table 3-5.

A total of 6,191 participants with radiographic OA were analysed for the primary endpoint, of which 3,183 were on topical NSAIDs and 3,008 were on placebo. All but one study recruited participants with knee OA. The remaining study consisted of 385 participants with hand OA (Altman et al., 2009).

The study by Yoo et al. (1996) included only women. All other studies included both men and women, but all had a higher proportion of female participants. Overall, 65% of the participants analysed for the primary outcome were women.

The majority of studies were commercially funded and were conducted primarily in North America and Europe. Four trials were published in a language other than English. Two were in Japanese (Yamamoto et al., 1979, Kageyama, 1984), one was in German (Rose et al., 1991), and one was in Korean (Yoo et al., 1996). The data were extracted by a native speaker for the Japanese and German publications (Drs Haruna Takahashi and Neele Dellschaft). A native Korean speaker was not found and the data were extracted using Google Translate for this publication.

A further four studies, consisting of 813 participants were included in the sensitivity analysis combining pain as a continuous and dichotomous measure (n=7,004 in total). Aside from the study by Barthel et al. (2009) these studies differed from those analysed in the primary analyses as they were published earlier and used hospital-based participants. Furthermore, the study by Yamamoto et al. (1979) was the only one to include participants with ankle OA. The four studies included only in the sensitivity analysis with dichotomous outcomes are indicated with * in Table 3-5 and Table 3-6.

Table 3-5 - Description of the studies included in the analysis of topical NSAIDs in OA

Trial Name	Year	Country of Study	Funding source	Design	Setting	Joints Affected
Yamamoto 1979*	1979	Japan	Unclear	Parallel	Hospital	Ankle, knee
Kageyama 1984*	1984	Japan	Unclear	Parallel	Hospital	Knee
Rose 1991*	1991	Germany	Unclear	Parallel	Hospital	Knee
Dreiser 1993	1993	France	Unclear	Parallel	Community	Knee
Yoo 1996	1996	Korea	Unclear	Parallel	Community	Knee
Sandelin 1997	1997	Finland, Sweden	Unclear	Parallel	Community	Knee
Grace 1999	1999	Canada	Public, Commercial	Parallel	Unclear	Knee
Rovensky 2001	2001	Slovakia	Unclear	Parallel	Community	Knee
Ottillinger 2001	2001	Hungary, Czech Republic, Switzerland, Germany	Commercial	Parallel	Community	Knee
Bruhlmann 2003	2003	Switzerland	Unclear	Parallel	Community	Knee
Roth 2004	2004	USA	Commercial	Parallel	Community	Knee
Trnavsky 2004	2004	Czech Republic	Commercial	Parallel	Community	Knee
Bookman 2004	2004	Canada	Commercial	Parallel	Community	Knee
Baer 2005	2005	Canada	Commercial	Parallel	Community	Knee
Niethard 2005	2005	Germany	Commercial	Parallel	Community	Knee
Rother 2007	2007	Germany	Commercial	Parallel	Community	Knee
Barthel 2009*	2009	USA	Commercial	Parallel	Community	Knee

Continued overleaf

Trial Name	Year	Country of Study	Funding source	Design	Setting	Joints Affected
Simon 2009	2009	Canada, USA	Commercial	Parallel	Community	Knee
Altman 2009	2009	USA	Commercial	Parallel	Unclear	Hand
Baraf 2010	2010	USA	Commercial	Parallel	Community	Knee
Kneer 2013	2013	Germany, Poland, Serbia, Croatia	Commercial	Parallel	Community	Knee
Rother 2013	2013	USA	Commercial	Parallel	Community	Knee
Varadi 2013	2013	Switzerland	Commercial	Parallel	Community	Knee
Conaghan 2013	2013	Czech Republic, Germany, Poland, UK	Commercial	Parallel	Community	Knee
Shoara 2015	2015	Iran	Public	Parallel	Community	Knee

*Only included in sensitivity analysis combining continuous and dichotomous pain outcomes (section 3.3.2.1.1.3.2)

Year is year of publication

The studies examined a variety of topical NSAIDs and the details of these interventions and their comparators are shown in Table 3-6.

Various topical NSAIDs, such as ketoprofen, diclofenac sodium, and ibuprofen, were applied for one to 12 weeks duration. Diclofenac sodium was the most commonly examined topical NSAID (nine studies), followed by ketoprofen (six studies). The topical NSAIDs were available as gels, creams, solutions, and patches (plasters). Carriers, such as DMSO and Transfersome gel, were commonly used in the intervention and vehicle preparations. These vehicles were regarded as placebo for this MA.

Studies commonly had more than two treatment arms, with a maximum of six treatment arms. Where appropriate, multiple arms were combined prior to the MA (indicated with ** in Table 3-6).

Table 3-6 - Description of the interventions and comparators in the studies included in the analysis of topical NSAIDs in OA

Trial Name	Intervention (Carrier)	Strength	Formulation	Frequency per Day	Comparator(s)
Yamamoto 1979*	Indomethacin	1%	Ointment	3	Placebo
Kageyama 1984*	Ketoprofen	3%	Ointment	3	Placebo
Rose 1991*	Piroxicam	0.5%	Gel	4	Placebo
Dreiser 1993	Diclofenac hydroxyethylpyrrolidine	180mg	Patch	2	Placebo
Yoo 1996	Ketoprofen		Patch	2	Placebo
Sandelin 1997	Eltenac	1%	Gel	3	Placebo (oral + gel), oral diclofenac
Grace 1999	Diclofenac sodium (pluronic lecithin)	2%	Gel	3	Placebo (pluronic lecithin vehicle)
Rovensky 2001	Ibuprofen	5%	Cream	3	Placebo
Ottillinger 2001	Eltenac	1%**, 0.3%**, 0.1%**	Gel	3	Placebo
Bruhlmann 2003	Diclofenac epolamine	180mg	Patch	2	Placebo
Roth 2004	Diclofenac sodium (DMSO)	1.5%	Solution	4	Placebo (45.5% DMSO)
Bookman 2004	Diclofenac sodium (DMSO)	1.5%	Solution	4	Placebo (4.55% DMSO)**, placebo (45.5% DMSO)**
Trnavsky 2004	Ibuprofen	5%	Cream	3	Placebo
Niethard 2005	Diclofenac diethylamine	1.16%	Gel	4	Placebo

Continued overleaf

Trial Name	Intervention (Carrier)	Strength	Formulation	Frequency per Day	Comparator(s)
Baer 2005	Diclofenac sodium (DMSO)	1.5%	Solution	4	Placebo (45.5% DMSO)
Rother 2007	Ketoprofen (Transfersome)	110mg	Gel	2	Placebo (gel + oral), oral celecoxib
Barthel 2009*	Diclofenac sodium	1%	Gel	4	Placebo
Altman 2009	Diclofenac sodium	1%	Gel	4	Placebo
Simon 2009	Diclofenac sodium (DMSO)	1.5%	Solution	4	Placebo (oral + modified 2.3% DMSO)**, placebo (45.5% DMSO)**, oral diclofenac**, oral diclofenac + topical diclofenac sodium
Baraf 2010	Diclofenac sodium	1%	Gel	4	Placebo
Rother 2013	Ketoprofen (Transfersome)	100mg	Gel	2	Placebo (4.4 g Transfersome Vehicle)
Conaghan 2013	Ketoprofen (Transfersome)	50mg**, 100mg**	Gel	2	Placebo (oral)**, placebo (2.2 g Transfersome)**, placebo (4.4 g Transfersome)**, oral celecoxib
Varadi 2013	Ibuprofen-VALE	10%	Cream	2	Placebo
Kneer 2013	Ketoprofen	25mg**, 50mg**, 100mg**	Gel	2	Placebo
Shoara 2015	Diclofenac sodium	1%	Gel	3	Placebo, chamomile oil

* Only included in sensitivity analysis combining continuous and dichotomous pain outcomes

** Multiple treatment arms combined prior to analysis, as per Table 3-1

DMSO, dimethylsulfoxide

3.3.1.1.3 Risk of bias assessment

A risk of bias assessment was completed for the included studies and is presented in Figure 3-4. The main deficiencies were failure to use ITT analysis and incomplete outcome reporting.

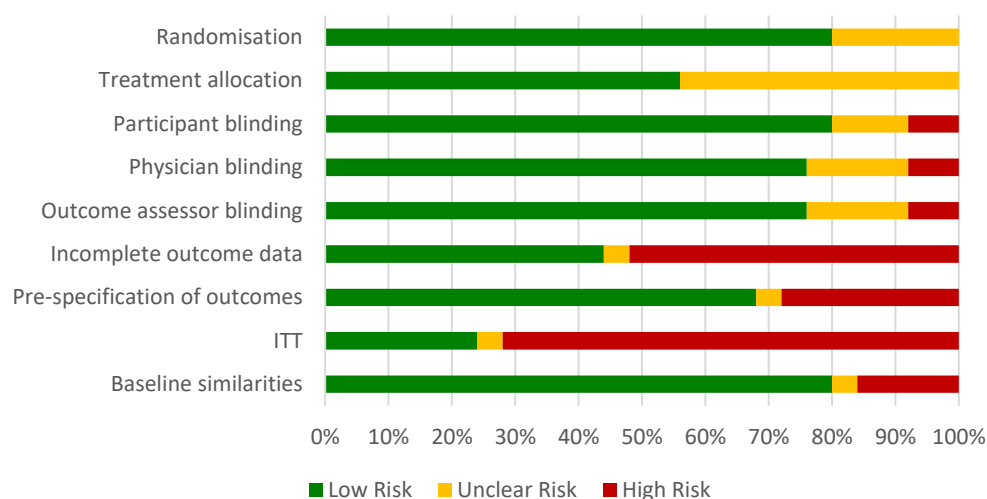


Figure 3-4 - Risk of bias assessment for topical NSAID versus placebo trials in OA

The method of randomisation was adequate in 80% of trials, with most describing computer-generated randomisation schedules. Imbalances in the treatment arms at baseline were evident in 16% of trials (four studies). Two studies (Trnavsky et al., 2004, Shoara et al., 2015) had significant differences in the number of women in the treatment and the placebo groups (92% vs 64% in Trnavsky et al., 2004 and 67% vs 83% in Shoara et al., 2015). One study (Varadi et al., 2013) reported higher WOMAC and pain scores at baseline in the treatment compared to the placebo group. The final study (Ottillinger et al., 2001) reported greater lateral JSN and longer duration of pain in the placebo group.

Treatment allocation concealment was deemed adequate in 56% of studies. The remaining studies did not describe procedures for allocation concealment. It was not possible to estimate the associated risk of bias without sufficient information.

Although all studies were described as blinded in their publications, four did not provide sufficient detail of the blinding procedure for its adequacy to be judged. A further two studies did not adequately blind the participants or investigators. Of these, one study (Conaghan et al., 2013) compared topical and oral treatments, but did not use a double-dummy technique. The other (Shoara et al., 2015) compared a topical NSAID *gel* with a placebo *oil* and used different packaging for the treatments.

In 18 of 25 studies, participants were excluded from analysis post-randomisation (using PP analysis or “modified ITT analysis”). This was the risk of bias domain most commonly associated with a high risk of bias.

3.3.1.2 Capsaicin

3.3.1.2.1 Literature search

The systematic literature search identified eight RCTs of capsaicin in OA. Two studies (Schnitzer et al., 2012, Fish et al., 2008) compared capsaicin to an active comparator, one study (McCarthy and McCarty, 1992) combined participants with OA and RA, and the remaining five studies were included in the analysis of capsaicin versus placebo. The flow diagram of the results of the systematic literature search are described in Figure 3-5.

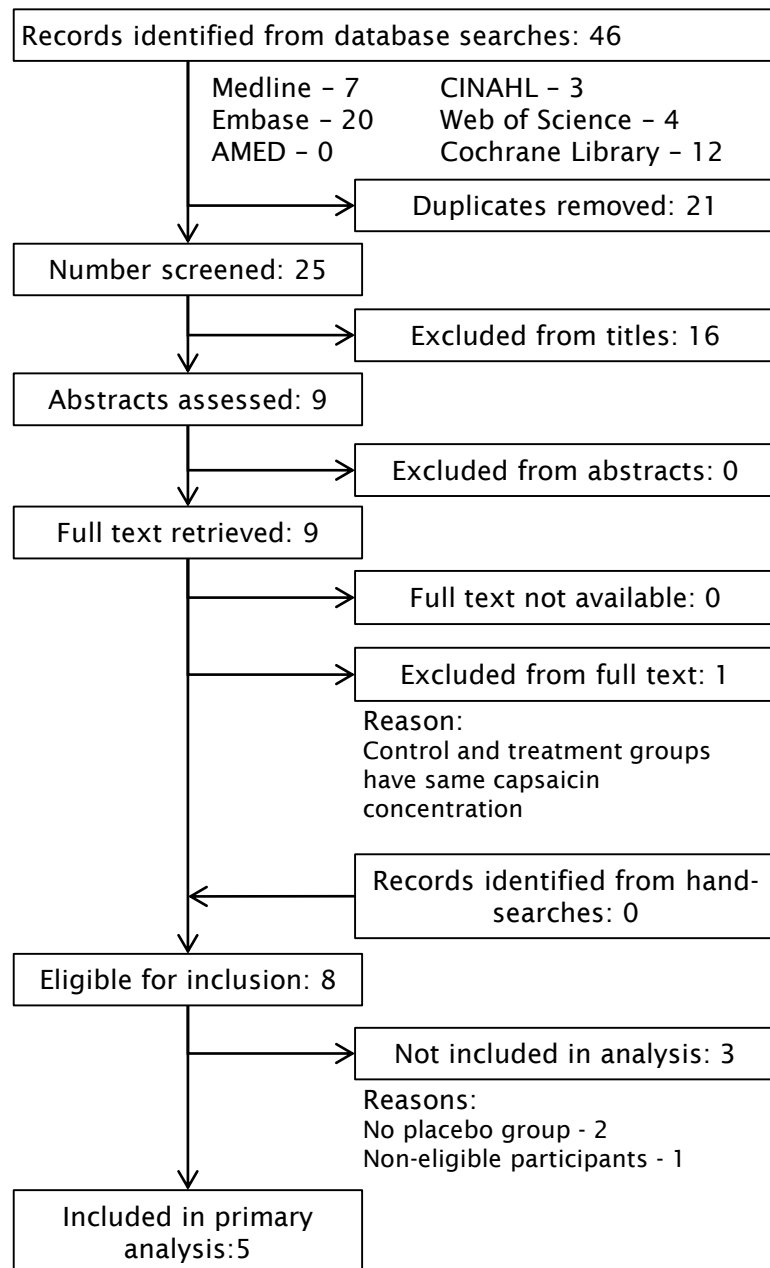


Figure 3-5 - Flow Diagram: Literature search results for capsaicin in OA

3.3.1.2.2 Description of the studies

The five placebo controlled trials contained 415 participants. Of these, 206 were on capsaicin and 209 were on placebo. The median number of participants analysed per study was 80 and this ranged from 42 to 130 participants per trial. Participants had radiographic OA affecting a variety of joints, including the hand, elbow, shoulder, hip, knee, and ankle. Four of the five trials were of parallel design. One study (Kosuwon et al., 2010) was a cross-over trial and only the first treatment period was extracted for analysis. The study by Kosuwon et al. (2010) was also the only study that was limited to female participants. All trials were published in English and are described in Table 3-7.

Capsaicin was compared to placebo in all five studies, and details of the interventions are given in Table 3-8.

All but one study (Kosuwon et al., 2010) investigated 0.025% capsaicin cream applied to the affected joint four times per day. Four trials were two-arm trials, comparing capsaicin only to placebo. One study compared capsaicin to placebo, glyceryl trinitrate (GTN), and a GTN-capsaicin combination. Only the capsaicin-only and placebo groups were analysed for the MA.

Table 3-7 - Description of the studies included in the analysis of capsaicin in OA

Trial Name	Year	Country	Funding Source	Design	Setting	Joints Affected
Deal 1991	1991	USA	Public	Parallel	Community	Knee
Altman 1994	1994	USA	Commercial	Parallel	Community	Wrist, Elbow, Shoulder, Knee, Ankle
Schnitzer 1994	1994	USA	Commercial	Parallel	Unclear	Hand
McCleane 2000	2000	UK	Commercial	Parallel	Community	Hand, Shoulder, Hip, Knee
Kosuwon 2010	2010	Thailand	Mixed	Crossover	Community	Knee

Year is year of publication

Table 3-8 - Description of the interventions and comparators in the studies included in the analysis of capsaicin in OA

Trial Name	Intervention	Capsaicin concentration	Formulation	Frequency per day	Comparator(s)
Deal 1991	Capsaicin	0.025%	Cream	4	Placebo
Altman 1994	Capsaicin	0.025%	Cream	4	Placebo
Schnitzer 1994	Capsaicin	0.025%	Cream	4 (for first 3 weeks) then 2	Placebo
McCleane 2000	Capsaicin	0.025%	Cream	4	Placebo, GTN, GTN + capsaicin
Kosuwon 2010	Capsaicin	0.0125%	Gel	3	Placebo

GTN, glyceryl trinitrate

3.3.1.2.3 Risk of bias assessment

The results of the risk of bias assessment for the capsaicin versus placebo trials are presented in Figure 3-6. Overall, the included studies were associated with high risks of bias.

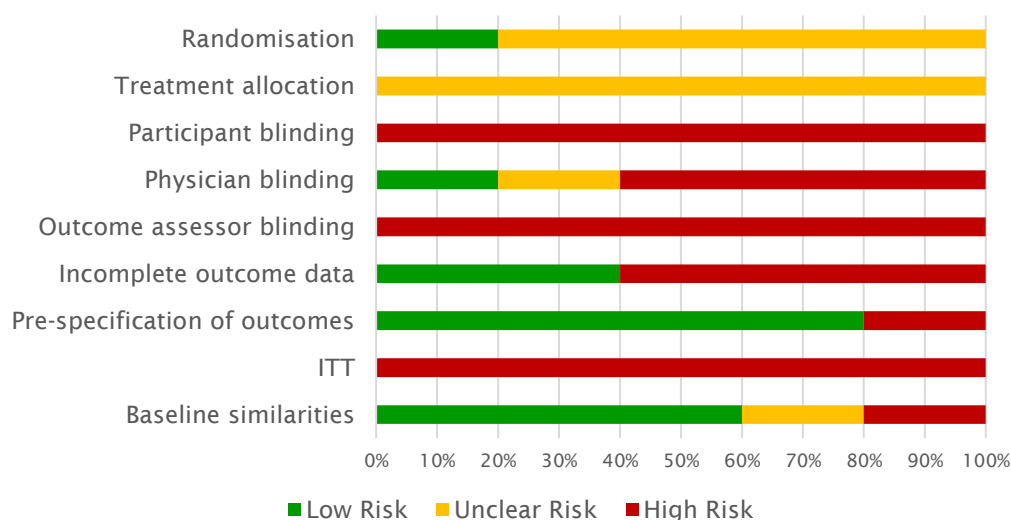


Figure 3-6 - Risk of bias assessment for capsaicin versus placebo trials in OA

The domains associated with the highest risk of bias were inadequate blinding and failure to use ITT analysis. Despite claims of being double blinded, the trials were likely inadequately blinded due to the warming sensation commonly experienced with initial applications of capsaicin, but not with placebo. All studies excluded some participants from analysis post-randomisation.

Only one study (McCleane, 2000) described a computer-generated randomisation procedure. The remaining studies did not provide sufficient detail regarding the randomisation sequence generation for the risk of bias to be determined. Similarly, no studies mentioned precautions taken to conceal allocation.

For the majority of studies, outcomes were reported as described in the methods. This was the domain associated with the lowest risk of bias.

3.3.2 Specific treatment effects

The following section describes the specific treatment effects of topical NSAIDs and capsaicin in OA, i.e., the difference between treatment and placebo groups for pain relief.

3.3.2.1 Topical NSAIDs versus placebo (CMA)

3.3.2.1.1 At or nearest to four weeks

21 studies with 6,191 eligible participants were analysed for the primary outcome of topical NSAID versus placebo at or nearest to four weeks of treatment.

Topical NSAIDs were statistically significantly superior to placebo (ES 0.31, 95%CI 0.20 to 0.41) (Figure 3-7). The ES is equivalent to a 0.8 cm (95%CI 0.5 to 1.0) greater decrease in VAS pain score in the topical NSAID group compared to the placebo group. Substantial and significant heterogeneity ($I^2 = 72.1\%$, $p < 0.001$) was evident. There is moderate confidence in the effect estimate due to inconsistency (Appendix 3 – GRADE assessment: Topical NSAIDs in OA).

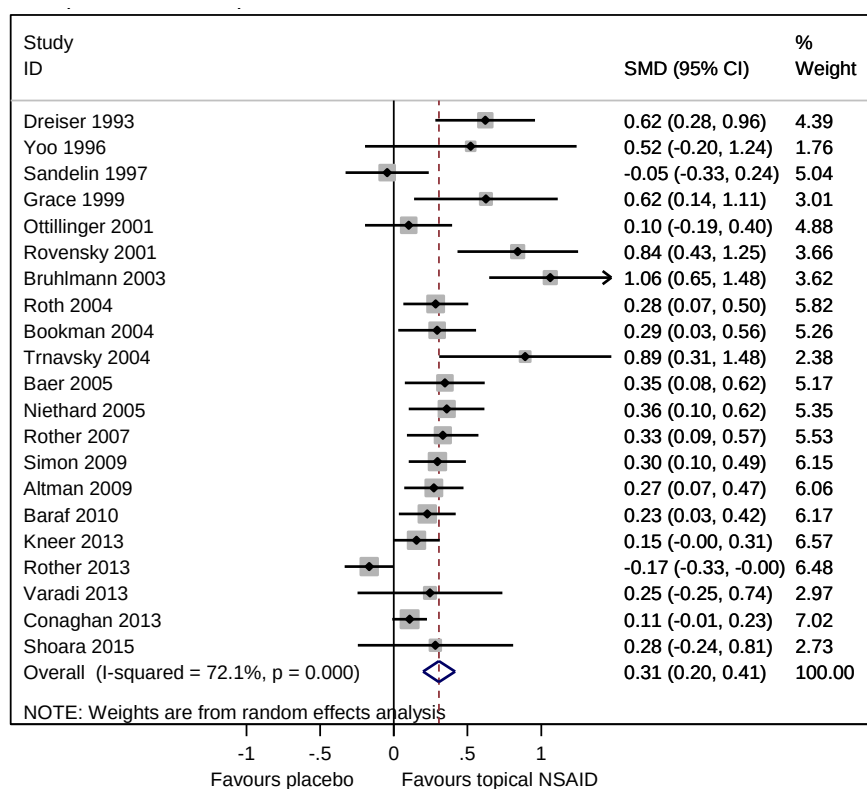


Figure 3-7 - Topical NSAIDs versus placebo in OA. Outcome: pain at or nearest to four weeks

3.3.2.1.1.1 Cumulative meta-analysis

The oldest included study was published in 1993 (Dreiser and Tisne-Camus, 1993) and the newest study was published in 2015 (Shoara et al., 2015). A cumulative MA (Figure 3-8) was conducted to determine how the ES has changed with increasing amounts of evidence.

Although still remaining statistically superior to placebo, the difference between topical NSAID and placebo has decreased at a constant rate in the last decade as new trials have been conducted. The precision of the estimate has increased with the addition of new evidence.

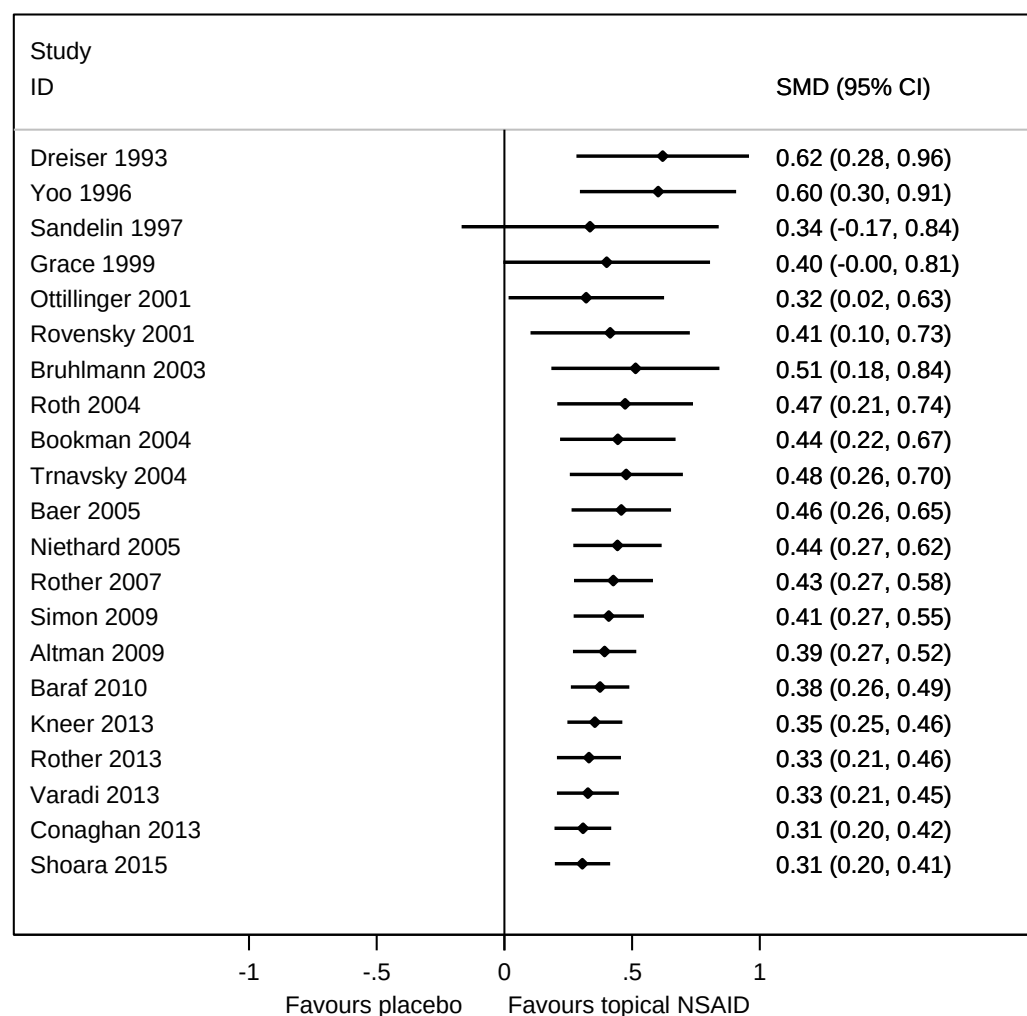


Figure 3-8 - Cumulative MA of topical NSAIDs versus placebo in OA. Outcome: pain at or nearest to four weeks

3.3.2.1.1.2 Publication bias and small study effects

The funnel plot of the topical NSAID studies is presented in Figure 3-9. Clear asymmetry is evident on visual inspection of the funnel plot. The presence of small study effects was confirmed by Egger's test ($p=0.002$). The intercept of Egger's test (2.82) indicated that smaller studies were associated with larger effects.

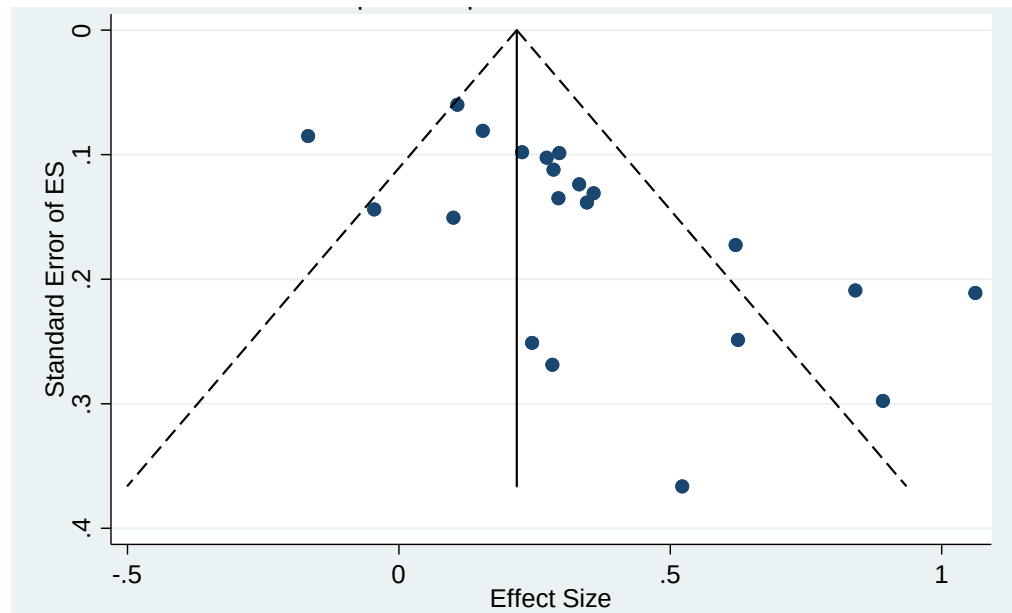


Figure 3-9 - Funnel plot with pseudo 95% confidence limits for topical NSAIDs in OA

ES and SE of ES for topical NSAIDs versus placebo in OA. Outcome: Pain at or nearest to four weeks

Using Duval and Tweedie's nonparametric "trim and fill" method to account for publication bias (Duval and Tweedie, 2000) the pooled random ES was lowered to 0.19, but remained statistically significant (95%CI 0.07 to 0.30).

The results of the "filled" funnel plot are presented in Figure 3-10. As in the traditional funnel plot, each study included in the MA is represented by a data point. The data points surrounded by a square represent the "missing" studies that were imputed using Duval and

Tweedie's methods. The “missing” studies mirror the extreme positive studies in order to create symmetry within the funnel plot.

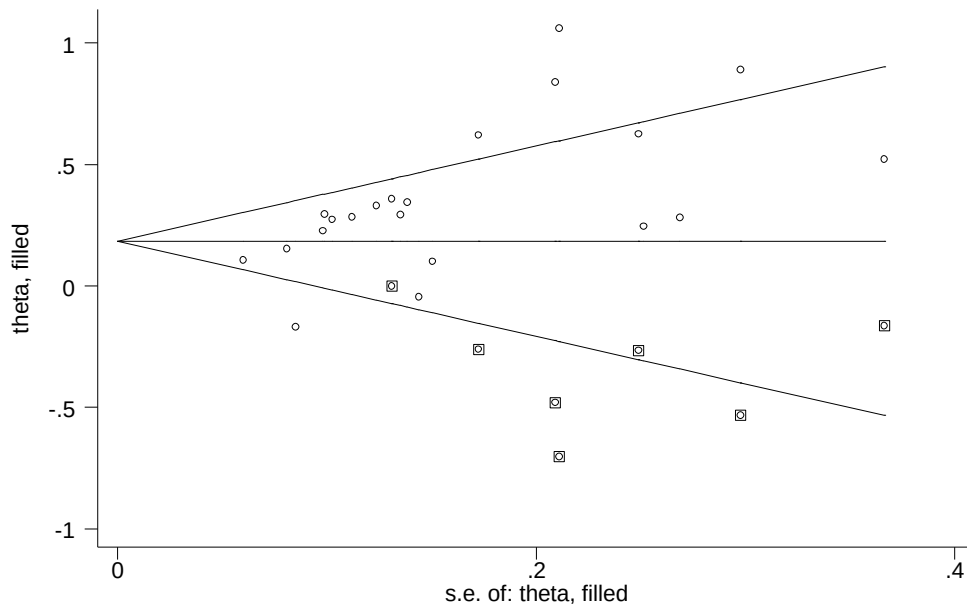


Figure 3-10 - Filled funnel plot of topical NSAIDs in OA

Outcome: Pain at or nearest to four weeks. Each included trial is indicated by a circle. Circles within squares indicate the new, imputed studies that fill the funnel plot to create symmetry

Meta-regression of the treatment effect at or nearest to four weeks and trial sample size identified a statistically significant association ($p=0.018$). Trials with smaller sample sizes reported larger treatment effects (coefficient = -0.0004576 , 95%CI -0.0008292 to -0.000086).

3.3.2.1.1.3 Sensitivity analyses

3.3.2.1.1.3.1 By study quality

Sensitivity analyses were conducted to examine the ES of topical NSAIDs excluding lower quality trials (Table 3-9). Two indicators were taken as surrogates for overall trial quality:

1. A low risk of bias score for allocation concealment
2. A sample size greater than 100 participants per treatment arm

Excluding lower quality trials did not significantly alter the ES of topical NSAIDs at or nearest to four weeks in OA. Including only studies with adequate allocation concealment yielded a largely unchanged ES estimate and 95% CI (0.32, 95%CI 0.23 to 0.42). Including only large studies produced a smaller ES (0.20) which remained statistically superior to placebo (95%CI 0.10 to 0.31).

**Table 3-9 - Sensitivity analyses by study quality for topical NSAIDs in OA.
Outcome: Pain at or nearest to four weeks**

Studies included	Trials n		ES (95%CI)	I ² (p-value)
All studies	21	6,191	0.31 (0.20 to 0.41)	72.1% (<0.001)
Allocation concealment	11	2,676	0.32 (0.23 to 0.42)	24.9% (0.207)
n>100 in treatment arm	10	4,893	0.20 (0.10 to 0.31)	65.7% (0.002)

Presented as Hedges' effect size (ES) and associated 95% confidence interval (95%CI). I² (the variation in ES attributable to heterogeneity) and associated p-value are presented
n, number of participants; Trials, number of trials

Due to the presence of small study effects, studies with adequate allocation concealment were also examined using the previously described "trim and fill" approach. No trimming was required and no "missing" studies were imputed. The ES was therefore unchanged by the "trim and fill" approach (ES 0.32, 95%CI 0.23 to 0.42).

3.3.2.1.1.3.2 By combining continuous and dichotomous outcomes

Four studies did not present pain as a continuous outcome and were not included in the primary analyses. Of these, one presented the

percent of participants with no pain on passive movement (Rose et al., 1991), two presented the percent of participants with improvement in pain (Kageyama, 1984, Yamamoto et al., 1979), and one presented the percent of people meeting a response according to OARSI criteria (Barthel et al., 2009).

Sensitivity analysis was conducted by combining the ES derived from the calculated OR for these studies with the ES used for the primary analysis of continuous data. The results are presented in Figure 3-11.

The ES of pain presented as a dichotomous outcome (0.45, 95%CI 0.27 to 0.63) was higher than pain presented as a continuous outcome (0.31, 95%CI 0.20 to 0.41). When all available sources of data were combined (25 trials, n=7,004), the effect estimate was marginally raised to 0.33 (95%CI 0.22 to 0.43).

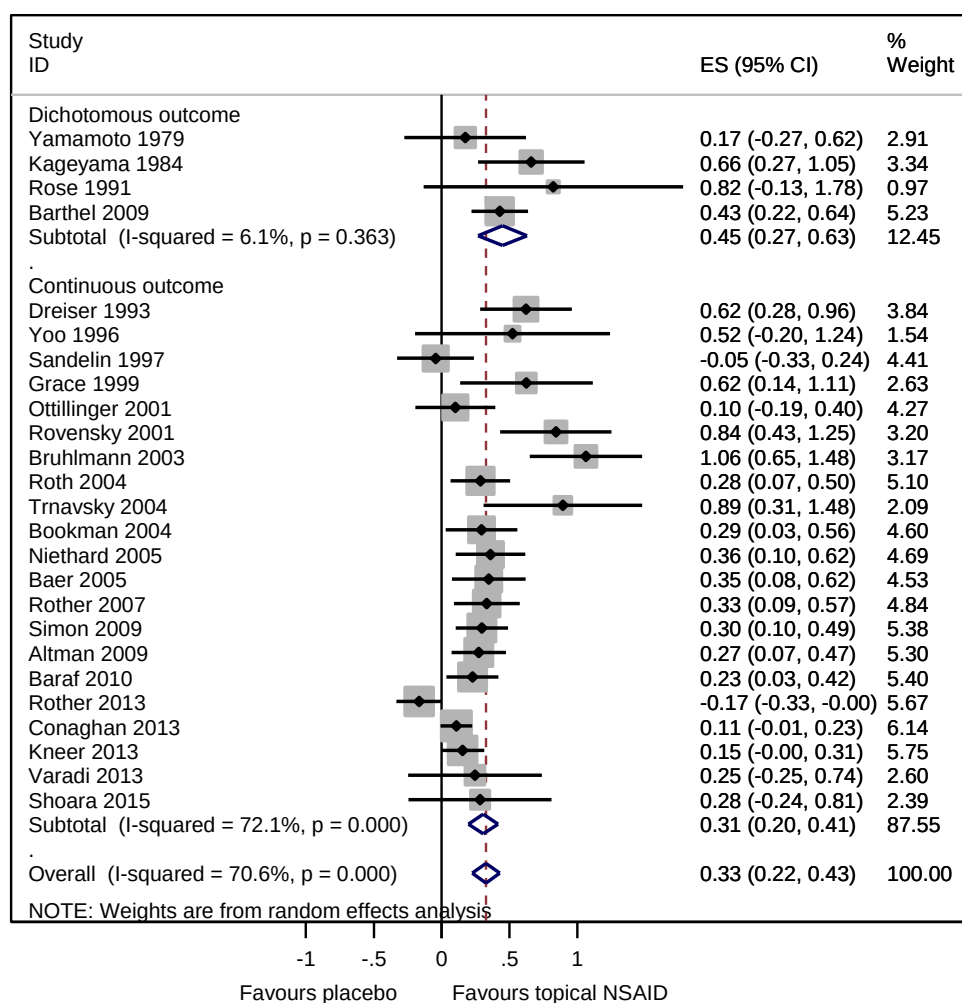


Figure 3-11 - Topical NSAIDs versus placebo in OA. Sensitivity analysis combining continuous and dichotomised pain outcomes at or nearest to four weeks

3.3.2.1.1.4 Subgroup analysis by duration of treatment

Figure 3-12 provides the summary estimates of the ES of topical NSAIDs versus placebo by weeks of treatment. The MAs included different studies at each time point, ranging from ten studies at week one to four studies at week three. Topical NSAIDs were statistically better than placebo in studies with follow-ups at 1, 2, 4, and 9-12 weeks. The effect of topical NSAIDs was highest in studies with follow-up at one week (ES 0.46, 95%CI 0.21 to 0.70).

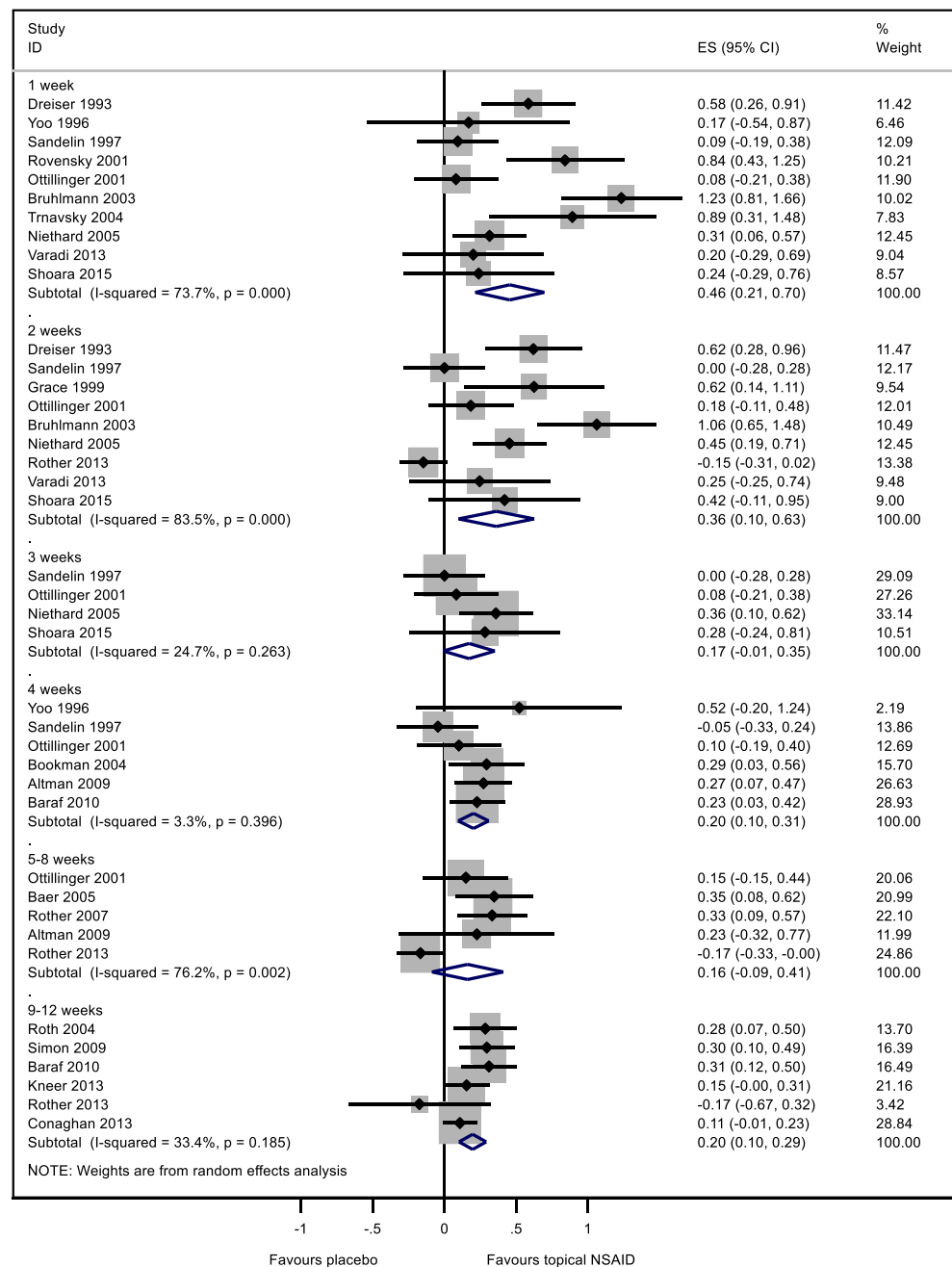


Figure 3-12 - Topical NSAIDs versus placebo in OA. Outcome: pain by weeks of treatment

Examination of the characteristics of the studies included at each time point found that the studies with shorter durations of treatment were generally smaller than those with longer durations of treatment. The median number of participants and interquartile range (IQR) that were analysed at each time point are presented in Table 3-10.

Table 3-10 - Median number of participants analysed per study by duration of treatment

Duration of treatment	Number of participants	
	Median	IQR
1 week	101.5	56.0 to 203.0
2 weeks	142.0	68.0 to 234.0
3 weeks	218.5	129.5 to 235.5
4 weeks	240.5	203.0 to 385.0
5-8 weeks	265.0	234.0 to 385.0
9-12 weeks	512.5	419.0 to 866.0

IQR, interquartile range

3.3.2.1.1.4.1 Meta-regression of sample size and duration of treatment

Meta-regression was conducted to examine the association between sample size, duration of treatment, and treatment effect. The results are presented in Table 3-11.

Table 3-11 - Meta-regression of sample size and duration of treatment versus specific treatment effect

Model/Covariate	Correlation coefficient	p-value
Independent model for each covariate		
Sample size	-0.0005452	0.002
Duration of treatment	-0.023703	0.039
Multivariable model including sample size and duration of treatment		
Sample size	-0.0005709	0.027
Duration of treatment	0.0019207	0.900

Sample size and treatment duration associated with the treatment effect in univariate analyses. In multivariable meta-regression modelling, the association between sample size and treatment effect remained significant whilst the effect of duration did not. Topical NSAIDs exhibited smaller specific effects at longer durations of follow up, however, this association was not seen when adjusting for sample size.

3.3.2.2 Capsaicin versus placebo

3.3.2.2.1 At or nearest to four weeks

Five studies with 415 eligible participants were analysed for the primary outcome of capsaicin versus placebo at or nearest to four weeks of treatment. The results are presented in Figure 3-13.

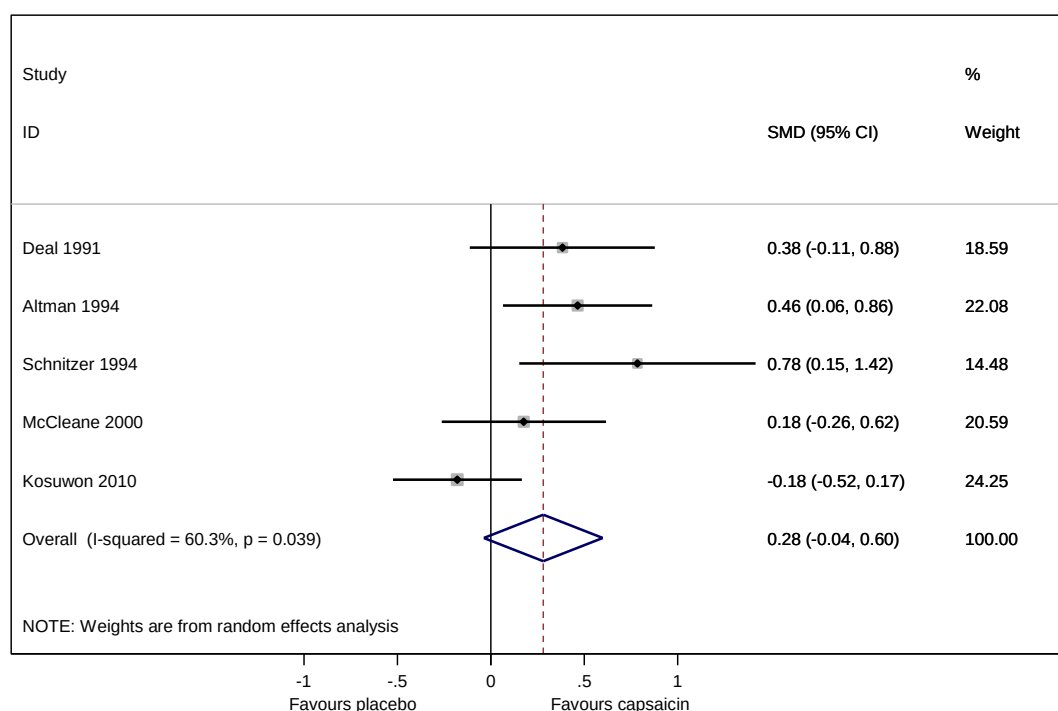


Figure 3-13 - Capsaicin versus placebo in OA. Outcome: Pain at or nearest to four weeks

The ES for capsaicin versus placebo was 0.28 in favour of capsaicin, however, the difference was not statistically significant (95%CI -0.04 to 0.60). This is equivalent to a 0.7 cm (95%CI -0.1 to 1.5) greater decrease in pain for those on capsaicin compared to placebo. The

estimate was associated with substantial and significant heterogeneity ($I^2=60.3\%$, $p=0.039$). The quality of the evidence was rated as very low due to risk of bias, inconsistency, imprecision, and publication bias (Appendix 4 – GRADE assessment: Capsaicin in OA).

3.3.2.2.1.1 Cumulative meta-analysis

The publication years of the trials of capsaicin in OA ranged from 1991 to 2010. The results of the cumulative MA are presented in Figure 3-14.

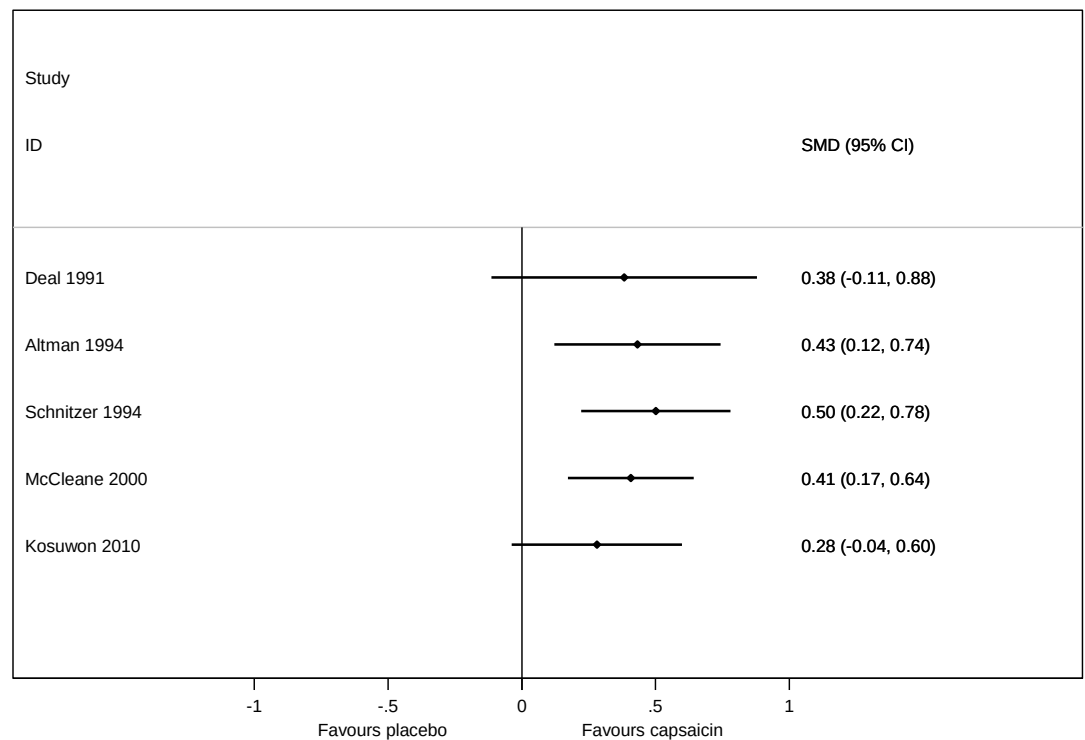


Figure 3-14 - Cumulative MA of capsaicin versus placebo in OA. Outcome: Pain at or nearest to four weeks

The cumulative MA showed that following an initial peak in the ES of capsaicin using the data published up to the mid 1990's (maximum ES 0.50, 95%CI 0.22 to 0.78) the more recent contributions have lowered the ES to a statistically non-significant level. The estimate for capsaicin has remained imprecise throughout all stages of evidence accumulation, and the addition of the most recent trial has further widened the 95% CI.

3.3.2.2.1.2 Publication bias and small study effects

A funnel plot for capsaicin in OA was plotted (Figure 3-15), but further tests to formally examine for funnel plot asymmetry were not conducted due to the low number of trials available (Higgins JPT, 2011). Visual examination of the funnel plot indicates probable funnel plot asymmetry, although the small number of trials available impedes the interpretation.

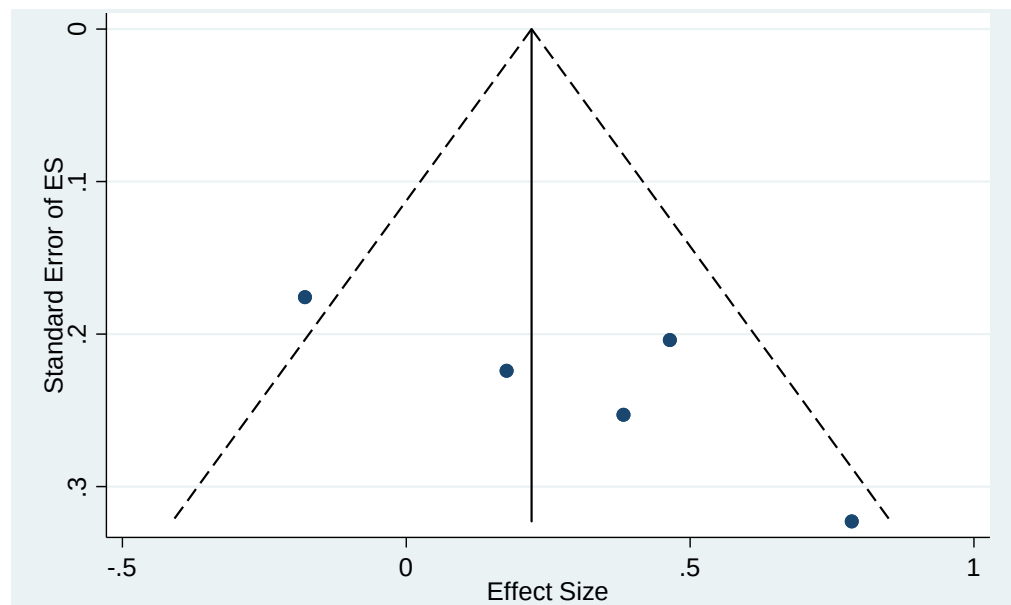


Figure 3-15 - Funnel plot with pseudo 95% confidence limits for capsaicin in OA
ES and SE of ES for capsaicin versus placebo in OA. Outcome: Pain at or nearest to four weeks

3.3.2.2.1.3 Sensitivity analyses

All capsaicin versus placebo trials had less than 100 participants per treatment arm and were found to have unknown or high risk of bias for allocation concealment. It was therefore not possible to conduct sensitivity analyses using only high quality trials.

3.3.2.2.1.4 Subgroup analysis by capsaicin dose

From examination of the trial descriptions it was found that the trial by Kosuwon et al. (2010) differed substantially from the remaining four trials in design, participant characteristics, and the nature of the intervention. Whilst the remaining four studies examined 0.025% capsaicin cream applied four times daily, the study by Kosuwon et al. (2010) used a lower dosing regimen in the intervention group (half the concentration, used three times per day). A subgroup analysis was conducted to examine only studies that used 0.025% capsaicin four times per day. The results are presented in Figure 3-16.

Including only studies using 0.025% capsaicin applied four times per day yielded an ES of 0.41 (95%CI 0.17 to 0.64). The heterogeneity was reduced to 0% ($p=0.473$).

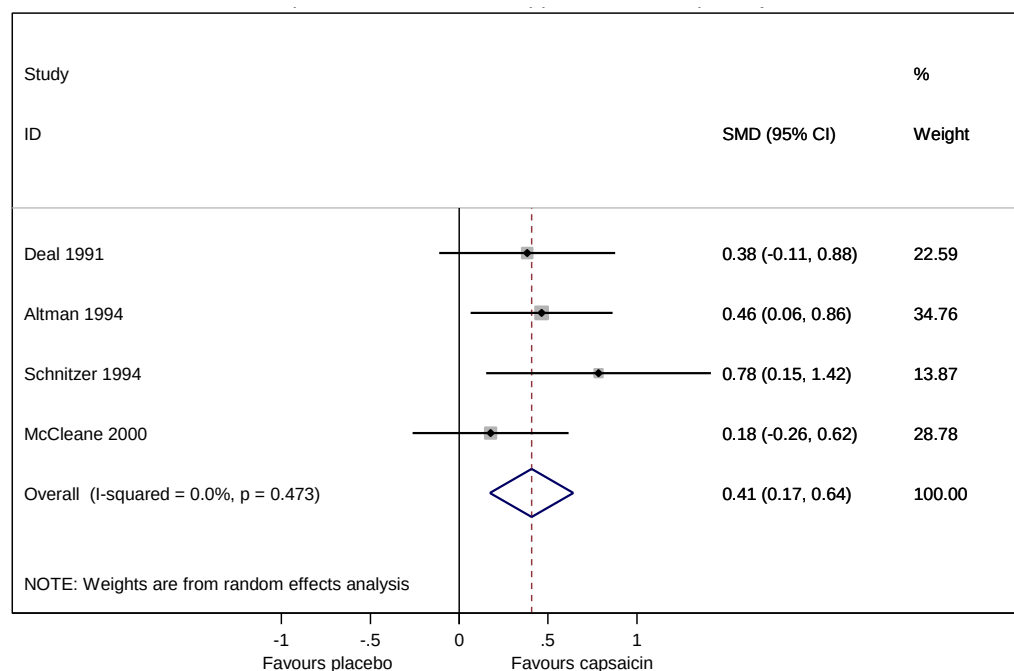


Figure 3-16 - Subgroup analysis: Capsaicin 0.025% four times per day versus placebo in OA. Outcome: Pain at or nearest to four weeks

3.3.2.2.1.5 Subgroup analysis by duration of treatment

The effect of capsaicin was examined at each duration of follow-up available (Figure 3-17). Follow-up durations over four weeks ranged between five and 12 weeks, and were grouped together. The trial by Altman et al. (1994) followed up responders for up to 12 weeks, but non-responders only for four weeks. The comparison at >4 weeks therefore includes participants assessed at weeks four (non-responders), eight (responders), and 12 (responders).

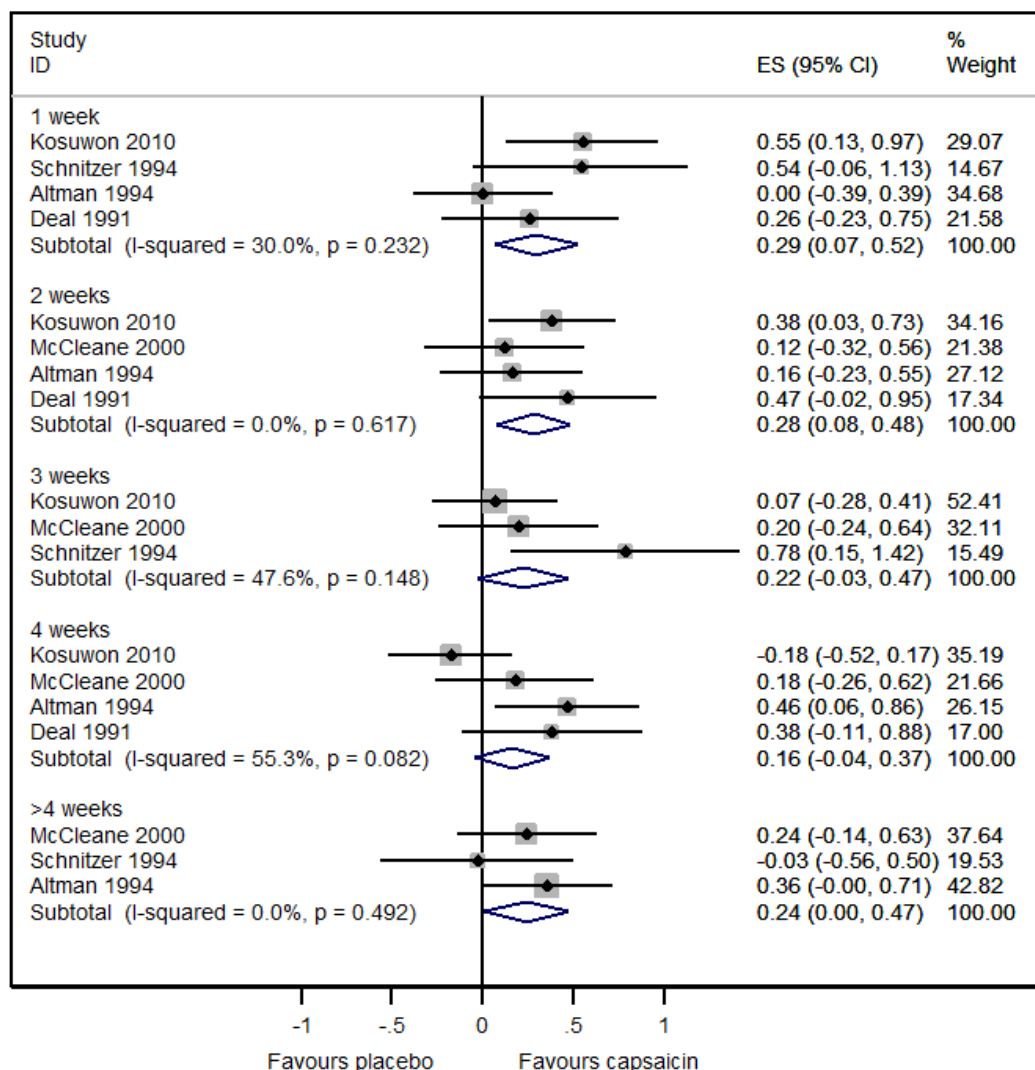


Figure 3-17 - Capsaicin versus placebo in OA. Outcome: Pain by duration of treatment

Capsaicin was statistically significantly better than placebo in the first two weeks of treatment, but not at longer durations of follow-up. The ES was highest at week one (0.29, 95%CI 0.07 to 0.52) and decreased with longer duration of treatment to 0.16 (95%CI -0.04 to 0.37) at four weeks. The ES increased again at durations of follow-up over four weeks (ES 0.24, 95%CI 0.00 to 0.47).

No trend was apparent in the sample sizes across the studies analysed at each time point. The median number and IQR of participants analysed at each time point is presented in Table 3-12.

Table 3-12 - Median number of participants analysed per study by duration of treatment for capsaicin in OA

Duration of treatment	Number of participants	
	Median	IQR
1 week	82.0	55.5 to 101.0
2 weeks	91.0	73.5 to 116.0
3 weeks	80.0	42.0 to 130.0
4 weeks	89.5	72.0 to 114.5
5-8 weeks	80.0	38.0 to 95.0

IQR, interquartile range

3.3.3 Topical NSAIDs versus capsaicin (NMA)

The following section describes the results of the NMA, where topical NSAIDs and capsaicin were compared indirectly using placebo as a common comparator.

3.3.3.1 Description of the trial network

Figure 3-18 shows the treatment comparisons included in the NMA. No trials were available comparing topical NSAIDs directly to capsaicin and no common comparators aside from placebo were available to link topical NSAIDs and capsaicin.

26 trials were included in the primary NMA of topical NSAIDs and capsaicin in OA, of which 21 compared topical NSAIDs to placebo and five compared capsaicin to placebo. The analyses were based on 3,217 participants on placebo, 3,183 on topical NSAIDs, and 206 on capsaicin.

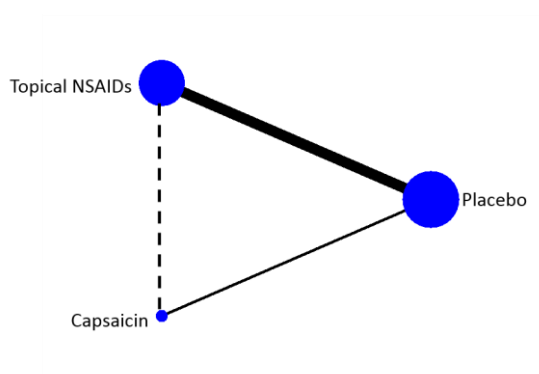


Figure 3-18 - Network diagram for primary NMA in OA

Nodes (circles) are weighted to represent the number of participants using each intervention. The solid lines represent the direct comparisons of the treatments in RCTs. The dotted line represents indirect comparisons generated through the NMA. The lines are weighted to represent the number of comparisons

3.3.3.1.1 Subgroup analysis by recommendation

The subgroup analysis examining topical NSAIDs and capsaicin as recommended in the BNF was based on 16 trials. 12 trials compared diclofenac and ibuprofen at their therapeutic doses to placebo. Four trials compared capsaicin at the recommended dose to placebo. The subgroup analysis is based on 1,576 participants on placebo, 1,198 on topical NSAID, and 141 on capsaicin.

3.3.3.2 Topical NSAIDs versus capsaicin

The results of the frequentist and Bayesian NMAs are presented in Table 3-13. As with the CMA, topical NSAIDs were superior to placebo, whilst there was no significant difference between capsaicin and placebo. The results of the indirect comparison between topical NSAIDs and capsaicin favoured topical NSAIDs (ES 0.04 for topical NSAID), however, the difference was non-significant (95%CI -0.29 to 0.37). The results of the frequentist and Bayesian analyses were in agreement with only minor differences between the confidence and credible intervals. The confidence in the NMA estimate was very low due to imprecision, intransitivity, and very low quality evidence for capsaicin versus placebo (Appendix 5 – GRADE assessment: NMA in OA).

Table 3-13 - Bayesian and frequentist NMAs of topical NSAIDs and capsaicin

Comparison	Type	Bayesian		Frequentist	
		ES	95% CrI	ES	95% CI
Topical NSAIDs vs placebo	Direct	0.31	0.19 to 0.46	0.31	0.19 to 0.44
Capsaicin versus placebo	Direct	0.27	-0.03 to 0.58	0.27	-0.02 to 0.56
Topical NSAIDs vs capsaicin	Indirect	0.04	-0.29 to 0.37	0.04	-0.27 to 0.35

ES, effect size; CI, confidence interval; CrI, credible interval; NSAIDs, non-steroidal anti-inflammatory drugs

In both frequentist and Bayesian analyses, topical NSAIDs had the highest probability of being the best treatment (mean probability 60%). Capsaicin had approximately 40% probability of being the best treatment. Ranking the treatments using frequentist and Bayesian methods (from best to worst), yielded the following order (Bayesian 95%CrI presented): 1. Topical NSAIDs (95%CrI 1-2); 2. Capsaicin (95%CrI 1-3); and 3. Placebo (95%CrI 2-3).

3.3.3.2.1 Subgroup analysis by recommendation

The results of the subgroup analysis, including only trials with topical NSAIDs and capsaicin used as recommended in the BNF, are presented in Table 3-14. At their therapeutic dose, topical NSAIDs and capsaicin were both superior to placebo for pain relief in OA. The indirect estimate comparing topical NSAIDs to capsaicin marginally favoured capsaicin (ES 0.08 for capsaicin), however, the difference was not significant (95%CrI -0.34 to 0.20). Only minor differences were observed between the Bayesian and frequentist results.

Table 3-14 - Subgroup Bayesian and frequentist NMAs of topical NSAIDs and capsaicin used as recommended in the BNF

Comparison	Type	Bayesian		Frequentist	
		ES	95% CrI	ES	95% CI
Topical NSAIDs vs placebo	Direct	0.34	0.25 to 0.45	0.33	0.25 to 0.41
Capsaicin versus placebo	Direct	0.41	0.16 to 0.66	0.41	0.17 to 0.64
Topical NSAIDs vs capsaicin	Indirect	-0.08	-0.34 to 0.20	-0.08	-0.33 to 0.17

CI, confidence interval; CrI, credible interval; ES, effect size; NSAIDs, non-steroidal anti-inflammatory drugs

When comparing topical NSAIDs and capsaicin at their recommended therapeutic dose, capsaicin had the highest probability of being the best treatment (71% in Bayesian and 73% in frequentist analyses). The probability of topical NSAIDs being the best treatment was 29% in Bayesian and 27% in frequentist analyses. Capsaicin was ranked as the best treatment (95%CrI 1-2), topical NSAIDs as the second-best (95%CrI 1-2), and placebo as the worst (95%CrI 3-3).

3.3.4 Overall treatment effect and PCE (CMA)

3.3.4.1 Topical NSAIDs

3.3.4.1.1 At or nearest to four weeks

3,183 participants (21 studies) were analysed for the overall treatment effect of topical NSAIDs at or nearest to four weeks. The overall treatment effect was 1.23 (95%CI 1.06 to 1.41) (Figure 3-19). This is equivalent to approximately 3.1 cm (95%CI 2.7 to 3.5) decrease on a 0-10 cm VAS pain scale. Considerable and significant heterogeneity was evident in the estimate ($I^2=92.3\%$, $p<0.001$).

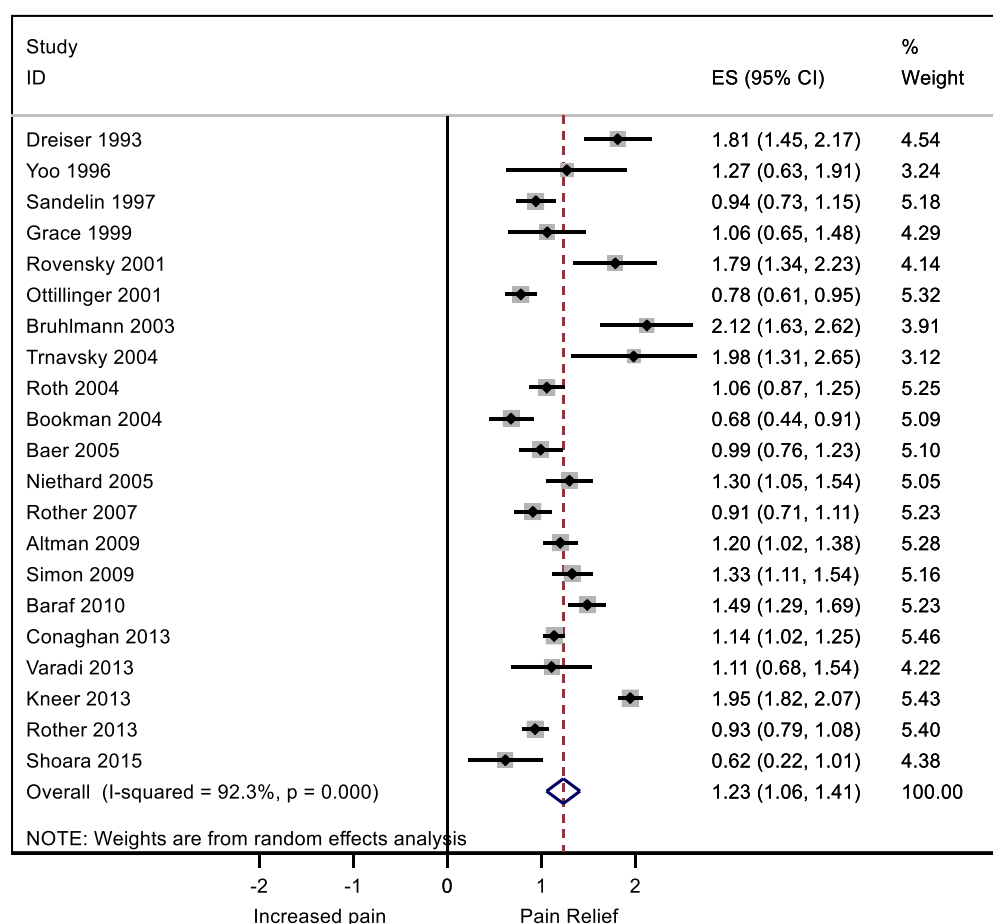


Figure 3-19 - Overall treatment effect of topical NSAIDs in OA. Outcome: Pain at or nearest to four weeks

This overall treatment effect of topical NSAIDs has remained largely unchanged with the addition of new trial data in the last decade (Figure 3-20). The PCE of topical NSAIDs was calculated as 68%.

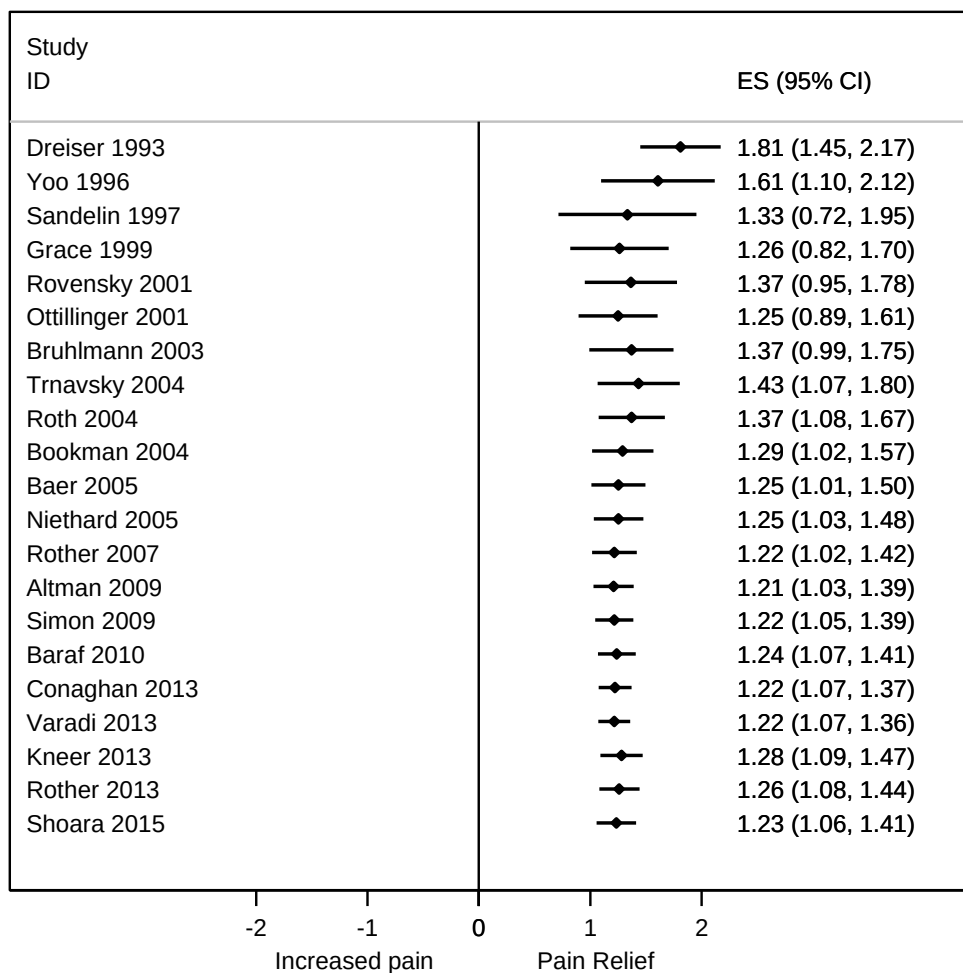


Figure 3-20 - Cumulative MA: Overall treatment effect of topical NSAIDs in OA.
Outcome: Pain at or nearest to four weeks

3.3.4.1.2 Sensitivity analyses

Sensitivity analyses comparing all trials to only high quality trials were conducted. The results are presented in Figure 3-21. The exclusion of lower quality trials did not substantially affect the overall ES of topical NSAIDs at or nearest to four weeks.

The absolute level of contextual effects remained largely unchanged between the primary analysis and the sensitivity analysis of studies with adequate allocation concealment. However, the specific effects were reduced, thereby marginally raising the PCE in studies with adequate allocation concealment (74% compared to 68%). Large studies showed greater absolute contextual effects and smaller specific effects. The PCE was highest in the large studies (79%).

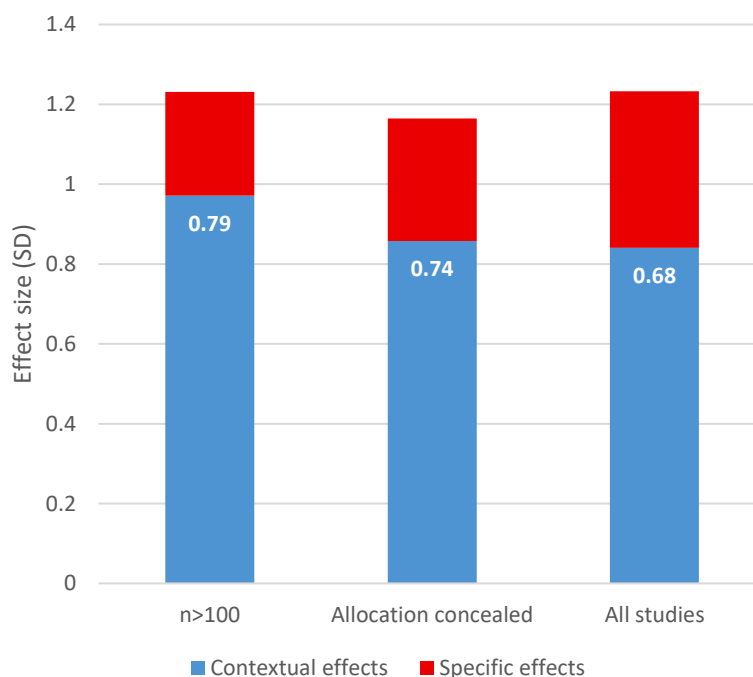


Figure 3-21 – Overall treatment effect (ES) and PCE of topical NSAIDs at or nearest to four weeks in OA. Sensitivity analysis by trial quality

Trial quality indicators used were n>100 (trials with at least 100 participants in the treatment arm) and allocation concealed (trials scoring low risk of bias for allocation concealment)

3.3.4.1.3 Subgroup analysis by drug class

Subgroup analysis of the overall treatment effects and PCE at or nearest to four weeks by drug class are presented in Figure 3-22. Four main drug classes were examined: diclofenac, ketoprofen, eltenac, and ibuprofen. The diclofenac class included diclofenac epolamine, diclofenac diethylamine, diclofenac sodium, and diclofenac hydroxyethylpyrrolidine. The overall treatment effects and PCE of diclofenac and ketoprofen were comparable. The results are based on 2,661 and 2,879 participants, respectively (total number analysed in topical NSAID and placebo arms). Only two trials were conducted using eltenac, which showed the smallest overall and specific effects. Ibuprofen showed the largest overall and specific effects, but this was only based on 214 participants.

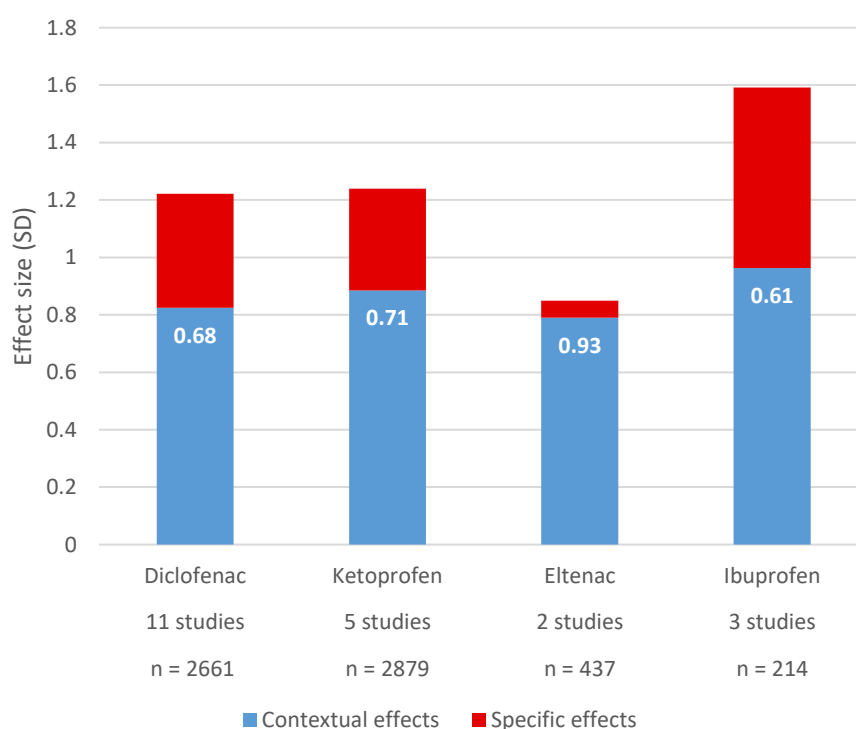


Figure 3-22 – Overall treatment effect (ES) and PCE of topical NSAIDs at or nearest to four weeks in OA. Subgroup analysis by drug class
n, number of participants; SD, standard deviation

3.3.4.1.4 Subgroup analysis by duration of treatment

The overall treatment effect and PCE by weeks of treatment are presented in Figure 3-23.

The PCE varied according to the time-point studied with the largest specific effects, and corresponding smallest PCE, found in the first two weeks of treatment. At or over three weeks of treatment, the overall treatment effect varied from 0.85 to 1.34, but the PCE remained constant at 80-82%.

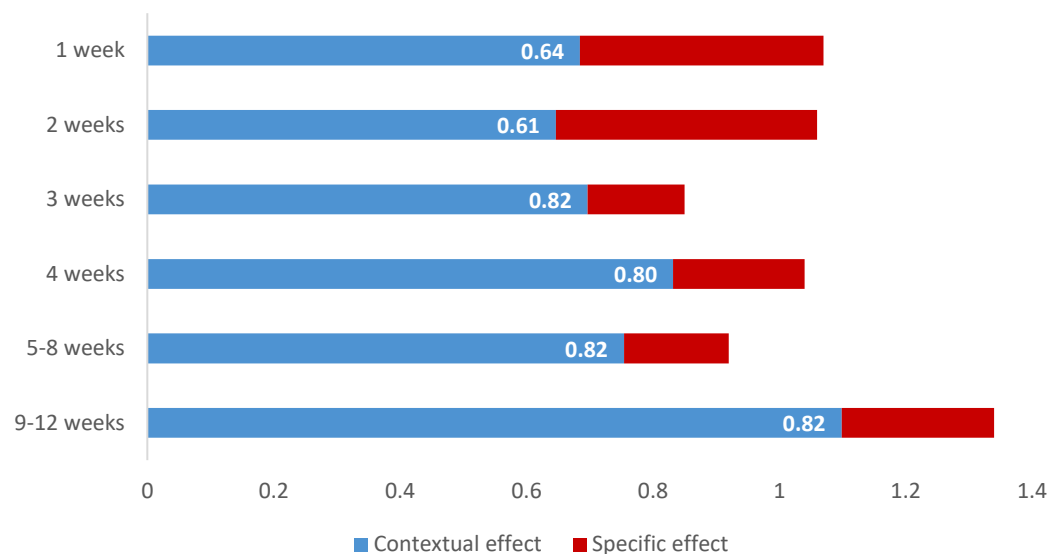


**Figure 3-23 - Overall treatment effect (ES) and PCE of topical NSAIDs in OA.
Outcome: Pain by weeks of treatment**

3.3.4.2 Capsaicin

3.3.4.2.1 At or nearest to four weeks

Five studies (206 participants) were analysed for the overall ES of capsaicin at or nearest to four weeks. The overall ES was 1.05 (95%CI 0.52 to 1.57) (Figure 3-24). The effect estimate is equivalent to approximately 2.6 cm (95%CI 1.3 to 3.9) decrease from baseline on a 0-10 cm VAS pain scale. The PCE of capsaicin was 56%.

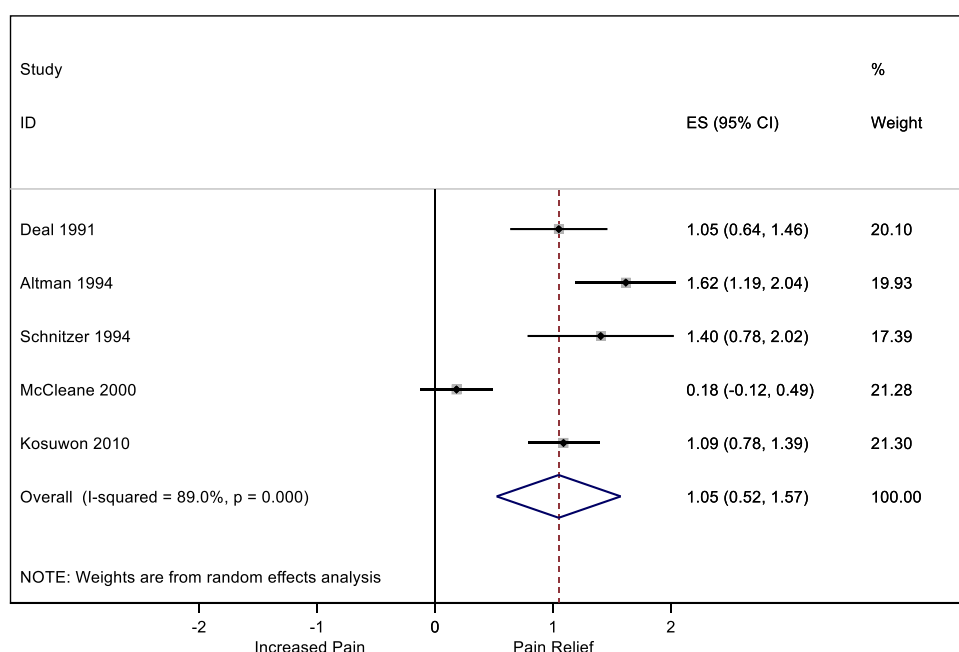


Figure 3-24 - Overall treatment effect of capsaicin in OA. Outcome: Pain at or nearest to four weeks

3.3.4.2.2 Sensitivity and subgroup analyses

No high quality trials were available for sensitivity analyses.

Subgroup analysis by capsaicin concentration found that the overall ES of 0.025% capsaicin was 1.04 (95%CI 0.33 to 1.76). The PCE was 46%.

Results of the overall treatment effect and PCE of capsaicin by duration of treatment are presented in Figure 3-25. Capsaicin produced statistically significant decreases in pain at all treatment time-points assessed in studies. The overall treatment effect of capsaicin was highest at three weeks of treatment where it was comprised of 63%

contextual effects and 37% specific effects. The PCE varied at different time-points from 61% to 72%.

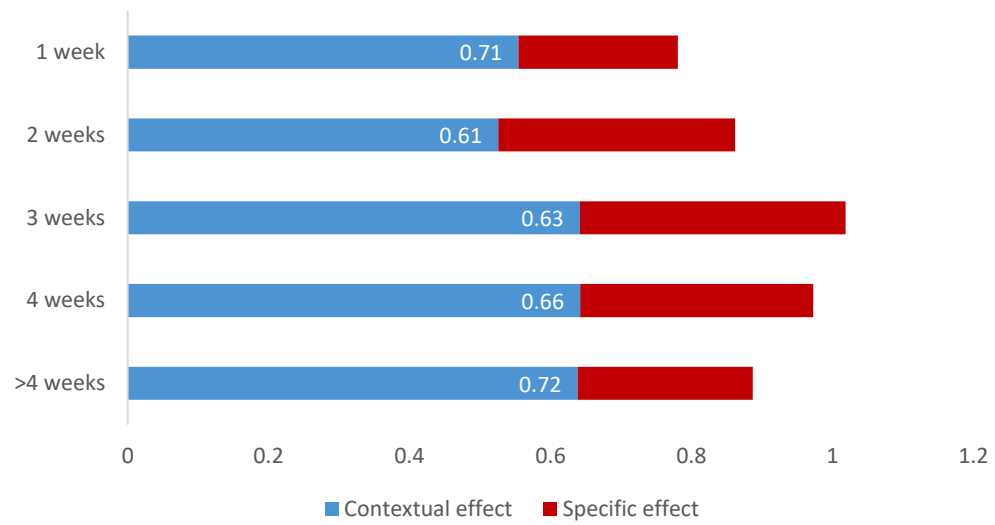


Figure 3-25 - Overall treatment effect (ES) and PCE of capsaicin in OA. Outcome: Pain by weeks of treatment

3.3.4.3 Topical NSAIDs relative to capsaicin

The overall ES and PCE at or nearest to four weeks for topical NSAIDs and capsaicin are presented in Figure 3-26.

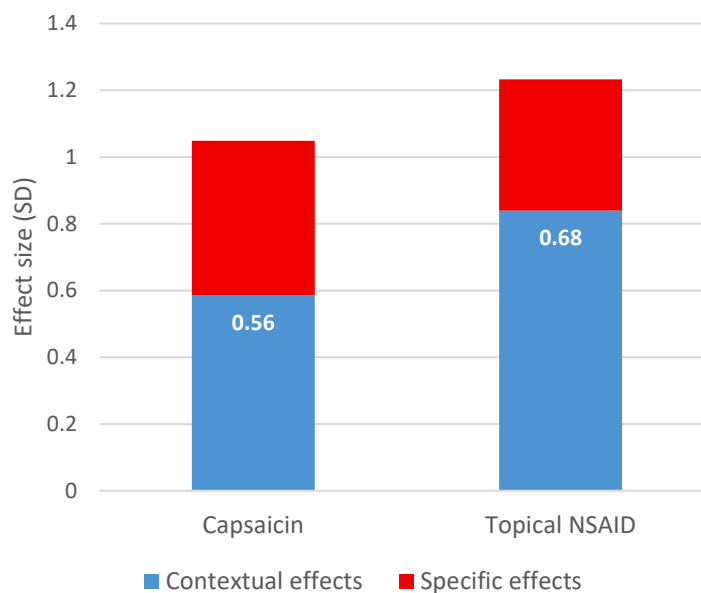


Figure 3-26 - Overall treatment effect (ES) and PCE of topical NSAIDs and capsaicin in OA. Outcome: Pain at or nearest to four weeks
NSAID, non-steroidal anti-inflammatory drug; SD, standard deviation

3.4 Summary of findings

Comparison	SMD treatment effect	0-10 cm VAS treatment effect	n (trials)	Confidence in effect estimate
Topical NSAIDs vs placebo	On average, pain scores in the topical NSAID group were 0.31 SDs (0.20 to 0.41) lower than in the placebo group	On average, pain scores in the topical NSAID group were 0.8 cm (0.5 to 1.0) lower than in the placebo group	6,191 (21)	⊕⊕⊕ ○ Moderate for inconsistency
Capsaicin vs placebo	On average, pain scores in the capsaicin group were 0.28 SDs lower (0.04 higher to 0.60 lower) than in the placebo group	On average, pain scores in the capsaicin group were 0.7 cm lower (0.1 higher to 1.5 lower) than in the placebo group	415 (5)	⊕○○○ Very Low for risk of bias, inconsistency, imprecision, publication bias
Topical NSAIDs vs capsaicin	On average, pain scores in the topical NSAID group were 0.04 SDs lower (0.29 higher to 0.37 lower) than in the capsaicin group	On average, pain scores in the topical NSAID group were 0.1 cm lower (0.7 higher to 0.9 lower) than in the capsaicin group	6,606 (26)	⊕○○○ Very Low as per lowest direct estimate, imprecision, intransitivity
Topical NSAIDs overall ES	On average, those in the topical NSAID group experienced 1.23 SDs (1.06 to 1.41) pain relief compared to baseline	On average, those in the topical NSAID group experienced 3.1 cm (2.7 to 3.5) pain relief compared to baseline	6,191 (21)	⊕○○○ Very Low for study type and inconsistency
Capsaicin overall ES	On average, those in the capsaicin group experienced 1.05 SDs (0.52 to 1.57) pain relief compared to baseline	On average, those in the topical NSAID group experienced 2.6 cm (1.3 to 3.9) pain relief compared to baseline	415 (5)	⊕○○○ Very Low for study type, risk of bias, inconsistency, publication bias

3.5 Discussion

3.5.1 Key findings and interpretation

3.5.1.1 Topical NSAIDs in OA

Moderate quality evidence suggests that topical NSAIDs are superior to placebo for pain relief at or nearest to four weeks. These findings are reflected also in high quality studies and in sensitivity analyses that combined continuous and dichotomous pain outcomes. This is keeping with the literature, which has consistently shown a greater effect for topical NSAIDs than placebo (Moore et al., 1998, Mason et al., 2004, Lin et al., 2004, Biswal et al., 2006, Bjordal et al., 2007, Zhang et al., 2010b, Derry S, 2012, Zeng et al., 2018). However, it must be noted that although the specific ES is statistically significant, it fails to reach the threshold for a minimum clinically important difference (MCID; ES 0.5) set by NICE (2014).

On examination of the topical NSAID arm only, the mean level of pain relief experienced by participants was approximately 3.1 cm on 0-10 cm scale. This is likely clinically significant, however, it is marginally lower than that reported by Zou et al. (2016) (3.4 cm). These differences may be explained by differences in the study population (current MA: 21 studies, 6,191 participants; Zou et al: 20 studies, 4,399) or calculation of the overall ES (differences in correlation coefficients). However, the discrepancy is most likely explained by differences in the outcome analysed. The present MA used pain data at or nearest to four weeks, whilst Zou et al. (2016) systematically chose the time point where the treatment group had its largest effect, thereby inflating the overall treatment effect.

The primary, *a priori* defined time point used was four weeks, but the effects of topical NSAIDs at all other durations of follow-up were also examined. As with previous literature (Lin et al., 2004, Derry S, 2012, Derry et al., 2016), topical NSAIDs exhibited the largest specific effect at shorter durations of follow-up. However, the effect of topical NSAIDs was maintained beyond four weeks. This is reflected in the overall ES,

where the PCE was lowest in the first two weeks of treatment, and then remained largely unchanged at longer durations of follow up. This may be due to wear-down effects, although it is likely an artefact of small study effects. In the present MA, studies conducting participant follow-ups at shorter durations were generally smaller in size. Multivariable meta-regression indicated that the association between duration and treatment effect was likely explained by the fact that smaller studies (with larger treatment effects) reported outcome data at shorter durations of follow-up. The observed trend for decreasing efficacy of topical NSAIDs over time may therefore be explained by small study effects, rather than a biological mechanism.

Although generally considered as a single treatment group, differences were observed in the overall ES and PCE between different drugs. Differences in effects between topical NSAIDs have previously been examined in a NMA by Zeng et al. (2018). Comparisons are somewhat limited as the NMA analysed each diclofenac formulation separately, whilst they were all pooled in the current analysis. In addition, comparisons between overall ES subgroup estimates should be done with caution as the analyses disregard randomisation. In the current work, the largest specific and overall ES and lowest PCE was found for ibuprofen. This is in keeping with the work by Zeng et al. (2018) who found ibuprofen to be the second best treatment for pain relief, behind the diclofenac patch. However, only a few studies with small numbers of participants were analysed for ibuprofen. This has the potential to overestimate the treatment effect (Nüesch et al., 2010). In contrast, over ten times more participants were included in the analyses for ketoprofen and diclofenac, perhaps partly explaining the apparent differences in effect. Finally, the effect of eltenac was almost wholly explained by contextual factors, indicating that stark differences in effect may occur between topical NSAID types. This may beg the question of whether or not pooling all topical NSAIDs is appropriate. However, exclusion of topical NSAID types, concentrations, and formulations not currently recommended by the BNF did not significantly alter the specific effect (ES: 0.31 for all versus 0.33 for recommended drugs).

3.5.1.2 Capsaicin in OA

Very low quality evidence indicates that capsaicin cream may not be statistically superior to placebo for pain relief at or nearest to four weeks. However, subgroup analysis indicated that capsaicin may be statistically superior to placebo when used at 0.025% concentration. Findings could not be verified in sensitivity analyses of high quality trials (no high quality trials were available). Neither primary nor subgroup analyses for specific ES met the threshold for clinically significant pain relief (ES 0.5) (NICE, 2014).

These findings contrast with previous evidence that capsaicin is effective in OA (Zhang and Li Wan Po, 1994, Mason et al., 2004, Cameron et al., 2009, Laslett and Jones, 2014). Discrepancies in the inclusion and analysis of the trial by Kosuwon et al. (2010) likely explain these differences. This cross-over trial found that, in the first period, participants on placebo experienced more pain relief at four weeks of treatment compared to those on 0.0125% capsaicin (ES favours placebo). In the second period and when the two periods were pooled, more pain relief was experienced using capsaicin cream than placebo (ES favours capsaicin). However, washout duration was one week and pain levels did not return to baseline level following washout. Using the combined cross-over data may therefore not have been appropriate. Published in 2012, this trial was not included in any of the earlier MAs. However, it was included by Laslett and Jones (2014), but they used the estimate from the combined treatment periods. Inclusion of the data from the first treatment period, as done in the present MA, explains why this is the first MA to find that capsaicin was not significantly superior to placebo.

The specific ES of capsaicin was greatest at one and two weeks' follow-up, where it was statistically superior to placebo, and decreased with longer durations of follow-up. The specific ES increased again after four weeks, however, this estimate is likely biased by the comparison conducted in the trial by Altman et al. (1994). Responders were followed up for up to 12 weeks, whilst non-responders were only followed up for four weeks. The comparison at >4 weeks for this study

therefore contains data at four (non-responders), eight (responders), and 12 (responders) weeks. Only one other MA has compared the effect of capsaicin at different durations of follow-up, although this was limited to weeks three and four (Cameron et al., 2009). No trend in effect was observed between the two time points.

Examination of the treatment arm alone indicated that capsaicin cream may offer clinically meaningful pain reduction overall. Only moderate contributions from contextual effects were found, however, interpretation is limited by the inadequacy of blinding.

3.5.1.3 Topical NSAIDs versus capsaicin

Very low quality evidence indicated that, at the study level, topical NSAIDs and capsaicin provide equal levels of pain relief at or nearest to four weeks. These findings are reflected in the subgroup analysis using only recommended doses/formulations of the drugs. This subgroup analysis was conducted to more closely reflect clinical practice and render the results more relevant to clinicians. Similarly, comparison of the overall ES indicate that both treatments may provide similar levels of overall pain relief.

No direct or indirect quantitative evidence for the relative efficacy of topical NSAIDs and capsaicin has been published previously. However, guidelines and reviews in the area may provide some clues regarding the perceived relative efficacy of the two drugs in OA.

On one hand, the OARSI and (knee-specific) EULAR guidelines provide equal recommendations for the two treatments, perhaps indicating a perceived equivalence in efficacy (McAlindon et al., 2014, Jordan et al., 2003). These findings are in keeping with the current NMA.

In contrast, other evidence suggests a hypothesised greater effect for topical NSAIDs. A narrative review examining topical treatments in OA concluded that topical NSAIDs were more effective than capsaicin (Altman and Barthel, 2011). However, the postulated superiority of topical NSAIDs was primarily based on their mechanism of action, rather than quantitative analysis. Capsaicin was thought to be less effective as it lacked significant tissue penetration and anti-

inflammatory effects. NICE, ACR, and (hand-specific) EULAR guidelines also favour topical NSAIDs, indicating a postulated greater efficacy for topical NSAIDs (NICE, 2014, Hochberg et al., 2012, Kloppenburg et al., 2018). Similarly, OARSI granted topical NSAIDs a greater mean benefit score (6.0/10) than capsaicin (5.1/10) (McAlindon et al., 2014).

However, guideline decisions are based not only on the perceived treatment efficacy, but on the *quality* of evidence. This is evidenced in the EULAR recommendations for hand OA, which list topical NSAIDs as the first line topical treatment, but do not provide a recommendation for capsaicin as they judge the blinding questionable (Kloppenburg et al., 2018). The preference for topical NSAIDs in the majority of guidelines may therefore reflect a greater confidence in the evidence, rather than a perception of a larger effect for topical NSAIDs. This is in line with the very low confidence established for capsaicin in the current MA.

3.5.2 Caveats

The present MAs are subject to several limitations. Firstly, the conclusions are limited by the scarcity of data available on capsaicin in OA. Only four trials examined capsaicin at its licensed dose and no trials directly compared topical NSAIDs to capsaicin. A large trial for capsaicin in OA has been published (Schnitzer et al., 2012), but was not included in the present MA as it was not placebo-controlled (control was 0.01% capsaicin). A large degree of uncertainty was therefore apparent in the effect estimate for capsaicin. The lack of significance of capsaicin overall and the equivalence of the two drugs may therefore be an artefact of the wide CIs.

Secondly, concerns over study quality were identified in the risk of bias assessments. Substantial risks of bias were identified particularly for the domains of complete outcome data reporting, use of ITT, and pre-specification of published outcomes. Risks of bias across these domains have the potential to overestimate the results of the MAs. Sensitivity analyses for topical NSAIDs indicated that the findings were unlikely to be biased by study quality, however, such analyses were not possible for capsaicin. Overall, moderate quality evidence was available

for topical NSAIDs, indicating that the true effect is likely to be close to the observed estimate, although there is a possibility that it is substantially different (Balslem et al., 2011). In contrast, only very low quality evidence was available for capsaicin. There is very little confidence in the observed estimate, which is likely to be substantially different from the true effect (Balslem et al., 2011).

Thirdly, due to inconsistent and incomplete outcome reporting, it was not possible to analyse all published evidence in the MAs. Asymmetry on visual inspection of the funnel plots indicated that this loss of data may have biased the results. However, efforts were made to transform data in order to avoid systematic losses of evidence.

Fourthly, because of the difficulty associated with blinding capsaicin, it was deemed a high risk of bias domain for all capsaicin trials. This may result in inherent differences in the placebo group across the NMA network, threatening the assumption of transitivity. In addition, although topical NSAID trials predominantly recruited individuals with knee OA, capsaicin trials enrolled participants with OA at sites such as the hand, wrist, elbow, shoulder, hip, and knee. Such differences in the study population may limit comparisons between the two treatments.

Finally, these study-level findings relate to populations of individuals with OA and may not be reflected at the individual patient level. In addition, it was not possible to examine treatment efficacy in distinct OA subgroups as these data were not available in the publications. No analyses could therefore be conducted for predictors of response. Analyses conducted at the individual patient level would be required for this.

3.5.3 Methodological comments

3.5.3.1 Estimation of specific treatment effects

The presentation of the specific treatment effects using SMD is subject to recognised limitations. Firstly, results presented as SMDs may be difficult to understand, and the results were therefore interpreted using accepted conventions (small, moderate, large effect) and back-translated to natural units of the most familiar instrument (0-10 cm

VAS) (Guyatt et al., 2013, Wandel et al., 2010). Additionally, SMDs are determined by the heterogeneity of the population in the studies. Assuming the same absolute MD, the calculated SMD is larger if the population is homogenous (small SD) and smaller if the population is heterogeneous (large SD) (Guyatt et al., 2013). Alternate approaches to the SMD have been suggested by the GRADE working group (Guyatt et al., 2013). These include using the WMD (back-translating outcomes to the most commonly used tool within studies before pooling) and conversion into relative and absolute effects (Guyatt et al., 2013). Examination of how these alternate approaches affect the interpretation and clinical significance of the findings may be warranted.

Clinical significance, at the study-level, was determined using the threshold of 0.5 SD established in NICE (2014) guidelines for OA. This threshold was chosen empirically and may be unrealistic as the effect of very few interventions are likely to be wholly above the threshold (NICE, 2014). In addition, the application of MCID thresholds to study-level data regarding the difference between treatments is subject to conceptual limitations. The MCID is based on pre- and post-intervention differences within an individual and its application to study-level differences between treatments may be misleading (Katz et al., 2015b). Using the overall ES and PCE may go some way to overcome this caveat as they allow examination of the pre- and post-intervention scores within the trial. Application of the MCID to the overall ES may therefore be more intuitive. However, it does not overcome the caveat of applying individual-level thresholds to study-level averages. A suggested approach to overcome this caveat is instead to compare the proportion of individuals that met the MCID threshold, however, this data is rarely available in trial publications. OMERACT-OARSI responder criteria have been implemented in some trials in OA (Pham et al., 2004), however, they are not outcome- or tool-specific. Standardising outcome reporting in OA trials to include the proportions meeting MCID may allow the generation of more interpretable and clinically relevant outcomes in MAs.

3.5.3.2 Examination of duration of treatment and treatment effect

Examination of treatment effects at different durations of treatment may not be appropriate in MA settings. Trends indicating decreased specific effects at longer durations of treatment for both topical NSAIDs and capsaicin were observed, however, different trials were analysed at each time point. It is therefore not possible to determine if these differences in ES are due to time or other differences between trials. For example, the effect of duration on the treatment effect of topical NSAIDs was shown to be mediated by sample size. Small study effects have been clearly documented in OA trials (Nüesch et al., 2010), however, future work needs to be conducted to determine whether it is possible to examine whether:

- a) the association between duration and sample size is reflected across different analgesics in OA and
- b) this precludes the examination of the treatment effect across different durations of follow-up in MAs.

3.5.3.3 Analysis of overall treatment effect and PCE

Examination of the overall treatment effect renders effectiveness data that is valuable for shared decision making. When making treatment decisions, an evidence-based response to “how much pain relief am I likely to achieve when I start using the treatment” may be more appreciated by a patient than an explanation of the difference between treatment and placebo. The overall treatment effect aims to address the apparent gap between research evidence and clinical findings (Zhang and Doherty, 2018).

However, analyses for the overall treatment effect are conducted within one treatment arm. The data is rendered observational in nature as no randomisation remains within the comparison. As such, the overall ES estimates should be interpreted with caution. Comparisons of overall ES estimates between treatments or subgroups do not control for confounding factors at the individual level (non-randomised comparisons) and are further influenced at the study-level by

differences in study populations, contextual factors, small study effects, etc.

Similarly, the PCE is a recently developed measure that increases the interpretability and clinical relevance of treatment effect output. When combined with the overall treatment effect, it allows the effect of a treatment to be easily understood as it relates to contextual and specific effects. An underlying assumptions required for its interpretation is that the contextual and specific treatment effects are clearly distinct and can be quantified accordingly. However, the current work with capsaicin raises the question whether criteria for the implementation of PCE need to be determined. For example, in such a setting where the intervention is not adequately blinded, the specific effects and contextual effects cannot be clearly distinguished in analyses. In these cases, the specific effect is likely comprised also of contextual effects, hindering the PCE estimation. The second setting in which the calculation of PCE may not be appropriate is where the distinction between placebo and nocebo effects is unclear. Whereas the term placebo effect is used to describe the beneficial contextual effects of a treatment, the term nocebo is used to describe these effects when they are harmful (Häuser et al., 2012). The determination of PCE explicitly excludes data where participants worsened from baseline under the assumption that this represents nocebo effects (Zou et al., 2016). This complicates the examination of PCE for a treatment such as capsaicin whose burning effect may be associated with increased placebo effects in some participants and increased nocebo effects in others. The current work therefore indicates that consideration of blinding and potential nocebo effects may be required in determining the appropriateness of the calculation of PCE.

4 CHAPTER 4: Predictors of response to topical NSAIDs and capsaicin in OA

4.1 Overview

Aggregate data MAs allowed examination of the average treatment effects of topical NSAIDs and capsaicin in OA. With the aim of guiding precision medicine, a movement away from study-level averages and to participant-level results is required.

The following chapter aims to identify predictors of response to topical NSAIDs and capsaicin in OA. It initially describes the methods for data collection and analysis. It then describes the data acquired, study population, and results of the IPD MA. It was only possible to conduct an IPD MA for topical NSAIDs and the reasons why an IPD MA could not be conducted for capsaicin are outlined in the chapter.

4.2 Methods

4.2.1 Data collection and preparation for analysis

Trial descriptions were extracted from the published reports of the included trials using methods described in Section 2.1.3.1. Similarly, risk of bias assessments were conducted as specified in Section 2.1.3.3. The collection and preparation of IPD are described below.

4.2.1.1 Data acquisition from trial custodians

Essential to the conduct of an IPD MA is the acquisition of participant-level data from trial data custodians. Data custodians may be the corresponding author or the pharmaceutical funder/sponsor of the trial. Unlike undertaking a CMA, where the data are readily available in published reports, conducting an IPD MA is dependent on forming successful collaborations with data custodians.

Eligible RCTs of topical NSAIDs and capsaicin were identified, as described in 2.1.1. All trials that were eligible for the CMA were eligible for the IPD MA. In addition, active-controlled trials, trials where data were not available for extraction from the published reports, and trials

with subpopulations of non-eligible participants were eligible for the IPD MA.

The methods for contacting authors of eligible publications are described in the following section (section 4.2.1.1.1). Pharmaceutical companies were contacted regarding identified trials and were also asked about the presence of unpublished data. The methods for this are described in section 4.2.1.1.2.

The IPD MA in OA was conducted in collaboration with the OA Trial Bank, an international consortium aiming to develop an IPD databank for OA to identify treatment-responsive subgroups. Participant-level data for OA trials were therefore shared with the OA Trial Bank. Where the data custodian declined collaboration with the OA Trial Bank or the trial was for NP, the collaboration was formed only between the data custodian and the University of Nottingham (UoN).

4.2.1.1.1 Contacting authors

The contact details for the first or corresponding author of all eligible trials were extracted from the publication or found through internet searching. Authors were contacted via email using a pre-specified disease-specific template, personalised to include the author name and study details. The study information sheet is available in Appendix 8 – Contact letter to authors. If the email address of the first or corresponding author was not found or was incorrect, the last author was emailed. If this was not found, contact details for all other authors were sought and used.

If no response was received, a second reminder email was sent to the author. If this did not result in a response, all other study authors were contacted via email and a second reminder email (i.e., third email) was sent to the first/corresponding author. If no response was received, the department where the research was carried out was emailed (Appendix 9 – Contact email to institutions). Final efforts were then made to contact the lead/corresponding author via postal letter and telephone. Where an author responded and recommended contacting

another author or the pharmaceutical sponsor, these recommendations were followed.

4.2.1.1.2 Contacting pharmaceutical companies

Contacting pharmaceutical companies took place via telephone calls, email, and online data sharing websites. If a company had pre-specified methods for requesting IPD, such as through the clinicaltrialdatasharing.gov website, this was followed. If no methods were pre-specified, attempts were made to contact the responsible department within the pharmaceutical company, either the data-sharing, medical information, or clinical trials team.

4.2.1.1.2.1 For published trials

The pharmaceutical sponsor of included trials were contacted directly using a pre-specified email in parallel with the second reminder email to the authors or in response to the authors stating that the data were held by the sponsor. If the company had been acquired or merged with another pharmaceutical company, the newest company was contacted to request the data.

4.2.1.1.2.2 For unpublished trials

To assess for the presence of unpublished data, all pharmaceutical companies that manufacture topical NSAIDs or capsaicin in the UK (identified through the BNF or eMC) or that have registered trials for the medications in OA or NP (identified through clinicaltrials.gov or clinicaltrialsregister.eu) were contacted. They were asked if any clinical trials had been conducted and if they were interested in collaborating on this IPD MA (Appendix 10 – Contact letter to pharmaceutical companies).

4.2.1.2 Data transfer

4.2.1.2.1 Data transfer agreements

Prior to collaboration, data custodians were required to sign a data transfer agreement (DTA) outlining the terms of the collaboration and transfer of the data. Data custodians of OA trials were sent the OA Trial

Bank's pre-specified DTA (Appendix 11 – OA Trial Bank DTA). DTA issues were resolved over email by liaising between the potential collaborator and the Erasmus MC legal department (through the OA Trial Bank). Where authors did not agree with the OA Trial Bank's DTA and the issues could not be resolved, they were invited to collaborate directly with the UoN instead. A DTA drafted by the legal department at UoN was used for this purpose (Appendix 12 – UoN drafted DTA).

Once finalised, the DTA was signed by the collaborating author and myself (on behalf of the OA Trial Bank) or the UoN legal department (for UoN-only collaborations). Once signed, the process of data collection commenced.

4.2.1.2.2 Data sharing procedures

On agreeing to collaborate, data custodians were asked to specify the variables measured in the study using an electronic form (Appendix 13 – List of variables from IPD MA trials). They were given the choice of sharing the whole database or only the variables pre-specified in the IPD MA protocol. The whole database was preferred, following recommendations by members of the OA Trial Bank, as it ensured that all necessary variables were received in the first instance. This was particularly convenient for predictors that could be assessed using various approaches, such as inflammation being assessed using US, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), or clinically.

Data custodians were asked for their favoured method for secure data transfer. The preferred method for receiving data was via email, however, other approaches were accepted in order to facilitate the process for collaborating authors.

On receiving the data, it was stored in a restricted-access folder on the University network. Once the datasheets were cleaned and merged, a final copy of the data were transferred to the OA Trial Bank.

4.2.1.3 Data clean and amalgamation

The process of preparing the data for analysis is illustrated in Figure 4-1.

An initial screen was conducted of the datasets to ensure that data for all participants had been received. This involved comparing the following numbers with those presented in the publication: total number of participants, the number of participants in the intervention group, the number of men/women, and the mean age. Any discrepancies were discussed with the collaborating author. The original data were then re-labelled and re-coded in order to standardise all studies for amalgamation and analysis.

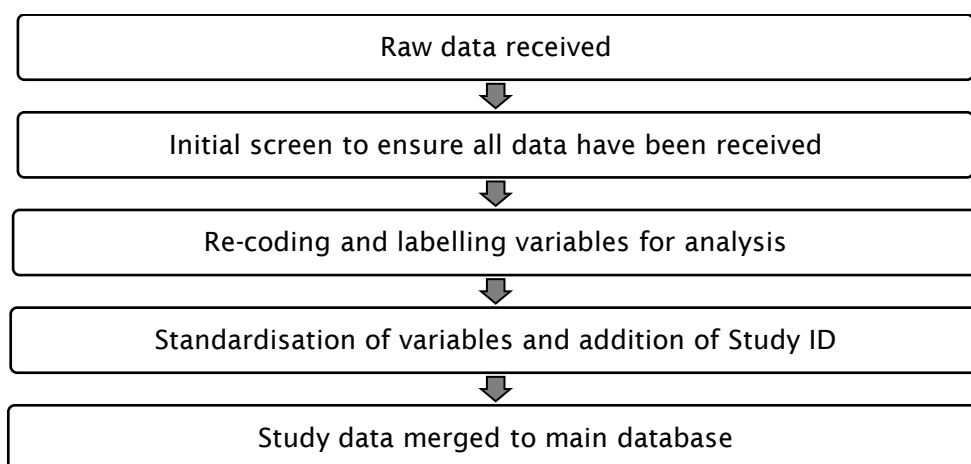


Figure 4-1 - Process of preparing original study data for merging and analysis

A database was created to hold all the variables required for analysis of the IPD MA, as listed in Appendix 13 – List of variables from IPD MA trials. A consistent approach to coding, to variable labelling, and for dichotomising continuous variables was established. A study ID was given to each individual trial and each participant was identified by the original study-specific participant identifier. The severity of radiographic changes were calculated for each participant on a 0-100 scale within each study. Evidence of clinical signs of inflammation (presence of effusion or swelling and warmth) were dichotomised. Biochemical inflammation was divided into tertiles within each study, based on the participant's ESR or CRP scores. Overall evidence of

inflammation was analysed as a dichotomous variable and involved combining clinical and biochemical signs of inflammation. For this, biochemical inflammatory markers (ESR or CRP) were dichotomised according to the study-specified normal threshold levels. Pain scores were standardised to a 0-100 scale within each study, assumed to be approximately equivalent to 0-100 mm VAS.

4.2.2 Statistical analysis

4.2.2.1 General approaches

The distributions of continuous variables were examined for normality using histograms. Markedly skewed data were described using medians and IQR, otherwise data were described using means and SD. Once all data were pooled, the amount of missing data were evaluated. Differences in baseline demographics were compared across participants with and without missing data. Where only a small amount of outcome data were missing and the distribution of baseline predictors were balanced across those with and without missing data, complete case analysis was conducted. To determine the representativeness of the data to the body of published evidence (Chapter 3), a random-effects CMA was conducted using data extracted from the included trials' publications or clinical trial results websites (www.novartisclinicaltrials.com).

IPD MAs were then conducted using the models and approaches illustrated in Figure 4-2. The analysis of the IPD was done using a one-stage and a two-stage approach. The one-stage approach was chosen as the primary methodological approach, whilst the results from the two-stage approach were used for sensitivity analyses.

The analysis was based on two treatment effects: the specific treatment effect (the difference between treatment and placebo) and the overall treatment effect (the change from baseline in the treatment arm) (Zou et al., 2016).

Predictors of the specific and overall treatment effects were examined. For the specific effect, the model was built using participants from both arms of placebo-controlled studies. The interaction between each

potential predictor and the treatment term were examined. For predictors of the overall treatment effect, placebo- and active-controlled trials were used, and only participants in the treatment group were analysed. Potential prognostic factors for an improved response on treatment were examined.

The primary time-point analysed was at or nearest to four weeks. Data were provided for participants at multiple durations of follow-up, and secondary analyses were conducted using data available at all time points. The treatment effect was first examined at each time point available. Analyses were then conducted by pooling data at all time points available per participant (repeated measures).

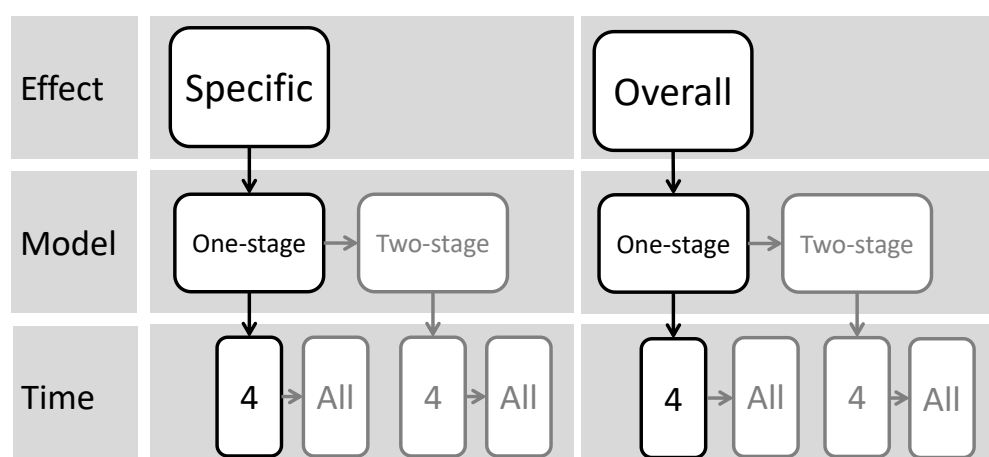


Figure 4-2 – IPD MA models and time points used for the analysis of the specific and overall treatment effect

Primary analysis (one-stage IPD MA at or nearest to four weeks) is shown in black. Secondary or sensitivity analysis are shown in grey (one-stage IPD MA using repeated measures and two-stage IPD MAs).

One-stage, one-stage IPD MA; two-stage, two-stage IPD MA; specific, specific treatment effect; overall, overall treatment effect; 4, at or nearest to four weeks; All, all time points

4.2.2.1.1 Forest plot generation

Although forest plot generation is possible for one-stage IPD MAs using the post-estimation command *ipdforest* (Kontopantelis and Reeves, 2013), it could not be successfully applied to this work. Firstly, the package cannot cope with the specification of separate residuals per study. Secondly, the command was not suitable for the presentation of the overall effect. Therefore, although the one-stage

methods were the primary approach used, forest plots were generated using the two-stage *ipdmetan* Stata command (Fisher, 2017).

4.2.2.1.2 One-stage modelling

A mixed effects multilevel model was built to account for clustering at the study level (all analyses) and patient level (repeated measures analysis). Data were analysed according to the hierarchical structure illustrated in Figure 4-3. Level 2 and 3 were included in the primary analyses using only one time point. For the secondary analyses, the model was extended to include repeated measures by specifying clustering at the participant-level (level 1) and adjusting for week (fixed common term).

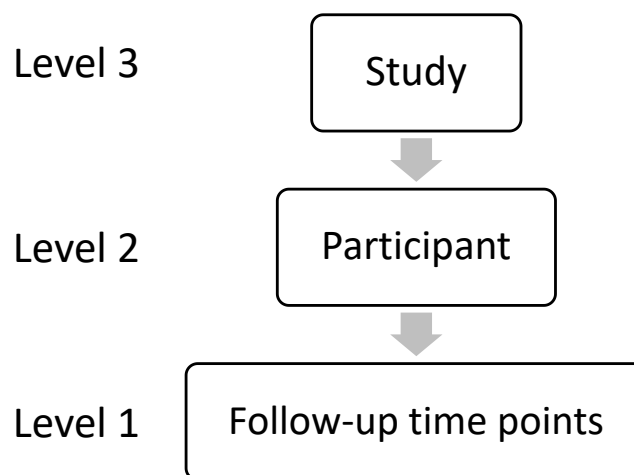


Figure 4-3 - Levels in multi-level IPD MA

Primary analysis at or nearest to four weeks included levels 2 and 3. Data at all three levels were used for the secondary analysis of repeated measures

4.2.2.1.2.1 Specific treatment effect

Follow-up pain scores were the dependent variable. The primary model (at or nearest to four weeks) was clustered at the study-level by specifying a random trial intercept (Equation 1) (Legha et al., 2018). Each trial was given a separate adjustment term for baseline pain, as well as separate residual variances (Burke et al., 2017). These assumptions are in line with the modelling of the two-stage IPD MA and were specified to ensure that both approaches were comparable. A random-effects distribution was specified for the treatment term to

allow for heterogeneity in treatment effect across included trials (Debray et al., 2015) and facilitate comparison between the IPD MA and CMA. The parameters were estimated using the restricted maximum likelihood (REML) approach (Legha et al., 2018).

Equation 1 – Specific effects one-stage IPD MA at primary time point

$$y_{Fij} = \alpha_i + \beta_i y_{Bij} + \theta_i Treat_{ij} + e_{ij}$$

$$\alpha_i \sim N(0, \delta^2)$$

$$\theta_i = \theta + u_i$$

$$u_i \sim N(0, \tau^2)$$

$$e_{ij} \sim N(0, \sigma_i^2)$$

y_{Fij} describes the follow-up pain score at or nearest to four weeks for person j in trial i , which is adjusted for baseline pain (y_{Bij}) separately within each study. α_i denotes a random effect on the study intercept. θ_i is the underlying treatment effect, which is given a random-effects distribution. e_{ij} denotes separate residual variances per study.

4.2.2.1.2.2 Predictors of the specific treatment effect

The specific effect model was extended to examine for potential predictors through the addition of a predictor (separate per trial) and a treatment-by-predictor interaction term (Equation 2). For computational efficiency, both treatment and interaction terms were assumed fixed. To quantify the presence of ecological bias, two interaction terms were included: a within-study interaction term (centred to the study mean) and an across-study interaction term based on the study mean (value that would be used in meta-regression) (Debray et al., 2015, Hua et al., 2017). The models were built using only one interaction term per model and were adjusted only for baseline pain.

Equation 2 - Interaction effects model used to examine for predictors of the specific treatment effect in a one-stage IPD MA at primary time point

$$y_{Fij} = \alpha_i + \beta_i y_{Bij} + \theta_j Treat_{ij} + \beta_j z_{Bij} + y_w Treat_{ij} * (z_{ij} - \bar{z}_j) + e_{ij}$$

$$\theta_j = \alpha + y_A \bar{z}_j + u_j$$

$$\alpha_i \sim N(0, \delta^2)$$

$$u_j \sim N(0, \tau^2)$$

$$e_{ij} \sim N(0, \sigma_j^2)$$

Equation 1 is extended to include interaction terms for predictor z (Equation 2). z_{ij} denotes the potential predictor, whose fixed-effect within-study interaction term is denoted by y_w and the across-trial interaction term is denoted by y_A . The potential predictor is centred about the mean covariate value ($\bar{z}_j$).

4.2.2.1.2.3 Overall treatment effect

As for the specific treatment effect, the primary model was clustered at the study-level by specifying a random trial intercept and separate residual variances per trial. The model was developed using only the treatment arm and a treatment term was therefore not used. An empty model was run using change from baseline pain scores as the dependent variable. The overall treatment effect was denoted by the intercept. Adjustment for baseline pain scores was not undertaken as the analysis was conducted within one group only and is not recommended in such settings (Lydersen, 2015). In addition, the unadjusted model provides a more clinically relevant output that can be interpreted as the overall treatment effect.

4.2.2.1.2.4 Predictors of overall treatment effect

As no treatment term was used in the overall effect model, a predictor-by-treatment interaction term could not be used. Instead, each potential prognostic factor was entered into the model individually as a covariate (common, fixed term). Significant predictors were subsequently examined together in a multivariable model.

As for the primary model, adjustment for baseline pain scores was not undertaken. Firstly, this was for clinical interpretability as we aimed to

examine prognostic factors that influence the change from baseline pain scores (i.e., overall treatment effect), irrespective of their independence from baseline pain levels. Secondly, this avoids potential biases associated with baseline adjustment in analyses of changes in health status (Glymour et al., 2005). Finally, this approach has been reported previously for the examination of prognostic factors associated with change from baseline pain scores following treatment for painful OA (Maricar et al., 2017).

4.2.2.1.3 Two-stage modelling

The specific treatment effect was calculated by running a regression model within each trial. Endpoint pain was used as the dependent variable and the model was adjusted only for baseline pain. The coefficient of the treatment term was then pooled across trials using a random effects MA.

For the overall treatment effect, the mean change from baseline score and associated SE were calculated per trial. The mean change and SE were then pooled in a random effects MA.

As for the one-stage approach, the predictor model was built with or without interaction terms depending on the definition of the treatment effect (dependent variable). For the specific treatment effect (i.e., the difference between treatment and placebo), an interaction term between treatment and the potential predictor was used. The partial regression coefficient of this interaction was pooled with the same interaction terms obtained from the other trials. For the overall treatment or placebo response (i.e., change from baseline) the partial regression coefficient of the prognostic factor was pooled across trials.

The two-stage specific effect model was extended to include repeated measures by specifying a two-level model within each trial. The treatment or interaction/prognostic terms were subsequently pooled in a random effects MA.

4.2.2.2 Potential predictors

4.2.2.2.1 Participant-level covariates

Table 4-1 lists all participant-level covariates that, dependent on data availability, were examined as potential predictors of treatment response.

Table 4-1 - Potential predictors of the specific and overall treatment effects

Potential predictor
Age
Sex
BMI
Baseline pain
Duration of complaints
Central sensitisation (e.g., pain elsewhere)
Evidence of inflammation (e.g., US, clinical effusion)
Radiographic findings
Type of pain (e.g., neuropathic-like pain)
Psychological assessments (e.g., pain catastrophising, depression, anxiety)
BMI, body mass index

Covariates found to be statistically significant predictors or prognostic factors were examined further in exploratory analyses. These involved subgroup analyses of the treatment effect by level of covariate and by time point, exploration of the robustness of the findings to different model assumptions, and further model adjustments for additional baseline covariates. The credibility of the participant-level subgroup effects were assessed by examination of the criteria listed in Table 4-2 (Sun et al., 2010).

Table 4-2 - Criteria used to assess the credibility of subgroup analyses

Criteria
Design
Is the subgroup variable a characteristic measured at baseline or after randomisation?
Is the effect suggested by comparisons within rather than between studies?
Was the hypothesis specified <i>a priori</i> ?
Was the direction of the subgroup effect specified <i>a priori</i> ?
Was the subgroup effect one of a small number of hypothesised effects tested?
Analysis
Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?
Is the significant subgroup effect independent?
Context
Is the size of the subgroup effect large?
Is the interaction consistent across studies?
Is the interaction consistent across closely related outcomes within the study?
Is there indirect evidence that supports the hypothesised interaction (biologic rationale)?
Criteria as per Sun et al. (2010)

4.2.2.2.2 Study-level covariates

Three study-level covariates were examined using the one-stage approach: the joint treated, sample size, and allocation concealment. An across-study interaction term between the covariate and treatment was generated for each participant. These were pooled in the examination of study-level predictors of the specific effect. For the overall effect, the study-level covariate was pooled across studies.

4.3 Results

4.3.1 Literature search and data acquisition

4.3.1.1 Topical NSAIDs

The 52 trials identified through the systematic literature search (Section 3.3.1.1.1) were eligible for inclusion to the IPD MA (Figure 4-4). A further four studies published only as conference abstracts and seven studies identified from clinical trial databases were identified. The IPD for 63 topical NSAID trials were sought.

It was not possible to contact five trials as no contact details were identified for the publication authors, institution, or trial sponsor. The first/corresponding author, co-author, institution, or sponsor were contacted for the remaining 58 trials. No reply was received for 12 trials. Of the 46 responses received, 32 did not result in a collaboration.

The predominant reason for not collaborating was a lack of interest from the pharmaceutical sponsor (10 trials) or lead author (1 study, no time for collaboration). For eight of the trials, the response received was not from the data custodian and the person responding was unable to locate the data or contact the data custodian. Five of these responses were from co-authors and three were from the institutions where the research had been conducted. Often, the IPD were simply not available. Two trialists explicitly responded that the data had been destroyed (retention period expired) whilst nine described the IPD as unavailable or lost. Four authors expressed interest, but ultimately did not collaborate. Two did not locate the data, one stopped replying after receiving the trial protocol, and one stopped replying shortly after informing us that they were on sick leave.

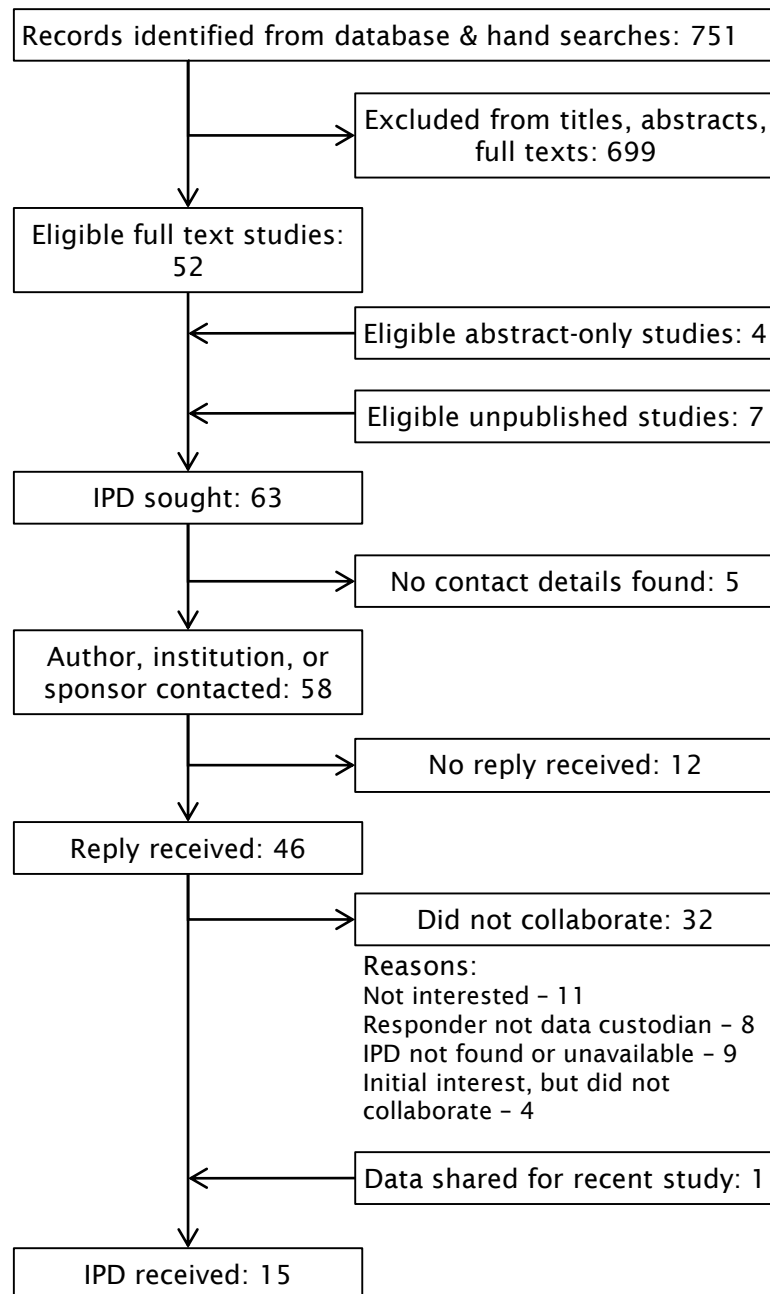


Figure 4-4 - Flow diagram of trial identification, contact, and data acquisition for IPD MA of topical NSAIDs in OA

The trial data were received from 15 trials. One of these trials was published after the date of the systematic literature search, however, the data were held by a collaborating co-author who offered to share this as well.

The predominant form of contact with data custodians was via email. Contact attempts were also made via telephone and letter, however, these methods did not prove fruitful. In addition, data were requested from Novartis (a subsidiary of GlaxoSmithKline, GSK) via clinicalstudydatarequest.com. This was followed up by telephone conversations with the relevant contact at Novartis/GSK and an application to GSK's research data sharing team.

Initial contact emails were sent in February 2016 and the final data were received approximately two years later, at the end of March 2018. DTA issues were a major hurdle to resolve for some potential collaborators and these took up to one and a half years to resolve. Data were shared using the OA Trial Bank DTA for 14 trials. One trial chose to collaborate only with the UoN using their own DTA, amended by UoN, as no consensus was reached with the OA Trial Bank (Appendix 14 – DTA for collaboration with VSM).

In parallel with study-specific contacts, pharmaceutical suppliers/producers/developers of topical NSAIDs were contacted for unpublished trial data. A total of 33 pharmaceutical companies were identified as suppliers of topical NSAIDs or sponsors of topical NSAID trials registered on trial databases. A subset of these were the funders or sponsors of trials previously identified (Figure 4-4). The pharmaceutical companies were contacted via their clinical trial data sharing team email address, medical information email address, contact email provided on trial registration websites, or website enquiry forms. Nine pharmaceutical companies responded. Two replied that they did not hold any trial data, two replied that the ownership of the topical NSAID had been transferred elsewhere, one explained that they did not conduct the planned trial, one asked for more information and then failed to reply to follow-up emails, and three were not interested in collaboration. Of those that were not interested, one

expanded upon this to explain that the product was out of their focus, actively promoted in only a handful of countries, and they were therefore not interested in a collaboration. In addition, application for the single published study they were linked to was not possible as they responded that as trial sponsors (and not funders) they did not hold the data.

4.3.1.2 Capsaicin

A total of 10 trials were identified as suitable for inclusion to the IPD MA, eight from the published MAs and two abstract-only publications (Figure 4-5). Trial authors were contacted for all 10 trials, however, none collaborated on the IPD MA. No responses were received for two trials, four trial authors did not have the trial data and were unable to access or locate it, and two responded that the data were lost or destroyed. A further two authors expressed an initial interest in collaborating, but did not respond to follow-up contact attempts. Contact with pharmaceutical suppliers of capsaicin did not yield any additional data.

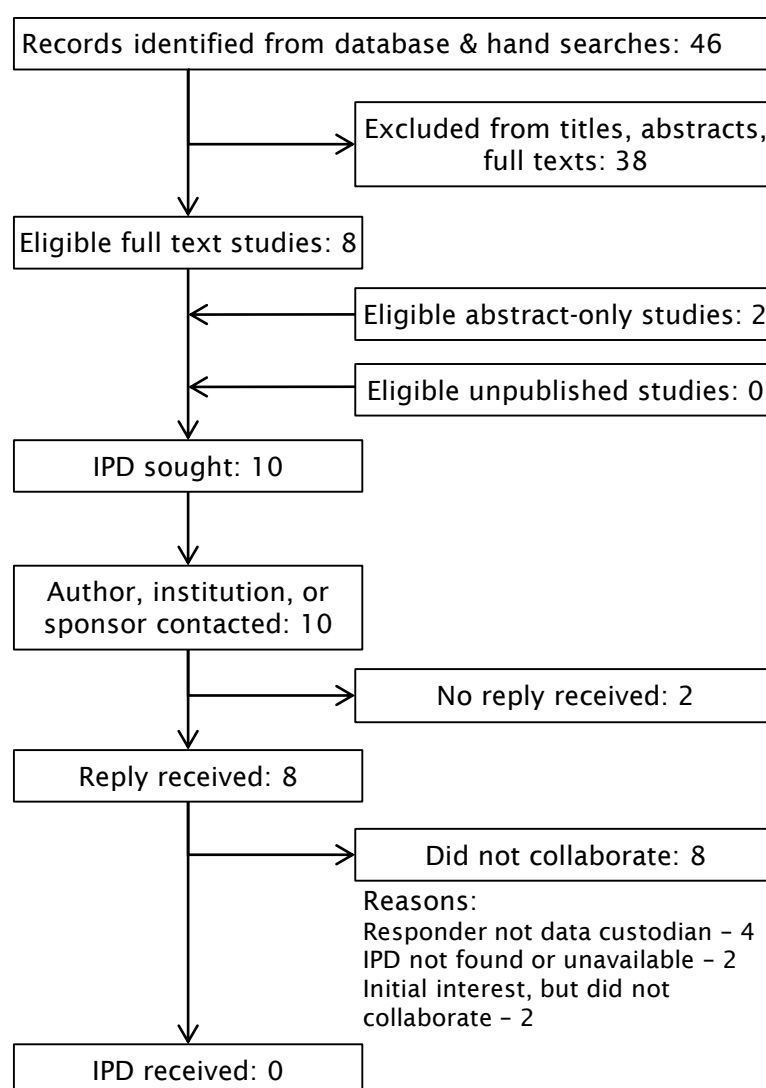


Figure 4-5 - Flow diagram of trial identification, contact, and data acquisition for IPD MA of topical capsaicin in OA

4.3.2 Predictors of response

4.3.2.1 Topical NSAIDs

4.3.2.1.1 Description of studies and participants

Data were obtained for 15 placebo- and active-controlled trials. The specific effect analyses were based only on placebo-controlled trials (11 trials, 3,140 participants), whilst participants on the topical NSAID arm of placebo- and active-controlled trials were analysed for the overall treatment effect (15 trials, 1,951 participants).

Checking the received IPD confirmed that the full datasets were received for 13 of 15 RCTs. The two remaining RCTs (Jabbari et al., 2016, Shoara et al., 2015) provided data only for participants that completed the trials. Data for participants that withdrew were sought from the data custodian, but were not available.

All trials were of parallel design and recruited community-dwelling individuals (i.e., not hospital inpatients) with OA (Table 4-3). Twelve trials were for knee OA (75% of IPD MA population) and the remaining three trials were for hand OA. The RCTs were undertaken in a range of countries, although participants were predominantly from the USA or Europe. Trials were primarily funded by pharmaceutical companies, including A.Vogel BioForce AG, IBSA Institut Biochimique SA, Inpellis, VSM, and Novartis (a GSK company).

The majority of trials assessed an NSAID gel, although patches and a cream were also assessed (Table 4-4). Diclofenac was the most commonly used topical NSAID, although ibuprofen and piroxicam were also assessed. One trial, by Underwood et al. (2008), did not specify which topical NSAID participants had to use, although recommended that they use ibuprofen.

Table 4-3 - Description of the studies included in the IPD MA of topical NSAIDs in OA

Trial Name	Year	Country of Study	Funding source	Design	Setting	Joint treated	Week analysed
Placebo-controlled trials							
Dreiser 1993	1993	France	Commercial	Parallel	Community	Knee	2
Bruhlmann 2003	2003	Switzerland	Commercial	Parallel	Community	Knee	2
Niethard 2005	2005	Germany	Commercial	Parallel	Community	Knee	3
Altman 2009	2009	USA	Commercial	Parallel	Community	Hand	4
Baraf 2010	2010	USA	Commercial	Parallel	Community	Knee	4
Varadi 2013	2013	Switzerland	Commercial	Parallel	Community	Knee	2
Shoara 2015	2015	Iran	Public	Parallel	Community	Knee	3
NCT00171626	2005*	USA	Commercial	Parallel	Community	Knee	4
NCT00171678	2005*	USA	Commercial	Parallel	Community	Knee	4
NCT00171652	2005*	Germany, France	Commercial	Parallel	Community	Hand	4
NCT01967550	2014*	Germany	Commercial	Parallel	Community	Knee	4
Active-controlled trials							
Van Haselen 2000	2000	UK	Commercial	Parallel	Community	Knee	4
Widrig 2007	2007	Switzerland	Commercial	Parallel	Community	Hand	3
Underwood 2008	2008	UK	Public	Parallel	Community	Knee	13
Jabbari 2016	2016	Iran	Public	Parallel	Community	Knee	4

Year is year of publication unless denoted by “*”. Those with “*” were not published as a stand-alone RCT publication, so year is date of trial.

Data from NCT00171652 was included in the publication by Barthel et al. (2010) which pooled two RCTs, whilst NCT00171626 and NCT00171678 were pooled in the publication by Baraf et al. (2011). NCT01967550 remains unpublished

Week analysed indicates week analysed in primary analysis (at or nearest to four weeks)

Table 4-4 - Description of the interventions and comparators used in the IPD MA of topical NSAIDs in OA

Trial Name	Intervention	Strength	Formulation	Frequency per Day	Comparator(s)
Placebo-controlled trials					
Dreiser 1993	Diclofenac hydroxyethylpyrrolidine	180mg	Patch	2	Placebo
Bruhlmann 2003	Diclofenac epolamine	180mg	Patch	2	Placebo
Niethard 2005	Diclofenac diethylamine	1.16%	Gel	4	Placebo
Altman 2009	Diclofenac sodium	1%	Gel	4	Placebo
Baraf 2010	Diclofenac sodium	1%	Gel	4	Placebo
Varadi 2013	Ibuprofen-VALE	10%	Cream	2	Placebo
Shoara 2015	Diclofenac sodium	1%	Gel	3	Placebo, chamomile oil
NCT00171626	Diclofenac sodium	1%	Gel	4	Placebo
NCT00171678	Diclofenac sodium	1%	Gel	4	Placebo
NCT00171652	Diclofenac sodium	1%	Gel	4	Placebo
NCT01967550	Diclofenac diethylamine	2.32%	Gel	2	Placebo
Active-controlled trials					
Van Haselen 2000	Piroxicam	0.5%	Gel	3	SRL homeopathic gel
Widrig 2007	Ibuprofen	5%	Gel	3	Arnica montana gel
Underwood 2008	Any topical NSAID	-	-	-	Any oral NSAID
Jabbari 2016	Diclofenac sodium	1%	Gel	3	Dwarf elder gel

NSAID, non-steroidal anti-inflammatory drug

4.3.2.1.1.1 Specific effect trials

Within the 11 trials included in the IPD MA, 1,587 participants were randomised to a topical NSAID and 1,553 participants were randomised to placebo. Overall, the trial population consisted of 2,162 (69%) women. Mean age was 62.7 years and mean baseline pain was approximately 68 mm on 0-100mm VAS scale.

Baseline demographics were balanced across topical NSAID and placebo arms (Table 4-5).

Table 4-5 - Distribution of baseline characteristics in the placebo-controlled trials analysed in the IPD MA of topical NSAIDs in OA

Baseline characteristic	n	NSAID	n	Placebo
n randomised		1,587		1,553
Women, n (%)	1,587	1,077 (67.9%)	1,553	1,085 (69.9%)
Age (years), mean (SD)	1,587	62.7 (10.2)	1,552	62.7 (10.4)
Baseline pain (NRS), mean (SD)	1,586	68.0 (17.9)	1,552	67.6 (18.0)
BMI (kg/m ²), mean (SD)	1,545	29.6 (6.2)	1,516	29.7 (6.4)
Weight (kg), mean (SD)	1,548	82.7 (19.4)	1,518	82.6 (19.6)
Inflammation present, n (%)	1,269	300 (23.6)	1,238	271 (21.9)
Clinical, n (%)	836	152 (18.2)	825	143 (17.3)
Biochemical, n (%) per tertile	1,151		1,117	
Lowest tertile		494 (42.9)		496 (44.4)
Middle tertile		325 (28.2)		300 (26.9)
Highest tertile		332 (28.8)		321 (28.7)
Knee joint affected, n (%)	1,587	1,187 (74.8)	1,553	1,170 (75.3)
Duration (months), median (IQR)	152	25 (12 to 60)	152	27 (10 to 57)
Standardised radiographic severity, mean (SD)	1,389	46.3 (15.2)	1,360	45.4 (15.2)

Clinical signs of inflammation: presence of swelling and warmth (1 study), presence of effusion (4 studies)

Inflammatory marker (divided into tertiles within each study): erythrocyte sedimentation rate (1 study), C-reactive protein (2 studies)

Standardised radiographic severity (0-100) calculated within each study from: KL grade (0-4; 7 studies), severity of changes (0-3; 1 study), number of changes (1-5; 1 study)

BMI, body mass index; IQR, interquartile range; n, number of participants for which baseline data were available; NRS, numeric rating scale; SD, standard deviation

4.3.2.1.1.2 Overall effect trials

Fifteen trials, with 1,951 participants on topical NSAIDs, were analysed for the overall treatment effect (Table 4-6). Almost 80% of participants came from placebo-controlled trials, whilst 19% were from active-controlled trials. Only one trial, with 1% of the overall IPD MA population, was placebo- and active-controlled.

The trial population was predominantly women (68%), with a mean age of 63 years. Three-quarters of the population (12 trials) had knee OA, whilst the remaining three trials recruited only hand OA. Participants reported, on average, baseline pain of 65mm on 0-100mm, although this was generally lower for participants in active-controlled trials.

Table 4-6 - Distribution of baseline characteristics in the placebo- and active-controlled trials analysed in the overall treatment effect IPD MA of topical NSAIDs in OA

Baseline characteristic	n	Placebo- controlled	n	Active- controlled	n	Placebo & active- controlled	n	All trials
n randomised		1,559		364		28		1,951
Women, n (%)	1,559	1,056 (67.7)	364	247 (67.9)	28	21 (75.0)	1,951	1,324 (67.9%)
Age, mean (SD)	1,559	62.9 (10.2)	364	61.7 (10.8)	28	52.7 (7.4)	1,951	62.5 (10.3)
Baseline pain (NRS), mean (SD)	1,557	68.2 (17.7)	362	51.6 (24.1)	28	56.4 (24.0)	1,948	65.0 (20.2)
BMI, mean (SD)	1,545	29.6 (6.2)	172	29.1 (4.7)			1,717	29.6 (6.1)
Weight (kg), mean (SD)	1,548	82.7 (19.4)	224	74.4 (15.2)			1,772	81.7 (19.1)
Inflammation present, n (%)	1,269	300 (32.6)	138	6 (4.4)			1,407	306 (21.8)
Clinical, n (%)	836	152 (18.2)	138	6 (4.4)			974	158 (16.2)
Biochemical, n (%) per tertile	1,141						1,141	
Lowest tertile		494 (42.9)						494 (42.9)
Middle tertile		325 (28.2)						325 (28.2)
Highest tertile		332 (28.8)						332 (28.8)
Knee joint affected, n (%)	1,559	1,159 (74.3)	364	265 (72.8)	28	28 (100)	1,951	1,452 (74.4)
Pain elsewhere, n (%)			136	30 (22.1)			136	30 (22.1)
Duration (months), median (IQR)	124	24 (12 to 60)	32	57 (24 to 84)	28	36 (12 to 60)	184	45.4 (42.4)
Standardised radiographic severity, mean (SD)	1,389	46.3 (15.2)	97	40.8 (16.6)			1,486	45.9 (15.4)

Clinical signs of inflammation: presence of swelling and warmth (1 study), presence of effusion (4 studies). Inflammatory marker (divided into tertiles within each study): erythrocyte sedimentation rate (1 study), C-reactive protein (2 studies). Standardised radiographic severity (0-100) calculated within each study from: KL grade (0-4; 7 studies), severity of changes (0-3; 1 study), number of changes (1-5; 1 study). BMI, body mass index; IQR, interquartile range; n, number of participants for which baseline data were available; NRS, numeric rating scale; SD, standard deviation

4.3.2.1.1.3 Risk of bias assessment

Risk of bias assessments were conducted for the 15 included trials using information available in the published reports (Figure 4-6). Three trials were only published as *post hoc* analyses of multiple RCTs conducted by the same pharmaceutical company using similar methods (Barthel et al., 2010, Baraf et al., 2011). These reports were used for the risk of bias assessment. One trial (NCT01967550) has not been published and risk of bias assessment was deemed “unclear” across all domains.

Random number sequence generation was fully described and deemed adequate in 12 of 15 trials. This was the domain associated with the lowest risk of bias. However, randomisation did not successfully balance the proportion of women (2 RCTs) and WOMAC scores (1 RCT) in the treatment groups at baseline. Three trials did not adequately blind participants and/or trial personnel (van Haselen and Fisher, 2000, Underwood et al., 2008, Shoara et al., 2015). These three trials were all active controlled.

The domain with the highest risk of bias was the use of ITT. Approximately two-thirds of trials did not include all analysed participants in the analyses in the published reports.

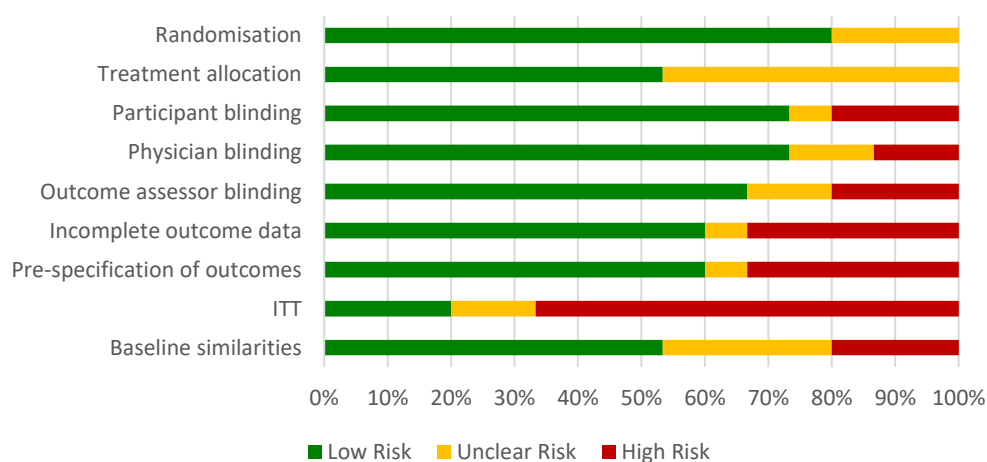


Figure 4-6 - Risk of bias assessment for topical NSAID trials analysed in the IPD MA in OA

4.3.2.1.2 Treatment effect of topical NSAIDs

4.3.2.1.2.1 Specific effect

Of 3,140 participants, 201 (6%) were missing pain data at the primary time point. Baseline demographics were balanced across those with and without missing data, and complete case analysis was used for all analyses.

Topical NSAIDs were statistically superior to placebo for pain relief at or nearest to four weeks. Participants on topical NSAIDs reported 6.35 mm (95%CI 3.81 to 8.89) less pain at endpoint than those on placebo (Table 4-7, Figure 4-7).

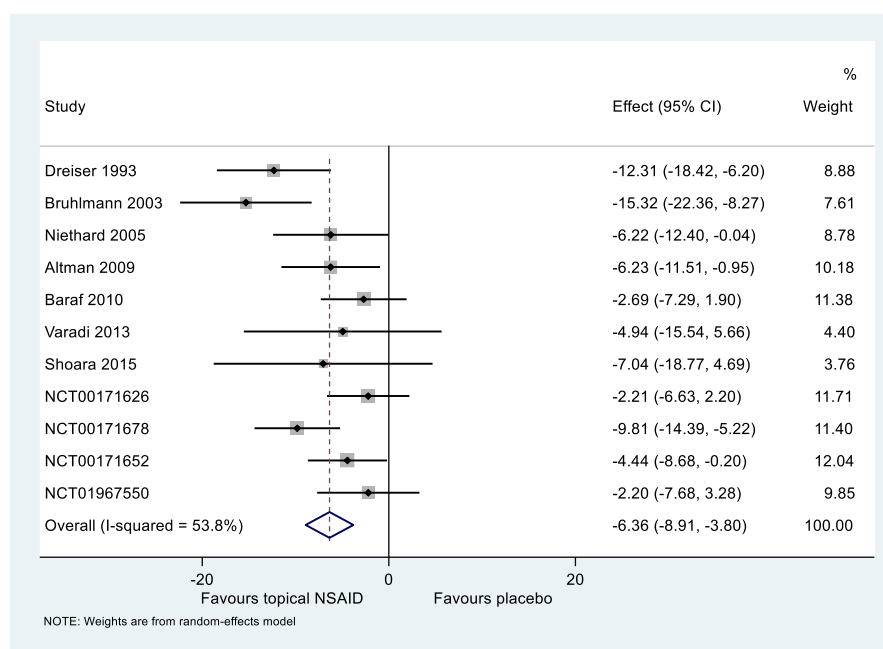


Figure 4-7 - Specific effect of topical NSAIDs versus placebo at or nearest to four weeks in two-stage IPD MA

Effect presented as MD on 0-100 mm VAS pain scale

Results from the two-stage IPD MA were in agreement with the primary one-stage results (Table 4-7). Results from the CMA of the subset of RCTs included in the IPD MA were comparable to the IPD MA findings, although the specific effect was marginally larger using the conventional versus IPD MA approach. Comparison with the overall CMA result (ES 0.31, 95%CI 0.20 to 0.41, section 3.3.2.1.1) indicated that the subset of RCTs included in the IPD MA were representative of the overall published evidence base, although results were slightly conservative.

Table 4-7 - Specific effect of topical NSAIDs in OA at or nearest to four weeks

Analysis	MD, VAS (mm)	95%CI
One-stage	-6.35	-8.89 to -3.81
Two-stage	-6.36	-8.91 to -3.80
Conventional*	-7.50	-10.75 to -4.25

*translated from Hedges' ES of 0.30 (95%CI 0.17 to 0.43) (Wandel et al., 2010)
CI, confidence interval; MD, mean difference; VAS, 0-100 mm visual analogue scale

4.3.2.1.2.1.1 Repeated measures

Data were available for multiple time points per study and these are listed in Table 4-8. The most commonly assessed time-point of assessment was at one week and the longest duration of follow-up was 12 weeks.

Table 4-8 - Time points (weeks) of assessment and associated specific effect of topical NSAIDs in OA

Week	Trials	n	MD, VAS (mm)	95%CI
1	11	3,067	-6.15	-9.49 to -2.81
2	8	1,649	-7.51	-10.88 to -4.15
3	3	576	-4.79	-8.64 to -0.93
4	6	2,356	-4.62	-7.13 to -2.10
6	2	716	-5.76	-9.26 to -2.26
8	5	1,918	-4.52	-6.75 to -2.29
12	3	1,151	-5.62	-8.55 to -2.68

CI, confidence interval; MD, mean difference; n, number of participants; VAS, 0-100 mm visual analogue scale; trials, number of trials

When including repeated measures data (11,433 observations), adjusted for week of assessment, topical NSAIDs remained statistically superior to placebo (Table 4-9). Some variation was observed between the one- and two-stage approaches, with the two-stage reporting a greater effect and wider CI.

Table 4-9 - Specific effect of topical NSAIDs in OA using all time-points of assessment (repeated measures)

Analysis	MD, VAS (mm)	95%CI
One-stage	-6.19	-8.36 to -4.03
Two-stage	-7.02	-9.97 to -4.06

CI, confidence interval; MD, mean difference; VAS, 0-100 mm visual analogue scale

4.3.2.1.2.2 Overall effect

Of 1,951 participants, 92 (5%) were missing pain outcome data at the primary time point. Baseline demographics were balanced across those with and without missing data, and complete case analysis was used for all analyses.

Participants on topical NSAID experienced, on average, 25mm decrease in pain on 0-100 mm VAS (Table 4-10, Figure 4-8).

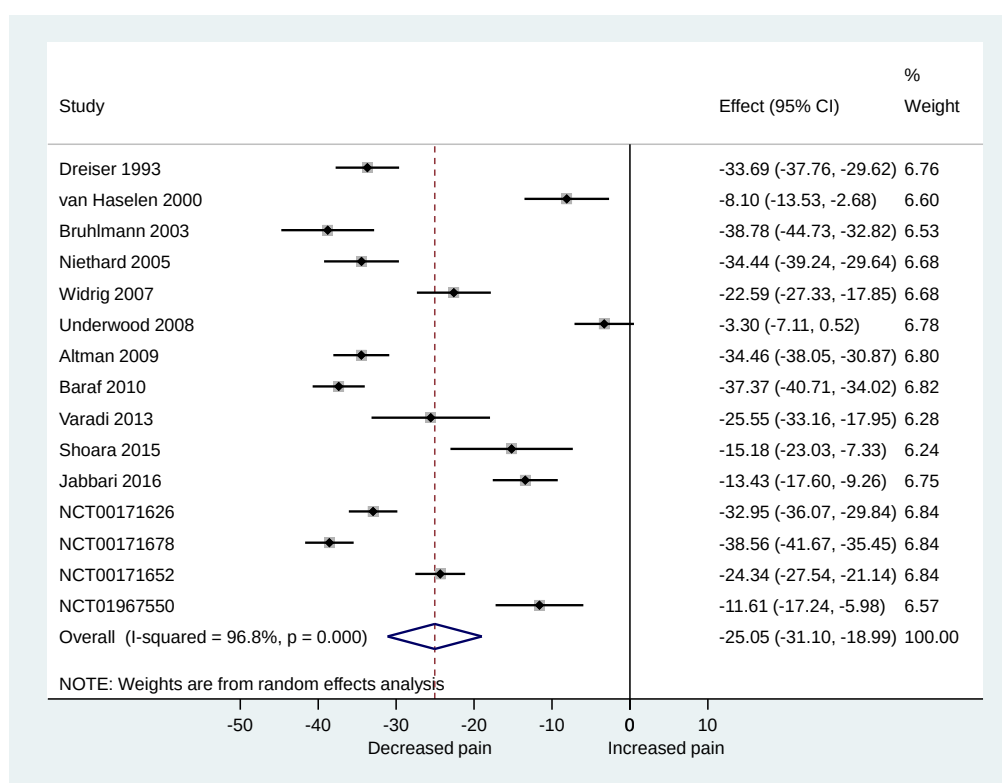


Figure 4-8 - Overall effect of topical NSAIDs at or nearest to four weeks in two-stage IPD MA

Effect presented as change from baseline scores on 0-100 mm VAS

Overall effect estimates from the one-stage and two-stage IPD MAs were in agreement. CMA of the included trials showed a similar overall effect, however, data for one study (Underwood et al., 2008) were not available for extraction from the published report. This study was consequently excluded from the CMA. The overall effect from the IPD MA trials was comparable to that of the body of published evidence (ES 1.23, 95%CI 1.06 to 1.41, section 3.3.4.1.1), albeit being slightly lower.

Table 4-10 - Overall effect of topical NSAIDs in OA at or nearest to four weeks

Analysis	Overall effect, VAS (mm)	95%CI
One-stage	-25.04	-31.14 to -18.95
Two-stage	-25.05	-31.10 to -18.99
Conventional ES*	-29.50	-34.50 to -24.50

*Translated from Hedges' ES of 1.18 (95%CI 0.98 to 1.38) (Wandel et al., 2010)
CI, confidence interval; VAS, 0-100 mm visual analogue scale

4.3.2.1.2.2.1 Repeated measures

Data were available for 11 different time-points of assessment, which are listed in Table 4-11. Shorter durations of follow-up were more commonly assessed in the included RCTs, although one trial (Underwood et al., 2008) followed up participants for up to two years.

Table 4-11 - Time points (weeks) of assessment in one-stage IPD MA trials of the overall effect of topical NSAIDs in OA

Week	Trials	n	Overall effect, VAS (mm)	95%CI
1	11	1,556	-21.25	-26.82 to -15.67
2	9	880	-23.14	-30.17 to -16.10
3	4	386	-20.61	-31.27 to -9.95
4	8	1,329	-25.25	-33.84 to -16.65
6	2	369	-34.46	-44.83 to -24.09
8	5	989	-38.26	-42.38 to -34.15
12	3	595	-42.78	-47.46 to -38.09
13*	1	128	-3.30	-7.11 to 0.52
26*	1	120	-2.28	-6.40 to 1.83
52*	1	108	-1.11	-5.74 to 3.52
104*	1	34	4.32	-3.62 to 12.27

* Only one trial provided data and analysis is therefore a regression model built only within that study
CI, confidence interval; VAS, 0-100 mm visual analogue scale; n, number of participants

Across the 6,494 observations collected at different durations of assessment for 1,925 participants, topical NSAIDs produced an overall

treatment effect of -19.44 mm VAS (95%CI -27.03 to -11.86) in the one-stage IPD MA. Two-stage IPD MA and CMA of the included RCTs were not conducted.

4.3.2.1.3 Predictors of response to topical NSAIDs

Data were available for nine potential participant-level determinants of outcome: sex (women vs men), age (years), baseline pain (0-100 mm VAS), BMI (kg/m²), evidence of inflammation (overall (y/n), evidence of clinical inflammation (y/n), biochemical markers of inflammation (tertiles)), duration of complaints (months), and radiographic severity (0-100 standardised scale). Of the potential predictors, only sex, age, and baseline pain were measured in all trials. Although only a subset of trials reported BMI and duration of complaints, these were all assessed on the same scale and were therefore directly comparable.

Most commonly, trials recorded the presence or absence of clinical signs of inflammation in the form of effusion (four studies) or warmth and swelling (one study) of the affected joint. Inflammation was also assessed using biochemical signs of inflammation (ESR and CRP). Although insensitive, these markers were used as surrogates for inflammation. In order to combine ESR and CRP, they were divided into tertiles per trial (highest to lowest level of blood marker) before pooling. In order to generate an overall indicator of inflammation, clinical and blood markers were combined. The blood markers were dichotomised using the participant-specific thresholds of normality specified within the primary trials.

No trials assessed the quality of pain or presence of neuropathic-like pain. Only one trial recorded the presence of pain elsewhere (a possible indicator for central mechanisms of pain) and this was therefore not analysed.

Radiographic OA severity was assessed in nine trials using different criteria. Seven studies presented KL grades, one study rated the severity of changes, and the last study reported only the number of changes. In order to pool the radiographic severity grades, these were first standardised to a 0-100 scale per trial.

4.3.2.1.3.1 Predictors of specific effect

In general, one-stage and two-stage approaches were in agreement, although more precise estimates (narrower CIs) were generated through the one-stage approach. Table 4-12 presents the β coefficient for the treatment-by-predictor interaction term for all potential predictors examined.

Table 4-12 - Interaction effect (β) and associated 95%CI for potential predictors of the specific effect of topical NSAIDs at or nearest to four weeks

Predictor	Trials	n	One-stage		Two-stage	
			β	95%CI	β	95%CI
Sex	11	2,939	-4.28	-7.98 to -0.58	-4.23	-7.88 to -0.58
Age	11	2,938	0.06	-0.11 to 0.22	0.05	-0.11 to 0.22
Baseline pain	11	2,939	0.05	-0.06 to 0.16	0.04	-0.06 to 0.15
BMI	10	2,863	0.03	-0.27 to 0.32	0.01	-0.29 to 0.31
Inflammation	7	2,339	-0.31	-4.98 to 4.36	-0.46	-5.09 to 4.17
Clinical	4	1,550	-1.84	-8.24 to 4.55	-2.47	-10.32 to 5.39
Biochemical	6	2,116	0.90	-1.53 to 3.33	1.07	-2.47 to 4.61
Duration	3	280	-0.04	-0.15 to 0.07	-0.05	-0.19 to 0.09
XR severity	7	2,576	-0.01	-0.12 to 0.14	0.01	-0.14 to 0.16

Statistically significant interaction effects are shown in bold
 β , beta-coefficient for interaction effect; BMI, body mass index; CI, confidence interval;
n, number of participants included in analysis; XR severity, radiographic severity

A statistically significant interaction was observed between sex and treatment (p for interaction = 0.023 in one-stage and two-stage IPD MAs) (Figure 4-9). Compared to men, women reported greater reductions in pain when using topical NSAIDs.

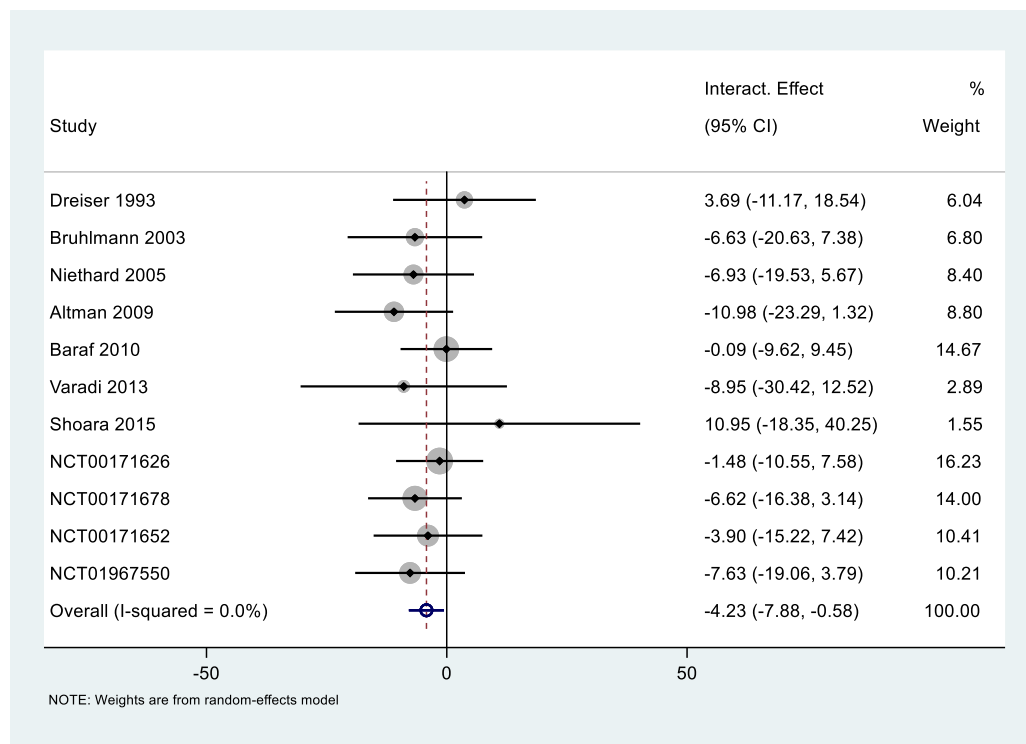


Figure 4-9 – Treatment-by-sex interaction term for topical NSAIDs at or nearest to four weeks in two-stage IPD MA

The specific effects of topical NSAIDs by sex are presented below (Table 4-13).

Table 4-13 - Specific effect of topical NSAIDs at or nearest to four weeks by sex

Sex	One-stage MD, VAS (mm)	95%CI	Two-stage MD, VAS (mm)	95%CI
Women	-7.43	-10.05 to -4.81	-7.43	-10.03 to -4.82
Men	-3.02	-5.96 to -0.09	-3.14	-6.07 to -0.20

CI, confidence interval; MD, mean difference; VAS, 0-100 mm visual analogue scale

4.3.2.1.3.1.1 Repeated measures

Potential predictors of response were assessed using data collected at all durations of assessment. No significant interactions were observed between any participant-level characteristic and treatment (Table 4-14).

Table 4-14 - Interaction effect (β) and associated 95% CI for potential predictors of the specific effect of topical NSAIDs using repeated measures data

Predictor	Trials	n	Obs	One-stage		Two-stage	
				β	95%CI	β	95%CI
Sex	11	3,097	11,433	-2.94	-6.07 to 0.20	-2.68	-5.68 to 0.33
Age	11	3,096	11,431	0.07	-0.08 to 0.21	0.04	-0.10 to 0.18
Baseline pain	11	3,097	11,433	0.01	-0.08 to 0.10	0.00	-0.09 to 0.09
BMI	10	3,019	11,195	0.07	-0.18 to 0.31	0.06	-0.22 to 0.33
Inflammation	7	2,477	9,524	-0.67	-4.57 to 3.23	-0.68	-4.41 to 3.04
Clinical	4	1,633	5,759	-2.50	-7.86 to 2.87	-2.48	-7.61 to 2.66
Biochemical	6	2,247	8,843	1.01	-1.04 to 3.06	1.28	-1.27 to 3.82
Duration	3	292	626	-0.00	-0.09 to 0.09	-0.00	-0.09 to 0.09
XR severity	7	2,719	10,637	-0.02	-0.12 to 0.09	-0.03	-0.17 to 0.12

β , beta-coefficient for interaction effect; BMI, body mass index; CI, confidence interval; n, number of participants; Obs, number of observations; Trials, number of trials; XR, radiographic severity

Sex did not significantly interact with treatment (p for interaction = 0.066 in one-stage and 0.081 in two-stage). The reasons for the difference in the significance of the treatment-by-sex interaction terms from the primary and secondary analyses were examined. The results of the three-level hierarchical model were robust to different model specifications, including the specification of random interaction terms, common baseline adjustment terms, and homoscedasticity (error term is the same across all values of the independent variable).

The specific effect of women versus men was examined separately at all durations of assessment (Figure 4-10). A larger specific effect and narrower CI was observed for women at all weeks. Topical NSAIDs were only superior to placebo at week one for men, but at all weeks for women. The difference between the sexes was greatest between two to four weeks, which were the time points analysed in the primary analysis. The analysis at the primary time point was based on data from 333 participants at week two, 294 participants at week three, and 2,513 participants at week four.

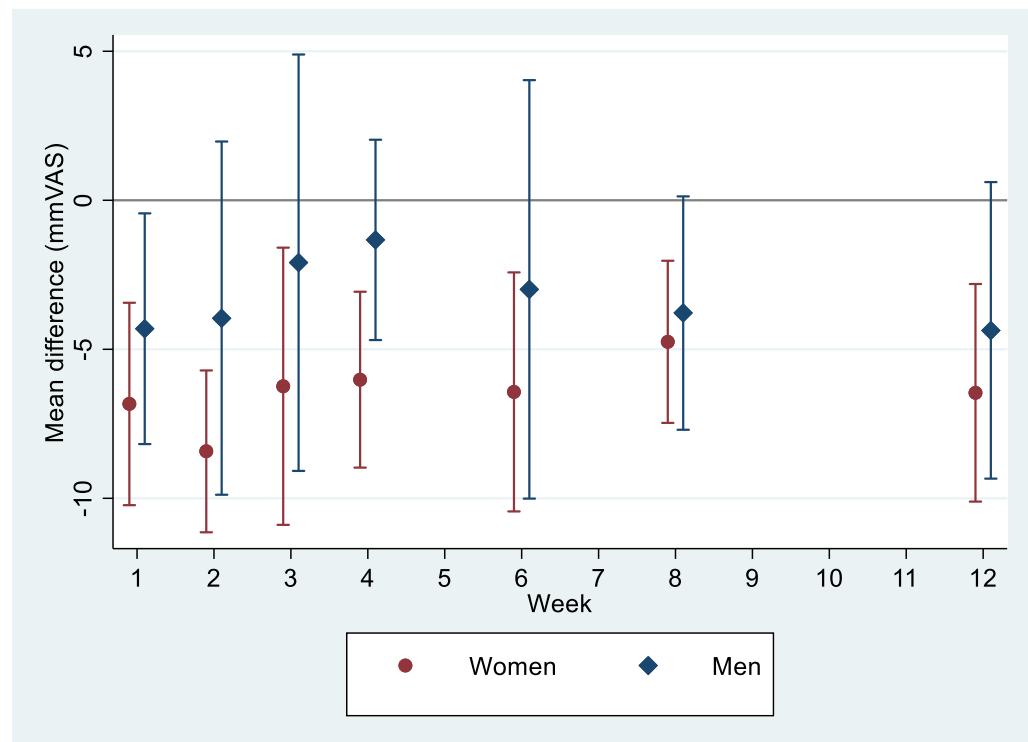


Figure 4-10 – Specific effect of topical NSAID, presented as MD (0-100mm VAS), for men and women at all durations of assessment

4.3.2.1.3.2 Predictors of overall effect

Nine potential predictors of the overall effect of topical NSAIDs were examined. Of these, only sex ($p=0.008$) and baseline pain ($p<0.001$) were significantly associated with response to topical NSAIDs (Table 4-15).

Table 4-15 - β coefficient and associated 95% CI for potential predictors of the overall treatment effect of topical NSAIDs at or nearest to four weeks

Factor	Trials	n	One-stage		Two-stage	
			β	95%CI	β	95%CI
Sex	15	1,857	-3.17	-5.53 to -0.81	-3.20	-5.57 to -0.83
Age	15	1,857	-0.02	-0.13 to 0.09	-0.01	-0.12 to 0.10
Baseline pain	15	1,857	-0.53	-0.59 to -0.47	-0.50	-0.63 to -0.37
BMI	12	1,633	0.09	-0.12 to 0.29	0.12	-0.23 to 0.46
Inflammation	8	1,329	-2.16	-5.42 to 1.09	-2.23	-5.87 to 1.42
Clinical	5	924	-3.03	-7.40 to 1.34	-3.42	-9.48 to 2.64
Biochemical	6	1,088	1.18	-0.55 to 2.90	1.24	0.98 to -3.45
Duration	4	181	-0.03	-0.09 to 0.03	-0.03	-0.10 to 0.03
XR severity	9	1,412	0.08	-0.00 to 0.17	0.08	-0.03 to 0.18

Statistically significant β coefficients are shown in bold

β , beta-coefficient for prognostic factor; BMI, body mass index; CI, confidence interval; n, number of participants; XR, radiographic severity

Individuals with higher levels of pain at baseline reported significantly greater levels of pain relief after treatment with topical NSAID. In addition, women reported greater pain relief than their male counterparts. The overall treatment effect of topical NSAIDs by sex are presented below (Table 4-16).

Table 4-16 - Overall effect of topical NSAIDs at or nearest to four weeks by sex

Sex	One-stage		Two-stage	
	VAS (mm)	95%CI	VAS (mm)	95%CI
Women	-25.77	-31.97 to -19.56	-25.78	-31.65 to -19.91
Men	-23.30	-29.42 to -17.19	-23.29	-29.76 to -16.82

CI, confidence interval; VAS, 0-100 mm visual analogue scale

One-stage multivariable analysis, including baseline pain and sex in the model, rendered the effect of sex non-significant ($p=0.162$), whilst baseline pain remained significant ($p<0.001$).

4.3.2.1.3.2.1 Repeated measures

As for the primary analysis, baseline pain and sex were statistically significant prognostic factors for the overall treatment effect when using data at all durations of assessment (Table 4-17). Those with greater baseline pain reported significantly more pain relief after using topical NSAIDs. Similarly, women reported more pain relief than men. However, in multivariable analysis the effect of sex was non-significant ($p=0.091$), whilst baseline pain remained significant ($p<0.001$).

Table 4-17 - β coefficient and associated 95% CI for potential predictors of the overall treatment effect of topical NSAIDs using repeated measures data

Factor	Trials	n	Obs	One-stage		Two-stage	
				β	95%CI	β	95%CI
Sex	15	1,925	6,494	-2.93	-5.12 to -0.75	-2.82	-4.92 to -0.73
Age	15	1,925	6,494	0.00	-0.10 to 0.10	-0.02	-0.13 to 0.10
Baseline pain	15	1,925	6,494	-0.55	-0.61 to -0.50	-0.48	-0.60 to -0.36
BMI	12	1,699	6,183	0.10	-0.08 to 0.28	0.12	-0.14 to 0.38
Inflammation	8	1,391	5,273	-1.66	-4.42 to 1.09	-1.82	-4.74 to 1.09
Clinical	5	959	3,343	-3.04	-6.77 to 0.69	-3.20	-6.87 to 0.48
Biochemical	6	1,143	4,537	0.65	-0.83 to 2.12	0.86	-1.06 to 2.79
Duration	4	182	389	-0.01	-0.07 to 0.05	-0.04	-0.10 to 0.02
XR severity	9	1,473	5,565	0.04	-0.04 to 0.12	0.05	-0.05 to 0.15

Statistically significant β coefficients are shown in bold

β , beta-coefficient for prognostic factor; BMI, body mass index; CI, confidence interval; n, number of participants; Obs, number of observations; St, number of studies; XR, radiographic severity

4.3.2.1.3.3 Study-level predictors

Three study-level predictors of response were assessed in the one-stage IPD MA: the joint treated, sample size, and allocation concealment (as determined from publications or trial registry) (Table 4-18). A small, but statistically significant, interaction between sample size and treatment was observed for the specific effect ($p=0.022$). Smaller trials reported larger specific ES, however, this pattern was not observed for the overall treatment effect.

Table 4-18 - β coefficient for potential study-level predictors of the specific and overall treatment effect of topical NSAIDs at or nearest to four weeks

Predictor	Specific effect		Overall effect	
	β	95%CI	β	95%CI
Joint	-1.07	-4.90 to 2.77	2.65	-13.02 to 18.32
Sample size (per 100)*	1.45	0.21 to 2.69	-2.83	-6.87 to 1.22
Allocation concealment	1.78	-1.56 to 5.12	-7.70	-19.89 to 4.49

Statistically significant β coefficients are shown in bold

*Effect presented per increase in sample size of 100 participants

β , beta-coefficient for interaction term (specific effect analysis) or factor (overall effect analysis); CI, confidence interval

4.4 Discussion

4.4.1 Key findings and interpretation

To our knowledge, this is the first IPD MA conducted for topical NSAIDs in OA. No data were received for capsaicin in OA and it was not possible to conduct an IPD MA for capsaicin.

Topical NSAIDs were statistically superior to placebo for pain relief, both at four weeks and using all available data (up to 12 weeks). Compared to placebo, participants on topical NSAIDs experienced approximately 6mm more pain relief on 0-100 mm VAS. Participants allocated a topical NSAID experienced, on average, a 25mm pain reduction from baseline at or nearest to four weeks. Examination of data at all durations of follow-up (repeated measures data) found that mean pain reductions on topical NSAIDs were approximately 20mm.

Seven potential participant-level predictors of response were examined: age, sex, BMI, baseline pain severity, inflammation, duration of complaints, and radiographic severity. Significant differences were observed only for sex and baseline pain. A significant treatment-by-sex interaction was found at the primary time point. Women reported greater specific treatment effect than men at or nearest to four weeks. Women also reported a larger overall treatment effect than men, but this association was not independent of baseline pain severity. Finally, those with greater pain at baseline reported larger overall treatment effects for topical NSAIDs.

4.4.1.1 Sex differences in response to topical NSAIDs

Differences in the response to topical NSAIDs were observed between the sexes, with women generally reporting greater treatment efficacy than men. The credibility of this potential subgroup effect was examined using the criteria reported by Sun et al. (2010).

There is increased confidence in the subgroup effect of sex as:

- a) the subgroup characteristic was measured at baseline,
- b) the effect was found within (rather than between) studies,

c) the interaction was consistent across studies ($I^2=0\%$ in two-stage IPD MA),

d) the interaction test suggested a low likelihood ($p=0.023$) that the subgroup effect was due to chance,

e) the relative size of the interaction effect was large (compared to the pooled specific effect), even though the absolute size is small, and

f) the interaction effect was independent of baseline pain. It was not possible to assess whether the effect was independent of other factors such as central sensitisation, type of pain, etc.

On the other hand, although the examination of sex was specified *a priori*, it was one of many covariates examined and the direction of the effect was not pre-specified. Additionally, although the direction of intervention effect was maintained when analysing data at all durations of follow-up, the findings were non-significant. Putative explanations for the difference between the primary and repeated measures analyses are provided below.

Firstly, it may be because of time-dependent differences in the interaction between sex and topical NSAID use. These differences may be true differences. For example, women may experience more pain relief than men at shorter durations of follow-up, but at longer durations they experience the same level of pain relief. However, the analyses at each time point are based on different studies and different participants. The observed variations in effect may be secondary to variations in response between populations and not time points. Secondly, differences in the statistical modelling and its assumptions are likely to influence the interaction effect. Although the inclusion of repeated observations increases the number of observations for analyses, it also adds significant complexity. Due to issues with non-convergence, the model assumptions for the repeated measures model differed from that of the primary model. Separate residual variances at the study-level were assumed in the primary model, whilst separate residual variances at the week-level were used in the repeated measures model. These variations may further explain the differences

in the findings. Thirdly, it may be that there is no significant interaction between sex and topical NSAID use overall, but that a spurious significant result was found at or nearest to four weeks by chance. The direction of the effect was maintained across a closely related outcome (pain at all durations of follow-up), however, the significance was not. This may reduce our confidence in the subgroup effect.

The final consideration for the credibility of the subgroup effect is its biological rationale, which is discussed below.

It is widely recognised that important differences in pain occur between men and women. Not only do many chronic pain conditions occur more commonly in women, but the experience and determinants of pain itself vary between the sexes (Greenspan et al., 2007, Bartley and Fillingim, 2013). Women consistently report greater levels of pain than their male counterparts across a variety of diagnostic groups (Bartley and Fillingim, 2013). The observed differences between men and women may result from a multitude of factors, including race/ethnicity/culture, biological mechanisms, physical variables, beliefs, coping mechanisms, and mood (Greenspan et al., 2007, Bartley and Fillingim, 2013). The treatment-by-sex interaction found in the present work was independent of baseline pain, indicating that the rationale for the subgroup effect is more likely explained by other factors.

Sex differences have not been reported for oral NSAIDs in postoperative, acute (third molar extraction), or OA pain (Barden et al., 2002, Averbuch and Katzper, 2000, Detora et al., 2001, Bingham III et al., 2011). The lack of sex difference for oral NSAIDs suggests that the difference observed may be explained by an interaction between sex and topical drug delivery of NSAIDs, rather than the drugs themselves.

In order to exert their effect, NSAIDs must penetrate through the skin and to deeper structures. Women have thinner skin than their male counterparts, and this difference may be even more noticeable in older age (Giacomoni et al., 2009). Perhaps this results in greater drug penetration and increased efficacy. Additionally, as the recommended

dose of topical NSAID is unaffected by the sex or size of the participant, it may be that women are applying more topical NSAID relative to their body size. These factors may all result in greater concentrations of topical NSAIDs at the site of action. Regrettably, no evidence is available on the presence of sex-specific differences in topical NSAID penetration or absorption.

A similar pattern in response between the sexes has been reported in an IPD MA for the 8% capsaicin patch in HIV neuropathy (Katz et al., 2015a). Women were more likely to experience a sustained or complete response than men. This might indicate that the increased efficacy of topical analgesics is due to a biological mechanism in drug penetration, such as thinner skin, which is consistent across topical analgesics. However, the work did not assess treatment-by-covariate interactions (only associations with response) and did not detect the same pattern for participants with PHN. The trial population for HIV was largely male (11% women), whilst the population for PHN was balanced for sex (54.8% women). Simulation work has shown that large imbalances in subgroup ratios has the potential to impact subgroup-specific test results (Brookes et al., 2001). The effect of sex was therefore likely to have been an artefact of subgroup imbalances in the trials rather than a genuine difference in effects. Unfortunately, the evidence available for topical capsaicin therefore does little to further our understanding of sex differences with regards to topical analgesics. Although the present IPD MA is imbalanced (70% women) with regards to sex, treatment-by-covariate interaction tests have been shown to be largely robust to imbalances in subgroup ratios (Brookes et al., 2001).

An alternative, and perhaps more feasible, rationale for the observed interaction may be that cultural differences between men and women affect the use of topical NSAIDs. Significantly more women report using personal care products, such as body lotion, than men (Biesterbos et al., 2013, Wu et al., 2010). Women may therefore be more compliant with topical analgesic use as they are more accustomed to applying care products. Alternatively, it may be that women are more likely to rub in the topical NSAID, as they would presumably do with body

lotion. Rubbing in topical diclofenac gel has been shown to increase diclofenac flux through the skin, although this finding is not consistent across all time points (Hasler-Nguyen and Fotopoulos, 2012). Finally, psychosocial differences may affect the differences. For example, men and women may have different treatment expectancies, which influence application and pain reporting.

On consideration of the evidence, women may experience more pain relief on topical NSAIDs than men. However, the difference between the sexes is minimal and unlikely to be clinically significant. These differences would, for example, probably not be captured if using a 0-10 NRS. Finally, although the interaction effect was independent of baseline pain, it was not possible to adjust for other factors such as central sensitisation or type of pain. The randomisation process should balance such confounders between arms for examination of the treatment effect. However, the effect of sex may be confounded as it has not been randomised (VanderWeele and Knol, 2011). We cannot strictly conclude that the difference observed is attributable to sex itself. It may be attributable to unmeasured differences between men and women.

4.4.1.2 Baseline pain

Participants with higher pain at baseline reported more pain relief on topical NSAID than those with less pain at baseline. As the comparison is within the topical NSAID arm only, worse baseline pain should be considered a prognostic factor for increased pain reduction on a topical NSAID, rather than a predictor of response to topical NSAIDs.

A significant association was found between baseline pain and overall effect, but no interaction between baseline pain and topical NSAID use was observed. This difference may arise from differences in the statistical modelling, but could also indicate that baseline pain is a prognostic factor for the contextual, rather than the specific, effect of topical NSAIDs.

Increased baseline pain severity has previously been shown to associate with increased placebo responses in OA at the study-level

(Zhang et al., 2008b). The present IPD MA reaffirms these findings and provides evidence that they are reflected at the individual-level as well. The most likely explanation for the observed effect is regression to the mean (Glymour et al., 2005), whereby unusually large (or small) measurements tend to be followed by values that are closer to the mean (Barnett et al., 2004). In practical terms, this will usually entail that individuals reporting high baseline pain levels are more to report lower pain levels (i.e., closer to the group mean) on follow-up.

4.4.1.3 Study-level predictors

A small, but statistically significant, interaction was found between sample size and the specific treatment effect of topical NSAIDs. Smaller trials reported larger specific effects for topical NSAIDs. This is in keeping with the CMA for topical NSAIDs (section 3.3.2.1.1.2) and previous work (Nüesch et al., 2010).

4.4.2 Caveats

The present IPD MA is subject to several limitations. Firstly, poor IPD retrieval rates meant that an IPD MA of capsaicin could not be conducted and only a subset of studies (25%) were analysed for topical NSAIDs. The present work may therefore be subject to data availability bias (Ahmed et al., 2012). Despite the poor study retrieval rate, the IPD MA still included 3,140 participants for the primary analysis, which is over half of the participants analysed in the primary CMA. In addition, comparisons between the captured data and all published evidence (CMA) indicated that the subgroup of studies captured were likely representative of the evidence base. Analyses combining IPD and aggregate data were not conducted as the primary objective of the IPD MA was to identify predictors of response, which is not possible with aggregate data.

Secondly, the examination of potential predictors of response was limited by data availability. Only covariates measured in the original trials could be assessed and potentially important unmeasured predictors could not be assessed. In addition, covariates were measured differently across trials such that pooled covariates may not

have been measuring the same characteristic across all studies. Furthermore, studies assessed different covariates and the analyses for each covariate are therefore based on different populations. Finally, RCT populations are generally relatively homogenous and may not reflect the true variation in important covariates found in the general population. Limitations such as these often affect IPD MAs that examine subgroup effects in RCTs. Only approximately a third of these IPD MAs identify significant subgroup-treatment interactions (Schuit et al., 2018). Significant interaction are more commonly found for participant-level variables and where a greater number of studies/participants are available for analysis (Schuit et al., 2018). Although IPD MAs have the scope to capture significant interactions, only a small proportion (<4%) have identified subgrouping factors associated with large differences in effect between the subgroups (Schuit et al., 2018).

Thirdly, although the participant characteristics examined were defined *a priori*, multiple covariates were examined in the analyses. It is widely accepted that, even where no association exists, one would expect to find that one in 20 comparisons are significant ($p < 0.05$) simply by chance (Kirkwood and Sterne, 2003). As such, the examination of predictors should be interpreted with caution and viewed as exploratory in nature.

Fourthly, in order to pool data, pain outcome scores were standardised from their original scales to a 0-100 scale. This approach has been used in previous IPD MAs in OA (Runhaar et al., 2017, Van Middelkoop et al., 2016), but may not be valid if the instruments have different measurement properties. The appropriateness of standardising WOMAC scores to 0-100 scale has, for example, been highlighted as an area requiring further research (Copsey et al., 2018).

Fifthly, the statistical methods are subject to caveats. The model specifications for the one-stage IPD MA were set to mirror those of the two-stage approach (Burke et al., 2017). Model assumptions were also guided by the complexity of the models and it was not always feasible to make more complex assumptions due to non-convergence.

Similarly, ITT analysis was not conducted because of the complexity in combining multiple imputation with multi-level regression modelling. The results of the present work are therefore limited by the assumptions made.

Finally, the work may be limited by potential risks of bias in the original studies. Assessment of the published manuscripts indicated some concerns over baseline imbalances, incomplete outcome reporting, and pre-specification of outcomes. These risks of bias were minimised in the analysis of raw data from the trials through adjustment for baseline covariates and the analysis of all relevant data. Examination of the remaining domains suggests low risks of bias for randomisation, allocation concealment, and blinding. Potential risks of bias remain for the absence of ITT analysis.

4.4.3 Methodological comments

4.4.3.1 IPD retrieval rates

A meta-epidemiologic study found that of the 760 IPD MAs published between 1987 and 2015, approximately 9% of publications reported receiving data for less than 30% of all available studies (Nevitt et al., 2017) (additional data received from author). Only one of these (0.13%), a Cochrane SR, reported not receiving any IPD (Nevitt et al., 2017, Stengel et al., 2005). The estimate is likely to underestimate the true prevalence of low IPD retrieval rates, as “unsuccessful” IPD MAs may simply not have been published (publication bias). Unfortunately IPD retrieval rates have not significantly improved in recent years despite increased awareness of IPD MA and data sharing requirements (Nevitt et al., 2017).

Despite adopting methods associated with increased odds for receiving a higher proportion of studies (inclusion of only RCTs and offering individual authorship) (Nevitt et al., 2017), our IPD retrieval rate was poor. One approach that may have increased collaboration is the inclusion of a complete DTA in initial contact emails (Polanin and Terzian, 2019). Addressing concerns such as storage and accessibility of the data set and assurances that the data would not be used for

purposes beyond the described protocol in the initial contact email has been shown to associate with increased willingness to share IPD (Polanin and Terzian, 2019). Further methods for improving IPD retrieval need to be developed and implemented.

In order to truly facilitate sharing IPD, the medical research community requires not only a shift in attitude, but also the provision of appropriate resources and information (Ohmann et al., 2017). Increasingly, recommendations highlighting the importance and practicalities of data sharing are becoming available for data custodians (Ohmann et al., 2017). Researchers are urged to develop plans for sharing the de-identified IPD in parallel with the development of the study protocol (Ohmann et al., 2017). Appropriate research infrastructure, policies, and consent should be planned prospectively to facilitate the process (Ohmann et al., 2017).

From personal experience, a commonly held misbelief is that IPD cannot be shared unless explicit consent for this is obtained for all original trial participants. This belief mistakenly disregards a central concept of IPD sharing, which is that it relates to *de-identified* or *anonymised* trial datasets. Regulations, such as the General Data Protection Regulation therefore do not apply to IPD (Recital 26) (GDPR, 2016). Practical guidance for anonymising trial datasets may be provided to trial authors to facilitate the process (Keerie et al., 2018).

4.4.3.2 Statistical considerations for IPD MA

The present work echoes the literature in its preference of the one-stage over the two-stage IPD MA approach (Debray et al., 2015). The application and specifications of the one-stage approach were more flexible and generated more precise effect estimates. Model specifications for the present work were chosen to reflect the two-stage approach, but such specifications need not necessarily be made in the future. Use of the one-stage approach was, however, more complex and required greater statistical consideration and advice. In addition, it did not easily translate to the presentation of forest plots or measures of heterogeneity that are comparable to CMAs. More work

is required to develop packages that facilitate the interpretation and presentation of one-stage IPD MAs

Examination of alternative approaches to pooling outcomes measured on different scales may be examined. An approach that has been suggested to retain the measurement properties of the scale is the use of standardised z-scores (Weitz et al., 2017). Alternatively, one can dichotomise the scales by using tool-specific MCID thresholds. This would, additionally, yield results that may be more clinically interpretable.

Data from repeated measures was analysed in a three-level one-stage IPD MA, however, methodological guidance on such methods are scarce. This extension allows the simultaneous analysis of all available follow-up data, which is invaluable in understanding the overall efficacy of a treatment. Choosing only one time point to examine may not give an accurate reflection of the treatment effect. Participants are, after all, concerned with treatment efficacy on an ongoing basis, not just limited to one time point. To our knowledge, this is the first IPD MA in OA to use three-level multilevel modelling to account for repeated measures in RCTs. Further statistical development of these methods is required.

Similarly, guidelines for the examination of the overall treatment effect and its prognostic factors are not available. Methods were developed for the present work, but may need to be refined. For example, following statistical advice, the covariate of interest was not centred to the study mean. Similarly, only one term was used for the covariate in the model, whilst the interaction models used a within-study and across-study term. Whether or not separate within- and between-study covariate terms should be included in the model could be examined.

Finally, the identification of responsive subgroups using IPD need not be constricted to traditional IPD MA methods. The use of machine learning as an alternative approach for precision medicine has been suggested (Deo, 2015) and methods for its implementation to randomised trial data are being developed (Buturovic, 2018). Perhaps these methods could be expanded and applied in an IPD MA setting.

4.4.3.3 Presentation and appraisal of IPD MA

As for aggregate data MAs, the confidence in the IPD MA output is dependent on the quality of the evidence analysed. The GRADE working group provide recommendations for appraising study evidence in a CMA and NMA setting (Balshem et al., 2011, Brignardello-Petersen et al., 2018, Puhan et al., 2014). However, no such recommendations are available for IPD MAs. One approach adopted has been to apply the traditional GRADE guidelines to the IPD MA, however, this may omit important quality considerations that relate specifically to IPD MAs. The quality assessment should include elements that are of importance to IPD MAs, however, the GRADE working group recommend against the modification of their guidelines (Guyatt et al., 2011a). As such, there is a need for the GRADE working group to provide an extension to GRADE guidelines that can be used to provide standardised assessments of the quality of evidence generated from IPD MAs. Elements specific to IPD MAs that should be addressed include:

- a) The recognition that traditional domains for risk of bias may not apply to IPD MAs in the same manner as CMAs. The risks of bias in domains such as selective outcome reporting and incomplete accounting for patients (failure to use ITT) may be minimised by IPD MA authors in their re-analysis and presentation of the IPD. Considerations of risks of bias should therefore focus on domains that cannot be controlled for or minimised in the IPD MA.
- b) The combined consideration of publication bias and data availability bias. Publication bias may affect IPD MAs and should be explored. Unpublished IPD may be incorporated in IPD MAs and this approach would increase the confidence in the IPD MA estimate. A more pertinent bias relating to a potentially biased subgroup of studies being analysed is data availability bias and this should be explicitly examined when the quality of evidence is appraised.
- c) The development of recommendations specific to the assessment of inconsistency in IPD MAs. This is pertinent particularly to one-stage IPD MAs that do not readily generate the heterogeneity measures described and utilised in GRADE guidelines, such as I^2 .

d) The additional considerations for indirectness that relate to IPD MAs. The comparison of raw data on baseline characteristics may facilitate the assessment of differences in populations. Rather than relying on summary data extracted from publications, such population differences can be directly observed using participant-level raw data. The indirectness of outcome measures should be assessed as for a CMA, however, the measurement of baseline characteristics should also be considered if these are used for interaction modelling. For example, it may be that surrogate markers for the characteristic of interest are used in the IPD MA which would not be applicable to clinical care.

5 CHAPTER 5: Individualising care and identifying predictors of response to topical NSAIDs and capsaicin in OA

5.1 Overview

The following chapter describes the pilot n-of-1 trial series of topical ibuprofen and capsaicin in participants with knee OA.

The study-level MAs established that licensed doses of topical NSAIDs and capsaicin are both effective for pain relief in OA. Furthermore, the NMA indicated that the treatments provide, on average, similar levels of pain relief. However, the aforementioned results relate to study-level average effects and the goal of precision medicine is to move from average effects to individualised care. N-of-1 trial series offer an excellent opportunity to implement precision medicine by examining treatment effects for each person individually. Furthermore, they allow the examination of potentially important predictors of response that are not available for analysis in IPD MAs. For example, because of the mechanisms of action of topical NSAIDs and capsaicin, SH and neuropathic-like pain were of particular interest. However, these are rarely measured in published trials.

Our hypothesis was that variation in response to treatment would be observed between individuals and that topical NSAIDs would be more effective for people with peripheral inflammation (increased SH), whilst capsaicin would be more effective for people with neuropathic-like pain.

Each n-of-1 trial conducted in the series was primarily used to establish which treatment was better for that person. The n-of-1 trials were then aggregated into a series to determine whether differences existed in the response to the treatments (primary aggregate analysis) and to explore the reasons for the differences in response (exploratory analyses). N-of-1 trials have not been widely implemented for this purpose in OA. The trial was planned as a feasibility study in this respect. Feasibility features examined in the series included recruitment strategy, follow-up rates, duration of intervention, number

of treatment cycles, and measures of outcome assessment. In order to generate output applicable to the aims of the present thesis, the trial series was designed and conducted as a randomised feasibility study (Eldridge et al., 2016). It is acknowledged that the trial *series* was not powered as a definitive study for the examination of predictors of response. However, each n-of-1 trial was regarded as a definitive trial for that person.

5.2 Methods

5.2.1 Research ethics, registration, and funding

N-of-1 trials are not generally considered clinical trials, but rather individualised studies done as part of clinical care. However, when aggregated into an n-of-1 trial series, they may or may not be classed as clinical trials depending on the purpose for pooling the results. Following correspondence with the Medicines and Healthcare products Regulatory Agency (MHRA), the present pilot n-of-1 trial series was classified as a non-CTIMP (not a clinical trial of an investigational medicinal product).

Research ethics approval was granted by the Faculty of Medicine & Health Sciences Research Ethics Committee (reference number B10022017) (Appendix 15 – Pilot n-of-1 trial series UoN ethics approval). The final approved protocol, participant information sheet, and consent form are available online at the address below:

<https://www.nottingham.ac.uk/research/groups/osteoarthritisandcrystalarthritis/studies/index.aspx>

The trial was registered on the clinicaltrials.gov portal (NCT03146689) and was conducted in compliance with the guidelines of the Declaration of Helsinki (2013 revision). The research was funded by Nottingham University Hospital Charity.

5.2.2 Trial design

Twenty-two participants with radiographic, painful knee OA were recruited to this open label pilot n-of-1 trial series.

Participants were enrolled to an n-of-1 trial consisting of a maximum of three treatment cycles. Each cycle involved treatment periods of 5% ibuprofen gel and 0.025% capsaicin cream in a randomly determined order (Figure 5-1). The duration of the treatment periods was four weeks and all treatment periods were separated by a washout period. The washout periods were individualised to each participant and were variable in length (maximum four weeks or until the participant felt their pain had returned to their usual level). Trial duration was maximum 44 weeks, however, this varied by the length of washout and treatment response.

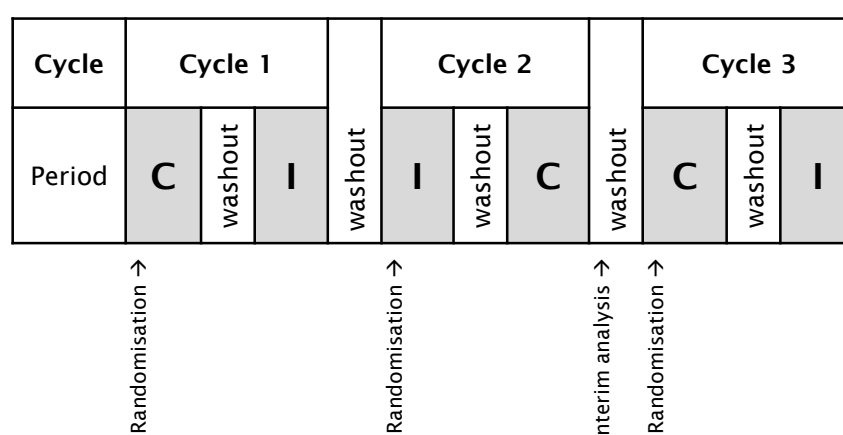


Figure 5-1 - N-of-1 trial design showing a hypothetical random treatment sequence for capsaicin cream (C) and ibuprofen gel (I)

An interim analysis was conducted for each participant at the end of cycle two to determine whether or not they progressed into cycle three. Those who met the pre-specified response criteria (i.e., showed the same treatment preference in cycles one and two) were given the option to complete the study at this point. Those who had different results in the first two treatment cycles or wanted to continue despite meeting the response criteria were randomised and entered the third treatment cycle.

Each participant, having tried both treatments in a random sequence, formed their own individual n-of-1 trial. The results of the trials were then aggregated across 22 participants to form an n-of-1 trial series.

5.2.2.1 Randomisation and treatment allocation

Participants were randomised at the beginning of each treatment cycle to determine the treatment sequence: 1) ibuprofen gel followed by capsaicin cream, or 2) capsaicin cream followed by ibuprofen gel. Each participant was therefore allocated, with equal probability, to one of the eight sequences shown in Table 5-1.

Table 5-1 - Possible treatment sequences for each n-of-1 trial

	Cycle 1		Cycle 2		Cycle 3	
	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
Sequence 1	C	I	C	I	C	I
Sequence 2	C	I	C	I	I	C
Sequence 3	C	I	I	C	C	I
Sequence 4	C	I	I	C	I	C
Sequence 5	I	C	C	I	C	I
Sequence 6	I	C	I	C	C	I
Sequence 7	I	C	I	C	I	C
Sequence 8	I	C	C	I	I	C

Randomisation for the paired treatment periods occurs at the beginning of each cycle
C, capsaicin cream; I, ibuprofen gel

Random sequence generation was conducted by Dr Daniel McWilliams using a web-based programme available at www.randomizer.org. He was not involved in participant recruitment, enrolment, assessment, or outcome collection. Randomisation data were kept strictly confidential and treatment allocation was recorded in sequentially numbered, sealed, opaque envelopes. Participants were sequentially allocated to the next assigned envelope at the beginning of the treatment cycle. The order was determined by the order of entry to the cycle, not by the order of enrolment to the trial. No envelope was taken out of order or opened prematurely.

5.2.2.2 Blinding

The participants and research staff were not blinded to the treatment.

Application of capsaicin cream may result in a warming sensation and erythema of the region. In order to effectively blind capsaicin, the control treatment should include an agent that mimics this reaction. The addition of such an ingredient to the topical NSAID was not felt appropriate as a) patients would not have this available to them in clinical practice, and b) the effect of the preparation would not be limited to that of topical NSAID alone. In addition, whilst Zacin 0.025% is a white cream, topical NSAIDs are not readily available in cream form.

5.2.3 Study participants

Study participants were recruited from the Nottingham Knee Pain and Related Health in the Community Study (KPIC) cohort (Fernandes et al., 2017). KPIC is an observational, prospective study conducted within Academic Rheumatology, UoN. It is briefly described below.

Individuals over the age of 40 years who were registered to general practices in Nottinghamshire were identified. Those with terminal illnesses, severe dementia, severe mental illnesses, or who were otherwise unable to give informed consent were excluded.

Approximately 40,000 postal questionnaires were sent out and 9,506 were returned at baseline. A second questionnaire was sent out one year later to 6,716 participants who consented to receiving a further questionnaire. Of these, 4,738 questionnaire responses were received. The questionnaires covered a diverse set of questions on basic demographics, knee pain, and related health.

Of those that returned postal questionnaires, a subset were invited to attend the department for clinical assessment at baseline, year one, and as part of other sub-studies. Clinical assessments included US and radiographic assessments. Using the data collected in the postal questionnaire and clinical assessment, individuals with radiographic, painful knee OA who consented to being contacted about future studies were recruited to the pilot n-of-1 trial series. The study eligibility criteria are outlined in Table 5-2.

Table 5-2 - Pilot n-of-1 trial series participant eligibility criteria

Eligibility criteria

Inclusion criteria

Men and non-pregnant women (pre-menopausal women must be on an acceptable contraceptive)

Aged 40-95

Knee pain between 4-8 on 0-10 NRS

Radiographic, painful knee OA (from KPIC data*)

Exclusion criteria

Inability to give informed consent

Daily use of oral NSAID for the last two weeks

Prior regular use of ibuprofen gel or Zacin cream on the affected knee(s)

Terminal or untreated major mental illness

Pregnancy or breastfeeding

Hypersensitivity or allergy to topical NSAIDs, capsaicin, or other ingredients in the preparations. This includes asthma attacks, urticaria, or acute rhinitis precipitated by NSAIDs

Current treatment for stomach or duodenal ulcers

Total joint replacement of affected joint

Renal failure

Taking anticoagulants

* Defined as individuals with knee pain (pain in and around the knee on most days for at least a month) and radiographic changes (definite joint space narrowing and definite osteophytes in the tibiofemoral and/or patellofemoral compartments) KPIC, Nottingham Knee Pain and Related Health in the Community Study; NRS, numeric rating scale; NSAID, non-steroidal anti-inflammatory drug

Initially, recruitment was aimed at individuals with predominantly neuropathic-like or predominantly inflammatory phenotypes (based on KPIC data). A predominantly neuropathic-like phenotype was determined by PDQ scores ≥ 13 and SH on US of < 4 mm. Inversely, a predominantly inflammatory-phenotype was determined by PDQ <13 and SH ≥ 4 mm. This hypothesis-driven recruitment approach was chosen in order to facilitate examination of treatment responses according to these two *a priori* defined potential predictors. However,

within the limited timeframe, resources available, and KPIC population, recruitment targets could not be met using these criteria. PDQ and SH threshold criteria were therefore dropped partway through recruitment. Subsequent participants were recruited to the trial irrespective of their SH and PDQ scores.

Eligible KPIC participants were contacted via postal letter, including a patient information sheet (PIS). Those that returned a ticked slip (using a pre-paid envelope) indicating that they would like to be contacted about the study were telephoned. Eligibility criteria were confirmed over the telephone, and those that were potentially eligible were invited to the department for an initial appointment. At the appointment, eligibility was re-established, the study was explained, and any questions were answered. If the person was eligible and interested, written informed consent was taken and they were enrolled in the study.

5.2.4 Interventions

5% w/w ibuprofen gel (Care, Thornton and Ross Limited) and 0.025% w/w Zacin (capsaicin) cream (Cephalon UK Limited) were used for the pilot n-of-1 trial series. Private prescriptions were issued by a General Medical Council registered physician (Professor David A Walsh) and dispensed in the outpatient Nottingham University Hospitals pharmacy.

Participants applied the medications four times per day to their painful knee(s). The recommended dose was a pea-sized amount of the capsaicin cream and an inch of the ibuprofen gel. Participants were recommended to gently rub the gel/cream in until it was no longer visible and to wash their hands carefully after application.

In order to standardise the contextual effects across participants, participants were all given the same explanation for the mechanism of action of the drugs:

"Capsaicin is the substance that makes chilli peppers spicy. Initially, application of capsaicin gives a warming or burning feeling, however with continued application it dampens down the pain signalling in and around the joint.

"Topical non-steroidal anti-inflammatory drugs are a group of drugs, such as ibuprofen and diclofenac. They act by reversibly blocking the production of substances that would otherwise cause pain and inflammation in and around the joint."

5.2.4.1 Concurrent therapies

Participants continued to use their regular medications, including oral analgesics, throughout the trial provided the frequency/dose had been stable for three months. For analgesics that were taken as required, participants were asked to maintain a similar routine throughout the study as they were on prior to enrolment. Non-permitted concomitant therapies were topical analgesics for the affected knee, regular oral NSAIDs, and arthroplasty. Participants were advised not to commence new therapy, including medications, exercise therapy, orthotics, or alternative therapies, during the trial.

5.2.5 Baseline and outcome assessments

5.2.5.1 Baseline measures

Data on potential predictors of the treatment response were assessed at baseline, prior to randomisation (Table 5-3). In order to avoid unnecessary radiation exposure and under the assumption that radiographic changes were unlikely to rapidly progress, baseline radiographic data were extracted from the KPIC study database (i.e., from radiographs obtained at varying time-points within the previous two years). All other baseline data were collected via questionnaire and clinical assessments at the baseline study visit.

Table 5-3 - Potential predictors of treatment response and their method of assessment for the pilot n-of-1 trial series

Predictor	Assessment tool
Questionnaire assessments	
Demographics	Age, sex
Knee pain intensity	NRS
Neuropathic-like pain	Modified PDQ
Function	ADL dimension of KOOS
Illness perception	Brief illness perception questionnaire
Expectation of treatment	Brief illness perception questionnaire
Anxiety	HADS
Depression	HADS
Fibromyalgia	Fibromyalgia screening questions
Central Mechanisms Traits	Various*
Clinical assessments	
Inflammation	US (SH, effusion, and Power Doppler)
Abnormal pain processing	QST – PPT and TS
Quadriceps muscle strength	Nicholas manual muscle tester
Obesity	BMI
Radiographic severity	NLDA and KL

*Eight questions from the original KPIC questionnaire that have been shown to indicate centrally augmented knee pain (Akin-Akinyosoye et al., 2018)
 ADL, activities of daily living; BMI, body mass index; HADs, Hospital Anxiety and Depression Scores; KOOS, Knee Injury and Osteoarthritis Outcome Score; NLDA; Nottingham Line Drawing Atlas; NRS, numeric rating scale; PDQ, painDETECT questionnaire; PPT, pressure pain threshold; SH, synovial hypertrophy; TS, temporal summation; US, ultrasound

5.2.5.1.1 Questionnaire assessments

Participants filled in a questionnaire at baseline. The design and content of the questionnaire was based on the KPIC baseline and year one questionnaires. Along with basic information on participant demographics, medical history, and drug history, the questionnaire also assessed the participant's baseline knee pain status. In addition, it encompassed various validated questionnaires evaluating potential predictors of the treatment response.

Pain intensity was determined using 0-10 NRS in response to the following question: “In the past week, on average, how intense was your knee pain rated on a 0-10 scale, where 0 is “no pain” and 10 is “pain as bad as could be””. NRS was chosen as the tool of choice as it is easy to implement via text messaging, over the telephone, and via questionnaire or participant diaries (Hawker et al., 2011). Baseline and outcome pain intensity data could therefore be easily collected using the same tool, regardless of the participant’s location and preference. In addition, it is often preferred by participants over alternative measures of pain intensity, such as VAS (Hawker et al., 2011). Finally, the tool has high test-retest reliability and construct validity (high correlation with VAS pain) (Hawker et al., 2011).

The presence of neuropathic-like pain was assessed using a modified PDQ (Freynhagen et al., 2006). Within the PDQ, various questions assess the nature of the pain experienced. The scores are summated and the total score indicates the likelihood of the participant’s pain being neuropathic-like in nature. Higher PDQ scores indicate a greater likelihood of the participant experiencing neuropathic-like pain. The thresholds of $PDQ \geq 13$ and $PDQ \geq 19$ were used to identify participants with possible or definite neuropathic-like pain, respectively. The PDQ scores were Rasch-transformed for the multilevel modelling in order to convert the scores into interval-level data with improved goodness of fit (Moreton et al., 2015).

Knee-related functional impairment was assessed using the ADL domain of the KOOS questionnaire (Roos and Lohmander, 2003). The ADL domain assesses the difficulty experienced when undertaking a range of ADLs, including rising from sitting and taking off socks. Higher scores indicate less functional impairment.

The participant’s illness perception and expectation of treatment were assessed using a modified Brief Illness Perception Questionnaire (Broadbent et al., 2006). The questions assess the participant’s perception regarding the impact of knee pain on their life, its likely duration, and how concerned they are by it. In addition, participants

are asked how much they think that the treatment will help their knee pain. Higher scores indicate a more threatening view of the illness.

The presence of anxiety and/or depression were assessed using the Hospital and Anxiety Depression Scale (HADS) (Zigmond and Snaith, 1983). The presence of anxiety and/or depression was determined using a pre-defined threshold of eight points (Bjelland et al., 2002).

The presence of fibromyalgia was assessed using screening questions covering two domains: 1) the number of painful areas (widespread pain index, WPI) as indicated in the body pain manikin and 2) the severity of symptoms such as fatigue (symptom severity, SS). The fibromyalgia severity (FS) score was calculated through the summation of the WPI and SS scores. In addition, participants were determined to likely have fibromyalgia if they met the following criteria (Wolfe et al., 2016):

1. Pain is present in at least four of five regions
2. Symptoms have been present for at least three months
3. $WPI \geq 7$ and $SS \geq 5$ or WPI 4-6 and $SS \geq 9$

Finally, self-reported items from the KPIC questionnaire that have been shown to indicate central pain augmentation (Central Mechanisms Traits) were included (Akin-Akinyosoye et al., 2018). The eight traits represented in the questionnaire were: anxiety, depression, catastrophising, neuropathic-like pain, fatigue, sleep disturbance, pain distribution, and cognitive impact. Validation and rewording of the Central Mechanisms traits questionnaire is currently underway. The scores for each trait were transformed to a 0-3 scale and summated. Higher scores indicate greater likelihood of central pain mechanisms.

5.2.5.1.2 Clinical assessments

Unless otherwise stated, clinical assessments were conducted by Dr Joanne Stocks in accordance with protocols used by the KPIC cohort study (Fernandes et al., 2017). Prior to conducting the assessments, these were explained to the participants. Where the participant enquired, the results of the assessments were explained (e.g., what was being assessed, if their values were higher/lower than usual, how they compared to the putative maximum/minimum range). It was

stressed that the relationship between the assessments and knee pain was not straightforward and the findings, as such, did not indicate OA severity or otherwise grade their condition. Interpretation of the results, such as statements like "this means you have inflammation in the joint", were generally avoided.

Anthropometric measures, in the form of BMI, were calculated by measuring height and weight.

The presence of inflammation in the knee was assessed using US by a trained physician, Dr Aliya Sarmanova (Sarmanova et al., 2017). US examination was conducted using a multi-frequency (7-12 MHz) linear array transducer. Due to a planned upgrade, two equipment were used during the study (Toshiba Aplio SSA-770A and Toshiba Aplio SSA-780A machines), however, agreement between the two machines was very good (Table 5-4).

Table 5-4 - Inter-machine agreement for the presence of US features

	Effusion	SH
Unweighted kappa	1.0	1.0
Concordance coefficient	0.97	0.99
(95%CI)*	(0.94 to 1.00)	(0.98 to 1.00)
MD in mm (old-new)	0.36	0.07
(SD, 95% limits of agreement)	(1.14, -1.86 to 2.59)	(0.66, -1.23 to 1.37)

Inter-machine agreement for one observer and two assessments conducted on the same day

*Lin's concordance correlation coefficient, test performed using the asymptotic point estimate and variance. P-value for effusion and SH $p < 0.0001$

CI, confidence interval; MD, mean difference; SD, standard deviation; SH, synovial hypertrophy

The supra-patellar recess and medial and lateral tibiofemoral spaces of both knees were assessed with knee flexion of approximately 20-30°. Changes were defined according to the definitions accepted by the OMERACT-7 Group (Wakefield et al., 2005). The maximal synovial thickness and effusion depth along the longitudinal axis were recorded in millimetres. As per EULAR recommendations, the presence of effusion and SH were dichotomised as absent ($< 4\text{mm}$) or present

($\geq 4\text{mm}$) (D'Agostino et al., 2005). The presence of a positive Power Doppler signal, an indicator of increased vascularity, was recorded.

Abnormalities in pain processing were evaluated using PPT and TS. PPTs were assessed using a hand-held pressure algometer (Somedic AB, Sweden) connected to a computer (HP ProBook 4520s). Increasing amounts of pressure were applied using the algometer until the participant indicated that the sensation changed from pressure to pain. Seven pre-specified bony anatomical sites were assessed per cycle of PPT testing. The sites assessed were: the sternum, the medial joint line bilaterally, the lateral joint line bilaterally, and the proximal tibia bilaterally. Three cycles were conducted and the average score of the cycles was used for the analysis. The site used for local, distal, and remote PPTs were the medial joint line in the index knee, ipsilateral proximal tibia, and sternum, respectively. PPTs over bony surfaces were used as they have been shown to be reproducible and are recommended for experimental tests of evoked bone-associated pain (Fernandes et al., 2017).

TS was measured using a 256 millinewton weighted pinprick stimulator. The pinprick was applied to the infero-medial border of the patella of the most affected knee. The participant was asked to rate the evoked feeling of pain or sharpness using a 0-100 NRS (0 = no pain, 100 = most intense pain or sharpness). Ten consecutive pinpricks were then applied to the same area at a rate of one pinprick per second, and the participant was asked to rate the pain or sharpness again using 0-100 NRS. The difference between the repetition and baseline scores were calculated (repetition minus baseline score). This process was conducted twice and the MD in repetition-baseline scores was calculated. Unlike the original KPIC protocol where a ratio between the scores was used, the MD was chosen as per Kurien et al. (2018). Using the difference, rather than ratio, between the scores yielded a less skewed measure and allowed the use of baseline scores of zero.

The maximum voluntary quadriceps contraction was measured using the Nicholas Manual Muscle Tester (Lafayette Instruments). Whilst sitting comfortably with their knee bent at 90° and the instrument held

against their shin, the participant was asked to push against the instrument as hard as possible in an attempt to straighten the leg. Three measures were taken and the average of these was used.

The severity of radiographic changes were scored by Dr Gwen Fernandes using the Nottingham Line Drawing Atlas (NLDA) (Wilkinson et al., 2005). Total NLDA scores and individual scores for JSN and osteophytes were used. For sensitivity analyses, radiographic severity was also graded using KL scores (Kellgren and Lawrence, 1957). Plain film radiographs of the tibiofemoral and patellofemoral compartments, taken using standard weight-bearing fixed flexion posterior-anterior and skyline 30 degree flexion views, were used from the KPIC database. The most affected compartment was determined using KL tibiofemoral and patellofemoral grading.

5.2.5.2 Outcome measures

Pain intensity was assessed using 0-10 NRS. Participants were asked for average pain intensity scores via telephone or text messaging at the baseline and end of each treatment period. Change in pain scores over the treatment period were analysed as the primary outcome. This was used to determine the preferential treatment response, overall treatment effect, and treatment difference.

Secondary outcomes were participant's end-of-cycle treatment preference for pain relief, their end-of-study overall preference, and weekly NRS pain intensity scores. Participant's treatment preference for pain relief was elicited at the end of each treatment cycle by asking the following question: "Considering only the pain relief experienced in this most recent cycle, which treatment do you feel provided satisfactory pain relief?". The participant's overall treatment preference was elicited at the end of the study using the following question: "Considering all the aspects of the treatment, including its effectiveness and ease of application, which treatment do you prefer?". Treatment preference options were ibuprofen, capsaicin, both, or neither. For analyses, responses of both and neither were combined to indicate no treatment preference. Finally, participants were also asked

to record their pain intensity (0-10 NRS) during treatment periods on a weekly basis in the participant diary.

5.2.5.3 Withdrawals and incomplete data

Efforts were made to minimise the amount of missing data and withdrawals through trial design. At the baseline department visit, questionnaires were checked to ensure all questions had been answered. Any questions that had been missed, most often by oversight or uncertainty, were addressed in conjunction with the participant. To avoid missing outcome data, they were collected in the format most convenient for the participant. Text messaging was the favoured method as this allowed data collection even when participants were busy or on holiday. A participant diary was used in parallel, which permitted the collection of outcome data when it was mutually agreed that telephone or text contact were not possible during that period. Trial design and duration was tailored to the individual, with variable washout periods and an interim analysis, in order to maximise participant retention. Finally, rapport was established with participants in order to further reduce drop-outs.

Where a participant withdrew during a treatment period, the NRS pain score at the point of withdrawal was used in the analyses. In addition, participants that withdrew were asked for their overall and pain-specific treatment preferences at the point of withdrawal and this was used in the analyses. Preferential treatment responses, based on NRS, were determined using only completed treatment cycles. When pooling treatment periods across all participants, all treatment periods for which data were available were used.

5.2.6 Sample size justification

The n-of-1 trials consisted of three treatment cycles (six paired periods). Three cycles were chosen as this is the commonest number of cycles used in n-of-1 trials in OA and other conditions (Yelland et al., 2007, March et al., 1994, Gabler et al., 2011). After the first few cycles, each additional cycle contributes little to the precision of the trial (Zucker et al., 2010). Three cycles were therefore felt to provide

sufficient data without lengthening the trial beyond 44 weeks, thereby maintaining trial feasibility and participant retention.

Having established the number of treatment cycles within each n-of-1 trial, sample size calculations were conducted to determine the number of treatment periods required in the trial series. The trial series was powered to examine preferential treatment responses across trial participants. Based on a MCID of 0.5 SD between topical NSAID and capsaicin, the sample size was calculated assuming 80% power and 0.05 significance level. For this, 11 participants, generating 33 paired comparisons (three cycles), were required. Assuming only 50% of participants show a preferential response between the treatments, 22 participants were required. Withdrawal rates of 20% have been reported for n-of-1 trial series, and this was factored into the 50% that do not show a preferential response (Gabler et al., 2011).

The trial series was not powered for the exploratory multilevel regression modelling. However, assuming participants underwent an average of two treatment cycles, 88 change-from-baseline pain scores would be available for the regression modelling. This was felt to be sufficient to provide an indication of the feasibility of using n-of-1 trial series to examine predictors of response and provide some hypothesis-generating results.

5.2.7 Statistical analysis

Baseline characteristics of the study population were presented descriptively. Continuous variables were examined for normality through visual examination of the histograms. Variables for which the histogram markedly deviated from normality were described using the median and IQR. Normally distributed data were described using mean and SD.

A difference of 1 point on the 0-10 NRS was determined to be the threshold for a MCID. This is a conservative estimate of the MCID (0.5 SD) used by NICE OA guidelines (NICE, 2014). The MCID of 0.5 SD is equal to 1.25 cm on 0-10 cm VAS, or approximately 1 point on 0-10 NRS (Wandel et al., 2010). This threshold reflects a “slightly better”

change in status (Salaffi et al., 2004). Statistical significance was set at $p < 0.05$.

Statistical analysis was conducted first at the individual level to determine the preferential treatment response. Results of the n-of-1 trials were then aggregated across the series to examine the treatment effects across all trial participants. Exploratory analyses were then conducted to examine for baseline characteristics that associated with treatment preference, the overall effect of topical therapy, or response to ibuprofen or capsaicin.

Multilevel model specifications common to all analyses were: random intercepts to account for clustering, common variance structure, and fixed treatment effects (when a treatment term was included). These specifications provide the best model fit for multilevel analyses of n-of-1 trials (Zucker et al., 2010). Clustering at the participant-level was undertaken to account for intra-subject correlation.

5.2.7.1 Treatment responses at the individual level

Responses to treatment, per cycle, were plotted for each participant. Preferential treatment responses were determined by calculating the change from baseline pain scores for each treatment period. The difference in change scores between the two treatments were used to determine the cycle preferential response. A difference greater than 1 point between the treatments was used to indicate a cycle preference for the treatment with the greater change score.

Where a participant displayed a preference in favour of a treatment for two or more cycles, this was determined to be their overall preference for the preferential treatment response. If the treatments provided equal pain relief in at least two cycles, or where all three cycles showed different preferences (e.g., 1 for ibuprofen, 1 for capsaicin, 1 with no difference), the participant was deemed to have no treatment preference. Where the participant withdrew prior to meeting the above criteria, their preferential response was deemed "unclear".

5.2.7.2 Aggregated treatment responses

Treatment responses were aggregated across the n-of-1 trial series. The proportion of participants per preferential treatment response category were calculated. Then, the overall treatment effect of the drugs and the difference between the drugs were examined across all treatment periods.

The overall treatment effect (absolute change from baseline) was examined separately for each drug in a two-level regression model, including only periods where that drug was used. Change scores, per period, were the dependent variable. As only one treatment was examined at a time, a treatment term was not included. Data were clustered at the participant-level and the overall treatment effect of the drug was determined from the model intercept. As per the overall treatment effect analyses in the IPD MA (section 4.2.2.1.2.3), the model was not adjusted for baseline pain.

The difference between the treatments was compared in a three-level model, accounting for clustering at the cycle- and participant-level. End of period pain scores were the dependent variable and adjustment was undertaken for period baseline pain scores. A fixed treatment effect was assumed. A treatment effect greater than 0 favoured ibuprofen and less than 0 favoured capsaicin. A treatment effect of 0 indicated no difference between the treatments.

5.2.7.3 Examination of predictors of response

Exploratory analyses were conducted for associations between baseline characteristics and treatment responses. Where possible, baseline characteristics were examined as continuous variables. Associations were examined between baseline characteristics and 1) preferential treatment response, 2) overall response to topical therapy (regardless of drug), and 3) interaction with treatment.

First, associations between preferential treatment responses and baseline characteristics were examined in all participants where a preferential treatment response was established (i.e., excluding drop-outs). A multinomial logistic regression model was used, where

preferential treatment response (no preference, ibuprofen, or capsaicin) was the dependent variable and the participant baseline characteristic was the independent variable. One potential predictor was examined per model.

Second, significant prognostic factors for a response to any topical treatment were examined. All treatment periods (regardless of treatment) were included, and no treatment term was specified. Clustering was accounted for at the participant-level only as model fit was not improved when clustering at the cycle-level. Furthermore, clustering at the cycle-level was deemed unnecessary as comparisons between treatments (within each cycle) were not conducted. Change scores were the dependent variable and each potential prognostic factor was examined in univariate analysis. As per the IPD MA (section 4.2.2.1.2.4), the model was not adjusted for baseline pain. Significant prognostic factors were combined in multivariable analysis.

Finally, interactions between baseline characteristics and treatment were examined. For this, a three-level model was built as specified for the examination of the difference between treatments. End of period pain scores were the dependent variable and adjustment was undertaken for period baseline pain scores. Fixed treatment, baseline characteristic, and treatment-by-characteristic interaction terms were the independent variables. The treatment term was coded as ibuprofen =0 and capsaicin =1. To facilitate the interpretation, the p-value for the interaction effect was presented along with the treatment effect (difference in mean NRS between ibuprofen and capsaicin) according to subgroups of the predictor variable (e.g., treatment effect in men and women separately). For continuous predictor variables, the estimated treatment effect is presented for predictor scores of one SD below and one SD above the mean predictor value.

5.2.7.4 Sensitivity analyses

Sensitivity analyses were conducted for assessments of the aggregated treatment effect and predictors of response. Analysis was conducted as specified for the multilevel models above, with any additional changes outlined below.

Trial inclusion criteria were for participants with baseline pain recording 4-8 on the NRS. However, for a subset of individuals, pain levels decreased during the trial despite four weeks of washout between periods. Sensitivity analyses were therefore conducted by restricting analyses to periods where the period baseline pain intensity was 4-8 (inclusive).

Using the pre-determined MCID of 1 point, it was determined whether the change from baseline scores met this threshold per period. The number of periods for which this was met was compared between the treatments using multilevel logistic regression (MCID met/not met). Examination of predictors of response were conducted as previously detailed, except using multilevel logistic regression.

5.2.7.5 Carryover and period effects

Carryover effects were minimised in the study design through the use of washout periods. To further account for carryover effects, analyses accounted for baseline scores (through change scores or adjustment of baseline scores). The potential impact of carryover effects were further reduced in sensitivity analyses restricted to trials where baseline pain was 4-8.

The significance of period effects were tested in multilevel modelling, including fixed treatment, period, and treatment-by-period interaction terms. Likewise, the significance of treatment sequence was examined using a treatment-by-treatment sequence indicator interaction term.

5.2.7.6 Feasibility assessments

In order to provide guidance for future n-of-1 trials in OA, the outcome chosen, frequency of outcome collection, and number of cycles used were examined.

Agreement between treatment preferences were determined using Cohen's unweighted kappa (κ) (Cohen, 1960). The level of agreement was interpreted using the following criteria: 0 to 0.20 none, 0.21 to 0.39 minimal, 0.40 to 0.59 weak, 0.60 to 0.79 moderate, 0.80 to 0.90 strong, and >0.90 almost perfect (McHugh, 2012).

For each cycle, the agreement between preferential treatment responses using NRS criteria (primary) and participant's end-of-cycle treatment preference was assessed. Note that for these analyses, all cycles were considered independent. Next, the median (IQR) difference between treatments was calculated for each participant preference category. This was used to determine how large a difference between treatments was required for a participant to report a treatment preference.

Using the same criteria as for the NRS (≥ 2 cycles in favour of a treatment), the overall treatment preferences were calculated using 1) the participant's end-of-cycle preferences and 2) using their pain diary scores at each week of follow up. Agreement was assessed between the primary preferential treatment response and other methods of ascertainment. To better understand the participant's decision-making, the agreement between their overall end-of-study preference and their calculated preference (based on cycle-responses) was examined.

Weekly pain data, collected from participant diaries, were plotted and presented descriptively. The effect of time on pain relief (regardless of treatment) was examined in a three-level regression model clustered by person and cycle. Using data from the end of each period, the agreement between preferences determined using diary responses and the primary assessment method were examined. In addition, the agreement between preferences at different weeks of follow-up were assessed.

Whether the inclusion of the third treatment cycle was justified was examined. The number of participants in whom a third treatment cycle was required to determine a preference was determined. Finally, sensitivity analysis was conducted for the treatment difference, including only data from the first two treatment periods.

5.3 Results

5.3.1 Study population

Study invitation letters were sent to 122 potentially eligible KPIC participants (Figure 5-2). Responses were received from 77 participants, of whom 41 expressed an interest in participating in the trial. Following telephone contact, 25 were eligible and interested in taking part in the trial. They were invited to the department. One person did not attend the department visit as they received an intra-articular injection of steroid and were no longer experiencing pain. At the department visit, two people did not meet the eligibility criteria (pain level below threshold) and the remaining 22 participants were enrolled.

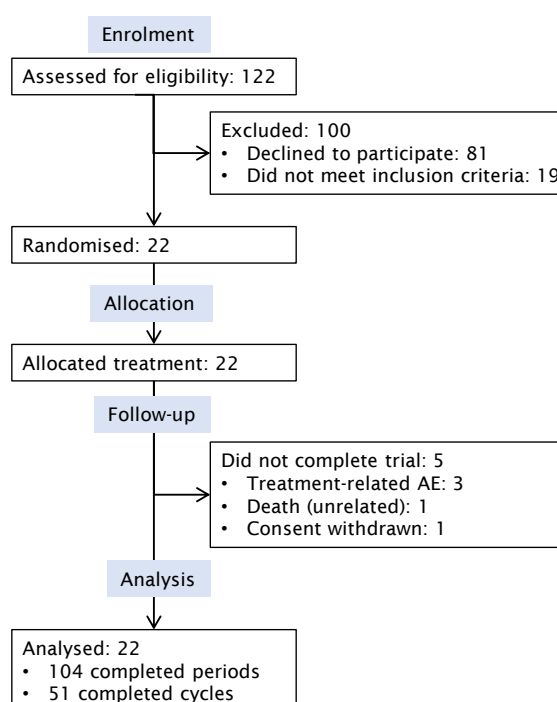


Figure 5-2 - CONSORT flow diagram of participant recruitment and retention in the pilot n-of-1 trial series

All enrolled participants were randomised and received their allocated treatment. The treatment sequence in the first cycle was ibuprofen→capsaicin for 12 participants, and capsaicin→ibuprofen for 10 participants. All 22 participants completed the first treatment cycle. The treatment sequence for the second cycle was ibuprofen→capsaicin for 10 participants, and capsaicin→ibuprofen for 12 participants. Five

participants discontinued during or at the end of the second cycle. Three participants withdrew because of AEs associated with ibuprofen gel use (localised skin rash). One participant withdrew consent following the recommendation by his GP as he was investigated for an unrelated dermatological condition. Finally, one participant died during the washout period prior to period four. Cause of death was ischaemic and hypertensive heart disease. The death was deemed unrelated to the study medications.

Ten participants met the response criteria at the interim analysis and were given the option to complete the study at this point. Eight of them chose to finish the trial after two cycles, whilst two people chose to continue into the third cycle. Nine participants entered and completed cycle three.

In total, the 22 participants completed 104 treatment periods (52 with ibuprofen and 52 with capsaicin) and 51 treatment cycles. Median (IQR) follow-up duration was 25.8 weeks (21.6 to 31.0).

5.3.1.1 Baseline characteristics

Baseline characteristics of all participants are presented in Table 5-5. The trial population was comprised of 12 women and 10 men. Approximately a tenth of participants met diagnostic criteria for fibromyalgia, one fifth met criteria for depression, and one third met criteria for anxiety. The most painful (index) knee was equally distributed between the left and right knee across the trial population. Participants reported a median baseline pain severity of 5.5 (IQR 4 to 7) on the 0-10 NRS. Almost 15% of participants reported features indicative of neuropathic-like pain. On US examination, almost 70% of participants had SH and 14% were Power Doppler positive. Radiographic OA changes (definite osteophytes and definite JSN) were evident in the tibiofemoral and patellofemoral compartments. Of the 21 participants that had tibiofemoral and patellofemoral KL grading, the tibiofemoral compartment was most affected in 14 people, the patellofemoral in four, and the compartments were equally affected in three.

Table 5-5 - Demographic and clinical baseline characteristics in the pilot n-of-1 trial series

Characteristic	All participants
Basic demographics	
Age; mean (SD), years	67.0 (9.3)
Sex; n (%) women	12 (55%)
BMI; mean (SD), kg/m ²	30.7 (5.6)
Questionnaire – comorbidities	
Anxiety; n (%) HADS anxiety subscale ≥ 8	9 (40.9%)
Depression; n (%) HADS depression subscale ≥ 8	4 (18.2%)
Fibromyalgia; n (%) diagnostic criteria met	2 (9.1%)
Questionnaire – OA features	
Index knee; n (%) left knee pain	11 (50%)
Baseline pain; median (IQR), NRS severity	5.5 (4 to 7)
Neuropathic-like pain;	
n (%) with definite NP (PDQ ≥ 19)	3 (13.6%)
n (%) with definite or possible NP (PDQ ≥ 13)	7 (31.8%)
Physical function; median (IQR), KOOS	58.1 (47.1 to 73.5)
Central Mechanisms traits; median (IQR), range (0-24)	9.4 (7.8 to 14.3)
Examination findings (index knee)	
Quadriceps strength; median (IQR), kg	16.1 (14.3 to 21.0)
Radiographic severity	
Total NLDA score; median (IQR)	13 (9 to 18)
NLDA osteophyte score; median (IQR)*	9 (5 to 15)
NLDA JSN score; median (IQR)	4 (3 to 5)
n (%) per tibiofemoral KL grade	
0**	2 (9%)
1	3 (14%)
2	7 (32%)
3	9 (41%)
4	1 (5%)
n (%) per patellofemoral KL grade	
0**	4 (19%)
1	9 (43%)
2	3 (14%)
3	5 (24%)
Continued overleaf	

Characteristic	All participants
US features	
Synovial thickness; mean (SD), mm	5.52 (2.93)
n (%) with SH (SH $\geq$ 4mm)	15 (68.2%)
Effusion; mean (SD), mm	8.89 (3.65)
n (%) with effusion (effusion $\geq$ 4 mm)	21 (94.5%)
n (%) Power Doppler positive	3 (13.6%)
QST features	
Localised PPT (MJL); median (IQR), kPA	393.0 (154.4 to 610.6)
Distal PPT (proximal tibia); median (IQR), kPA	390.6 (200 to 529.6)
Remote PPT (sternum); median (IQR), kPA	280.9 (137.7 to 401.2)
TS; mean (SD), 0-100 NRS	37.9 (24.2)

*One participant was missing skyline views and an aggregated osteophyte score was not calculated. Osteophyte scores in the available graded compartment were ≥ 2

**Participants with KL grading of 0 in one compartment had evidence of OA in the other compartment

BMI, body mass index; HADS, Hospital Anxiety and Depression Scale; IQR, interquartile range; KL, Kellgren-Lawrence; KOOS, Knee Injury and Osteoarthritis Outcome Score; MJL, medial joint line; NLDA, Nottingham Line Drawing Atlas; NP, neuropathic pain; NRS, numeric rating scale; PDQ, painDETECT questionnaire; PPT, pressure pain threshold; QST, quantitative sensory testing; SD, standard deviation; SH, synovial hypertrophy; TS, temporal summation

Examination of PDQ and SH scores revealed that only two participants (9%) had a predominantly neuropathic-like pain phenotype (PDQ $\geq$ 13 and SH<4), whilst ten (46%) had a predominantly inflammatory phenotype (SH $\geq$ 4 and PDQ<13).

5.3.2 Preferential treatment responses

Treatment responses, measured using the 0-10 pain intensity NRS at baseline and week four, are illustrated per treatment cycle for each participant in Figure 5-3.

A preferential treatment response was established within two treatment cycles (interim analysis) in 10 participants. Of the two participants that chose to continue after the interim analysis, the findings of the third cycle were in keeping with those of the first two cycles. Seven participants had conflicting results in the first two cycles and underwent a third cycle after the interim analysis. Two of these participants had different treatment preferences in each cycle and were classed as having no overall preference. For the remaining five participants, the third cycle was in agreement with one of the previous cycles. Five participants withdrew prior to the preferential treatment response being established using NRS criteria. All drop-outs completed at least the first treatment cycle.

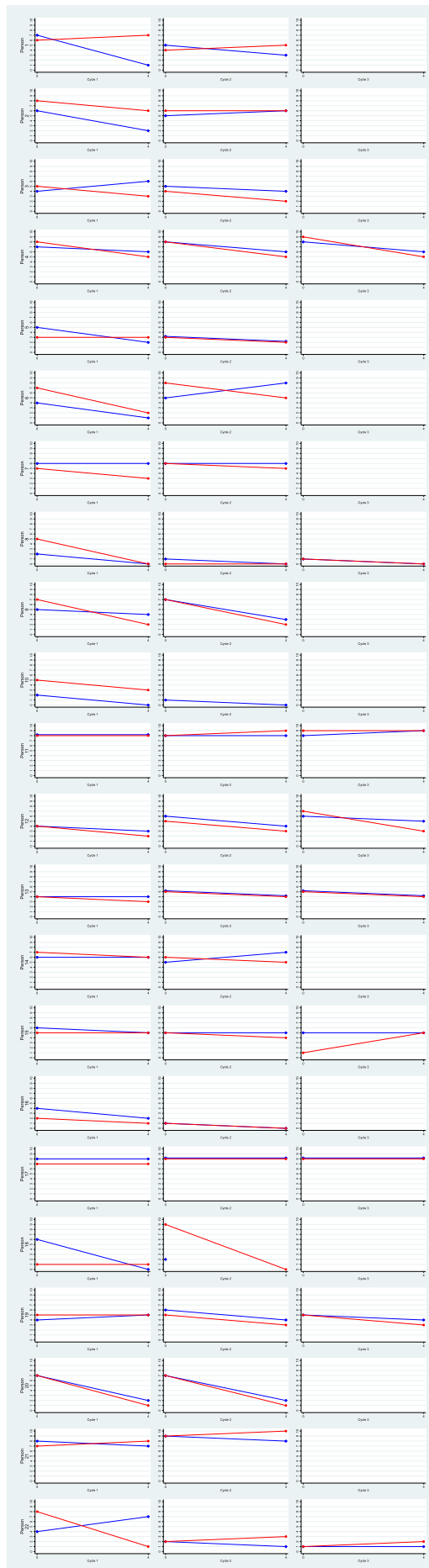


**Figure 5-3 -
Treatment responses
for topical ibuprofen
(blue) and capsaicin
(red).**

0-10 NRS pain
intensity at baseline
and week 4
illustrated per cycle.
Each participant is
represented on one
row. Cycles 1, 2, and
3 are presented in
the first, second, and
third column
respectively

Overall preferential treatment responses were for topical ibuprofen in four participants (18%), capsaicin in nine participants (41%), no preference in four participants (18%), and unclear preference in five participants (23%). Treatment preferences in the first treatment cycle for the drop-outs were for topical ibuprofen in four participants and equal in one participant.

5.3.3 Aggregated treatment effect

The treatment effect was aggregated across 104 completed treatment periods, accounting for clustering at the participant level. The overall treatment effect was 1.18 points (95%CI 0.52 to 1.83) for ibuprofen and 1.64 points (95%CI 0.85 to 2.44) for capsaicin. No significant differences were found between ibuprofen and capsaicin overall (treatment coefficient = -0.29, 95%CI -0.80 to 0.22, $p=0.271$). No period or treatment sequence effects were identified ($p=0.691$ and $p=0.773$, respectively).

A greater variance was observed between-participants than within-participants (participant-level variance: 4.2; cycle-level variance: $3.8e^{-12}$).

Sensitivity analyses were in keeping with the primary analyses.

Including only periods where baseline pain levels were 4-8 points (inclusive) on NRS, no significant difference was observed between the treatments (treatment coefficient = -0.60, 95%CI -1.25 to 0.04, $p=0.068$).

Overall, the level of pain relief met the MCID threshold in 66 of 104 treatment periods (64%). Of these, ibuprofen was used in 33 periods and capsaicin used in 33. The OR for reaching clinically important pain relief for capsaicin compared to ibuprofen was 1.02 (95%CI 0.39 to 2.64).

5.3.4 Predictors of response

5.3.4.1 Preferential treatment response

Associations between baseline characteristics and preferential treatment response were examined in the 17 participants where a

preferential treatment response (ibuprofen, capsaicin, or no preference) was established based on NRS criteria. Three characteristics significantly associated with the preferential treatment response: fibromyalgia severity score ($p=0.030$), quadriceps strength ($p=0.040$), and synovial thickness ($p=0.020$). Neuropathic-like pain did not associate with a treatment preference ($p=0.229$).

Table 5-6 - Distribution of significant baseline characteristics per preferential treatment response group (determined using NRS criteria)

Characteristic	Preferential treatment response		
	Ibuprofen n=4	Capsaicin n=9	No preference n=4
Fibromyalgia severity score; mean (SD)	5.75 (1.26)	10.89 (4.34)	9.75 (4.27)
Quadriceps strength; median (IQR), kg	24.4 (17.6 to 27.9)	16.0 (14.5 to 19.4)	14.9 (12.8 to 15.7)
Synovial thickness; mean (SD), mm	6.48 (1.42)	3.96 (3.16)	8.50 (2.52)

IQR, interquartile range; SD, standard deviation

5.3.4.2 Overall treatment effect (irrespective of treatment)

Three baseline characteristics significantly associated with an improved response to topical treatment in univariate analysis (Appendix 16 – Association between baseline characteristics and overall treatment effect of topical therapy in pilot n-of-1 trial series). Overall, women had 1.27 points (95%CI 0.12 to 2.42) more pain relief on topical therapy than men ($p=0.031$). Those with higher BMI reported greater reductions in pain on treatment (β 0.11, 95%CI 0.00 to 0.22, $p=0.043$). Finally, individuals with more pain at baseline reported greater reductions in pain (β 0.80, 95%CI 0.61 to 0.99, $p<0.001$). However, in multivariable analysis including sex, BMI, and baseline pain as covariates, only baseline pain remained significant (baseline pain: $p<0.001$, sex: $p=0.580$, BMI: $p=0.734$).

Excluding treatment periods where baseline pain levels were not 4-8 NRS (inclusive), sex and baseline pain significantly associated with treatment response (sex: β 1.30, 95%CI 0.01 to 2.59, $p=0.048$; baseline pain: β 0.96, 95%CI 0.59 to 1.32, $p<0.001$) (Appendix 16 – Association between baseline characteristics and overall treatment effect of topical therapy in pilot n-of-1 trial series). In multivariable analysis, only baseline pain remained significant (baseline pain: $p<0.001$; sex: $p=0.335$). Greater pain at baseline was associated with increased pain reduction after treatment.

In analyses based on meeting the MCID, a significant association was found only for physical function (OR 1.05, 95%CI 1.00 to 1.09, $p=0.035$). Participants with better function at baseline were more likely to achieve clinically important reductions in pain on topical therapy.

5.3.4.3 Treatment-by-covariate interactions

Examination for significant interactions between treatment and baseline predictors in all completed treatment periods did not yield any significant findings (Appendix 17 – Treatment-by-baseline characteristic interaction effects in pilot n-of-1 trial series). Neither neuropathic-like pain nor synovial thickness predicted a response to either treatment ($p=0.405$ and 0.101 , respectively).

Limiting analyses only to periods where baseline pain was 4-8 NRS (inclusive), a significant interaction was identified only for distal PPT ($p=0.049$) and TS ($p=0.020$) (Appendix 17 – Treatment-by-baseline characteristic interaction effects in pilot n-of-1 trial series). The treatment effect favoured capsaicin more at high distal PPT scores (β -1.14, 95%CI -1.98 to -0.31) than at low distal PPT scores (β 0.06, 95%CI -0.85 to 0.96, respectively). The treatment effect also favoured capsaicin more at lower TS scores (β -1.41, 95%CI -2.34 to -0.48) compared to higher TS scores (β 0.15, 95%CI -0.75 to 1.05). For each of these, the interaction indicated that greater pain relief was observed for capsaicin, relative to ibuprofen, when higher distal PPTs and lower TS were recorded (each indicative of less pain sensitivity).

Using the MCID as the outcome, a significant interaction was observed only between treatment and synovial thickness ($p=0.042$). Participants with lower levels of synovial thickness had greater odds of meeting the MCID in pain relief for capsaicin relative to ibuprofen (OR 3.08, 95%CI - 0.72 to 13.14). In turn, those with the highest levels of synovial thickness had lower odds of meeting the MCID on capsaicin relative to ibuprofen (OR 0.32, 95%CI 0.07 to 1.45). The interaction indicated that pain relief was more often experienced for ibuprofen, relative to capsaicin, in individuals with higher synovial thickness (indicative of peripheral inflammation).

5.3.5 Feasibility assessments

5.3.5.1 Outcome measures

Minimal agreement was found between treatment preferences at the cycle-level determined using NRS criteria (preferential treatment response) versus participant's own preference (50.98% agreement, $\kappa=0.2660$). For those cycles where a preference was reported for capsaicin, the median difference in NRS pain relief between treatments was 1 point in favour of capsaicin (range: 6 points for capsaicin to 1 for ibuprofen). Where ibuprofen was preferred, the median treatment difference was 0 (range: 10 points for capsaicin to 6 for ibuprofen). Finally, where no preference was reported, the median treatment difference was 0 (range: 3 points for capsaicin to 7 for ibuprofen).

Differences in treatment preference were evident when comparing overall treatment preferences using NRS criteria (preferential treatment response), the participant's own end-of-cycle preference (based on pain relief), and end-of-study preference (overall preference) (Table 5-7). The agreement between NRS-determined preferential treatment response and the participant's end-of-cycle treatment preference was greater ($\kappa = 0.3923$) than with the participant's end-of-study preference ($\kappa = 0.2373$). Only minimal agreement ($\kappa = 0.3732$) was observed between the participant's preference based for pain relief (end-of-cycle) versus their overall preference (end-of-study).

Of the 19 participants that recorded overall treatment preferences using at least two approaches, six showed the same treatment preference across all three methods. In three participants, the NRS preferential treatment response was in keeping with the end-of-cycle preference. In five, the end-of-cycle preference was in keeping with the end-of-study preference. In two people, the NRS preferential treatment response was equal to the end-of-study preference. Three people had a different treatment preference for each method, however, two of these participants recorded an “unclear” for one of the methods.

Table 5-7 - The number (%) of participants per treatment preference category using NRS preferential treatment response, participant’s end-of-cycle pain preference criteria, or participant’s end-of-study overall preference

Treatment preference	NRS criteria	Participant’s preference	
		End-of-cycle	End-of-study
Ibuprofen	4 (18)	7 (32)	10 (46)
Capsaicin	9 (41)	6 (27)	4 (18)
No preference	4 (18)	5 (23)	7 (32)
Unclear (dropout)	5 (23)	4 (18)	1 (5)

5.3.5.2 Weekly NRS pain scores

Fluctuations in weekly NRS pain scores, collected using participant diaries, were apparent (Appendix 18 – Weekly NRS pain scores). Participants reported greater levels of pain relief at longer durations of follow up (β 0.29, 95%CI 0.19 to 0.40, $p < 0.001$).

Compared to the primary outcome, using weekly NRS scores yielded missing data for an additional 16 (15%) periods. Treatment preferences were only established for 10 participants using data from week one, 10 participants from week two, 11 participants from week three, and 12 participants from week four when using diary scores. Agreement between preferences based on diary-recorded NRS scores versus investigator-enquired NRS scores at four weeks was minimal (54.6% agreement, $\kappa = 0.3642$).

Table 5-8 - The number (%) of participants per treatment preference category using primary NRS criteria and using weekly participant diary data at each duration of follow-up

Preference	Primary	Week 1	Week 2	Week 3	Week 4
Ibuprofen	4 (18)	2 (9)	4 (18)	3 (14)	2 (9)
Capsaicin	9 (41)	4 (18)	4 (18)	7 (32)	8 (36)
No preference	4 (18)	4 (18)	2 (9)	1 (5)	2 (9)
Unclear	5 (23)	12 (55)	12 (55)	11 (55)	10 (45)

Moderate agreement was found between diary-established preferences at week four and earlier weeks (week 1 $\kappa= 0.7161$; week 2 $\kappa= 0.6562$; week 3 $\kappa= 0.7250$).

5.3.5.3 Number of cycles

On limiting the trial to only two treatment cycles, a treatment preference was established in 10 of the 22 participants. The addition of the third treatment cycle yielded a treatment preference for an additional seven (32%) participants. Using treatment preferences established at the end of two cycles, no associations were found between baseline predictors and preference. The difference between treatments remained non-significant ($p=0.354$).

5.4 Discussion

5.4.1 Key findings and interpretation

This pilot n-of-1 trial series provides the first evidence for individual treatment preferences and potential predictors of response to topical NSAIDs and capsaicin in OA. Treatment preferences were determined for each participant individually, and a variation in treatment preferences was observed across the trial population. Treatment equivalence demonstrated at a study-level concealed clinically important differences between treatments at an individual level. Exploratory analyses identified several potential baseline predictors of the overall treatment effect and preferential treatment response that may warrant further examination. However, as the trial series was not

powered for these analyses, the findings, which are discussed below, should be interpreted with caution and only as hypothesis-generating.

Clinically meaningful pain relief was experienced in approximately 60% of periods using either topical NSAID or capsaicin. Prognostic factors associated with an improved response to topical therapy in primary univariate analysis were increased baseline pain, female sex, and increased BMI. However, the effect of sex and BMI were not independent of baseline pain in multivariable analysis. Only the association between baseline pain and response to topical therapy remained significant. In keeping with the overall effect analysis of the IPD MA (Chapter 4) and previous evidence for 0.075% civamide cream in OA (Schnitzer et al., 2012), those with more baseline pain experienced greater reductions in pain. However, a statistically significant association between baseline pain severity and the likelihood of achieving clinically important pain reduction was not found. This justifies the view that the observed association between baseline pain and treatment response may reflect regression to the mean (Glymour et al., 2005), rather than a biologically determined association.

Sensitivity analyses found that those with better physical function more often reported clinically meaningful reductions in pain following topical therapy. This is in keeping with the work by Bingham III et al. (2011) that showed that individuals with lower physical function scores at baseline had significantly lower odds for achieving OMERACT-OARSI response after taking oral NSAID or placebo. Self-reported physical function, especially as measured on the WOMAC or KOOS scale, represents a complex interaction between pain, pathology, and psychological factors. The association between physical function and overall treatment response may therefore reflect the conclusion that those with less severe OA show a greater response to therapy (Bruyère et al., 2015).

Although topical NSAIDs and capsaicin were equally effective for pain relief across the whole study population, variations in response were evident at the individual level. Approximately 60% of participants

showed an increased response for one treatment over the other. Baseline characteristics that associated with a preferential treatment response to topical NSAID, capsaicin, or neither were synovial thickness, fibromyalgia severity, and quadriceps strength. Of these, only synovial thickness also significantly interacted with the treatment term in multilevel sensitivity analyses, suggesting that people with greater synovial thickness may respond better to topical ibuprofen than to capsaicin. Additional significant interactions were also identified for distal PPTs and TS, however, these findings were not replicated across alternative analyses.

Topical NSAIDs show preferential distribution to areas of inflammation where therapeutic concentrations of the drug persist (Hagen and Baker, 2017). Those with greater synovial thickness may therefore benefit more from topical NSAIDs as a greater concentration of drug migrates to the target tissues where it is able to exert its anti-inflammatory effect. Alternatively, it may be that participants with more peripheral inflammation find the application of an alcohol-containing gel “cooling” or “soothing” as compared to a cream. The association may therefore reflect preferences in the formulation, rather than drug. No association was observed between preference for topical NSAIDs and effusion or power Doppler signal, alternative assessments of US inflammation. This may be because of the three US features of synovial inflammation, synovial thickness has been shown to have the strongest relationship with OA knee pain (Sarmanova et al., 2017). Alternatively, it may be because of the relative homogeneity in the effusion and Power Doppler measures across the study population. The majority of participants had effusions (>4mm) and were power Doppler negative, therefore making it difficult to identify predictors.

Increased fibromyalgia severity scores are indicative of more widespread pain and increased symptom severity (e.g., fatigue, waking unrefreshed, cognitive symptoms, headaches, lower abdominal cramps, depression), which are often taken to indicate increased contributions from central sensitisation. In addition, individuals with fibromyalgia have been shown to have significantly reduced muscle strength than matched controls (Larsson et al., 2018). This may

explain why baseline characteristics of reduced fibromyalgia severity and increased quadriceps strength favoured the same treatment.

Reduced distal PPTs and increased TS may be indicative of increases in central sensitisation or aberrant central processing of OA pain. In the present work, individuals with increased QST-assessed pain sensitivity had a better response to ibuprofen than capsaicin. This finding was not reflected across primary analyses, other markers of central sensitisation, or the Central Mechanism Traits score. In addition, the findings are further conflicted by the view that increases in fibromyalgia severity indicate increasing central sensitisation. Examination of the literature reveals that the evidence for associations between markers of central sensitisation and response to NSAIDs or capsaicin is equally inconsistent.

On one hand, Edwards et al. (2016) found that those with less impairment in CPM (i.e., less central sensitisation) reported increased response to topical diclofenac gel use. However, this response was not found for PPTs or TS, although the protocol for their assessment differed from the present study. In addition, the analysis was uncontrolled (i.e., less impairment in CPM is a prognostic factor for an improved response on topical NSAID) and included only 35 participants. Work by Arendt-Nielsen et al. (2016) failed to identify any QST-features that significantly predicted a response to oral NSAID. Finally, preliminary analyses by Petersen et al. (2017) indicate that responders to oral NSAIDs had increased levels of TS at baseline (i.e., more central sensitisation) compared to non-responders. A similar, albeit non-significant, association was found between lowered PPTs and increased response to oral NSAID. Caution in interpreting the work is required as it is preliminary and remains unpublished as a full text. In addition, the analyses were conducted in only 80 participants and were uncontrolled (i.e., increased TS is a prognostic factor for an improved response to oral NSAID). The evidence for central sensitisation and response to NSAIDs is very limited and conflicting. Oral NSAIDs are thought, in part, to be effective through inhibition of central mechanisms associated with central sensitisation (Arendt-Nielsen et al., 2016). However, as the systemic absorption of topical

NSAIDs is negligible, their central action is limited. The biological explanation for the observed interaction is therefore unclear. Perhaps topical NSAIDs act on individuals with central sensitisation by reducing prostaglandin levels in and around the joint, thereby normalising peripheral nociceptor hypersensitivity.

No evidence for central predictors of response to capsaicin in OA exist. Some evidence is available for the 8% capsaicin patch in NP which may be extrapolated to the low concentration cream in OA with extreme caution. As for topical NSAIDs, the evidence for central predictors of response to capsaicin is sparse and conflicting. Mainka et al. (2016) found that responders to capsaicin were more likely to have cold and mechanical hyperalgesia (i.e., more pain sensitivity). Katz et al. (2015a) found that increased allodynia (i.e., more pain sensitivity) associated with a more favourable response to capsaicin. In contrast, Gustorff et al. (2013) found that decreased allodynia (i.e., less pain sensitivity) associated with better response to capsaicin. In OA, it may be that the augmented sensitivity to a wide variety of modalities means that individuals with central sensitisation experience a more intense burning sensation on applying capsaicin cream. This may increase the placebo effect of capsaicin, thus reducing the corresponding overall treatment effects.

The present work was not powered for the exploration of potential predictors of response to the two most commonly used topical analgesics (NSAIDs and capsaicin). Accordingly, the findings should be interpreted with caution. A definitive study is warranted.

5.4.2 Recommendations for a definitive study

The present study has shown that it is feasible to conduct n-of-1 trial series to examine topical NSAIDs and capsaicin in individuals with knee OA. Recommendations for the design and conduct of the definitive study are outlined below.

Of the pool of potentially eligible participants in the present study, only 35% expressed an interest in participating in the trial. The recruitment strategy for a definitive study should therefore utilise a

sufficiently large pool of potential participants as the source population in order to meet recruitment targets. If a definitive trial were to be conducted within the UoN Academic Rheumatology department, the source population considered could be the KPIC participants that have not attended the department for study visits (i.e., that were not considered for the present study) or from other cohorts of community-dwelling individuals.

A potential barrier to the recruitment of participants may be the long duration of the trial (maximum 44 weeks). It may be possible to shorten the trial by reducing the length of treatment periods to three weeks. Of the pain data collected at each week of follow-up, data at three weeks showed the greatest agreement with the four-week data. Reducing treatment period length to three weeks would reduce the overall trial duration (maximum 38 weeks), and may facilitate recruitment to the definitive trial. An alternative approach to reducing treatment duration, by reducing the number of treatment cycles, is not recommended as this was shown to reduce the number of participants for which a preferential treatment response was identified.

It would not be recommended to recruit participants according to the distinct phenotypes of predominantly neuropathic and predominately inflammatory. It may not be possible to meet recruitment targets due to the relatively low proportion of participants identified that meet criteria for predominantly neuropathic-like pain (<10% of the study population). The present work has indicated that significant interactions may still be captured for inflammation without limiting inclusion criteria to distinct subgroups (high vs low SH). Recruitment should aim to include a spectrum of levels of synovial thickening or central sensitisation without dichotomising participants at inclusion, as long as a full spectrum of severity of each measure are covered. Recruitment from a community dwelling source population, where a large variation in the characteristic of interest would be expected, should be considered.

It is recommended that a definitive n-of-1 trial series implemented for the examination of predictors of response in OA focus on only one

potential predictor per trial. The present work indicated that US-measures of inflammation, QST-assessed measures of central sensitisation, and the presence of fibromyalgia may warrant further examination in a definitive trial. In order to maximise the potential of capturing predictors of response, such a trial may consider restricting analyses only to participants with baseline pain of 4-8 on a 0-10 scale per period.

The assessment of treatment responses using the 0-10 NRS for pain was easy to implement across telephone, text, and home visits. Response rates at the beginning and end of each treatment period were excellent, likely reflecting its ease of implementation. However, NRS preferential treatment responses showed only minimal agreement with the participant's own treatment preference. It may be that the NRS was not adequately capturing the properties of the pain experience that influenced a participant's treatment preference based on pain relief. Alternatively, participant preferences may have been affected by recall bias or have been based on more than just pain relief. A definitive trial may consider implementing a multidimensional pain tool. Alternative means of collecting outcome data, such as using an app or at home visits may be required.

Participant retention rates were in keeping with the literature for n-of-1 trial series, with 23% of participants withdrawing prior to trial completion. The planning of a definitive trial should consider these rates when conducting sample size calculations and implement approaches for handling missing data. The sample size required for a definitive trial should be specific to the primary aim and analysis of the study. Simulation methods provided by Lane and Hennes (2018) could be implemented once the statistical model and aim of the study is determined (e.g., which potential predictor is examined). The simulation would also require the input of the means, variances, and coefficients (i.e., covariances and correlations) for the effects at the cycle and participant-level of analysis generated from the present feasibility study (Lane and Hennes, 2018). Sample sizes required to determine interaction effects are larger than those required to assess treatment effects. In the example by Lane and Hennes (2018), a ten-

fold increase in the number of participants required to detect a significant interaction effect was required compared to the number needed to detect the main treatment effect. It may be that the sample size required goes beyond the resources available for the definitive trial, therefore emphasising the role of IPD MAs for detecting treatment interactions.

5.4.3 Caveats

The present work is subject to several limitations. First, the examination of potential predictors of response involved analysis of many covariates within a limited sample size. Although the list of potential covariates examined was pre-specified, the findings are subject to the problem of multiple comparisons. When examining 25 baseline characteristics in seven analyses at the 0.05 significance level, we would expect to find nine significant findings purely by chance (Kirkwood and Sterne, 2003). The present work identified 12 significant findings (nine separate characteristics), indicating that a large proportion of these findings may have been identified by chance. In addition, analyses conducted were not all pre-specified in the protocol. For example, the examination of predictors of the overall treatment effect (regardless of drug), analyses using the MCID, and sensitivity analyses using baseline pain scores of 4-8 NRS were not pre-specified. Furthermore, only a subset of the predictors (distal PPT, TS, SH) were identified through treatment-by-covariate interactions. However, even these interactions were only found in sensitivity, and not primary, analyses. Finally, imbalances in subgroup ratios (e.g., PD signal absent in 86% of population), may have biased the conduct of these subgroup-specific tests (Brookes et al., 2001). Significant findings observed may therefore be spurious and resulted from chance. The results should be interpreted with extreme caution and viewed as exploratory in nature. However, their conduct allowed exploration and further development of methods for the implementation of n-of-1 trial series in OA.

Second, the trial was not powered for the exploratory examination of predictors of response. Potentially important predictors may have been missed due to the relatively small number of participants analysed. The

absence of a significant effect therefore does not mean that a significant effect does not exist, but rather that it could not be detected in the present study.

Third, imbalances in the treatment preference of drop-outs limited examination of predictors of the treatment effect. 80% of the drop-outs favoured ibuprofen in their first treatment cycle, however, they were not included in the examination of overall treatment preferences as their preferential treatment response was not established. This imbalance may have biased the findings for the predictors of preference. These participants were, however, included in the multilevel modelling for treatment-by-covariate interactions.

Fourth, for a subset of participants, their baseline pain levels after washout remained at low or non-existent levels. This can be seen in rows 8, 10, 16, and 22 of Figure 5-3. Where baseline pain was 0 or 1 points, there was limited scope to record improvements in response to therapy. This may have lowered the apparent treatment effect and biased treatment comparisons if baseline pain levels were dissimilar within a treatment cycle. This limitation was accounted for in the present work by adjusting for baseline pain when comparing the treatments and conducting subgroup analyses of periods where baseline pain was 4-8 on NRS. The pattern observed may be due to insufficient washout durations, the Hawthorne effect, regression to the mean, natural fluctuations in OA pain, or the effect of participant's lifestyle changes. Many factors that affect a person's knee pain may change during a prolonged trial and this is a major limitation for trials with multiple cross-overs. It may be that OA knee pain may not be stable enough for examination in lengthy n-of-1 trials.

Fifth, it was not possible to examine the *a priori* defined comparison of predominantly inflammatory versus predominantly neuropathic-like phenotypes. It was not possible to recruit sufficient number of participants with such restrictive criteria. Most commonly, participants exhibited a mixed phenotype. Furthermore, calculation of a neuro-inflammation ratio as specified in the trial protocol was not done. Standardising US scores according to a maximum possible score was

not felt appropriate as no consensus was reached for the appropriate method. In addition, the Rasch-transformation of PDQ scores yielded negative values that did not lend themselves to the calculation of the ratio.

Sixth, although the trial population was reflective of the overall prevalence of neuropathic-like pain in the Nottinghamshire community, the trial reported a significantly larger proportion of individuals with evidence of US inflammation (Fernandes et al., 2018, Sarmanova et al., 2017). This may limit the applicability of the findings to the general community.

Seventh, participants and trial personnel were not blinded. Due to inherent difficulties in blinding capsaicin and in order to more closely reflect clinical care, the trial was open label. The results may therefore be biased by treatment expectancy and other contextual factors. In order to minimise this, a standardised explanation for the mechanism of action was given to participants. In addition, physical examination findings were not related to hypothesised treatment mechanisms. Nonetheless, separation of the specific treatment effects was not possible. Ideally future n-of-1 trial series should be blinded. If not feasible, treatment expectancy should be accounted for by asking participants about their treatment-specific expectancies at baseline. Treatment expectancy should relate to the drug and outcome measure used. The examination of treatment effects could then be adjusted for treatment expectancy.

Eighth, NRS was used as the primary outcome measure, due to its ease in applicability and interpretation. However, limited agreement between preferences determined using NRS and those captured by asking the participant themselves was observed. This may suggest that the absolute difference in pain intensity scores may not fully capture the complexity of participant preferences. Due to the unidimensional nature of NRS, it may not effectively summarise the participant's experience of a treatment. Informal discussions with trial participants highlighted the difficulty they experienced in summarising their whole pain experience with one number. The use of NRS may

further have been limited by difficulties with recall as participants were asked to rate their average scores over one week. Once again, participants voiced concerns over choosing a number that would accurately reflect a whole week. Finally, the tool may not be sensitive enough to detect small differences in treatment preference. It may be that clinically important differences between treatments were smaller than 1 point on the 0-10 NRS.

Finally, a MCID of 1 point was used to determine the treatment preference at a cycle-level and to determine the number of periods where pain relief was clinically important. However, the choice of MCID and its use is subject to limitations. The MCID should be specific to the instrument and should be based on calculations conducted in a very similar context (Katz et al., 2015b). The cut-off used by NICE was determined empirically, with the purpose of grading imprecision across the evidence (NICE, 2014). This value may therefore not be applicable for use at the individual level. In addition, using a pooled SD of 2.5 cm, the MCID used by NICE is comparable to 1.25 cm change on a 0-10cm VAS (NICE, 2014, Wandel et al., 2010). Direct conversion from VAS to NRS may not be appropriate, and the conservative use of 1 point on the NRS may therefore be questioned. In addition, the same threshold was used to compare pain relief between treatments (to determine a preference) and to assess absolute change from baseline scores for each period. MCID is primarily used to compare the proportion of individuals that meet the MCID between treatments, and it may not be appropriate to apply the same threshold to determine a treatment preference. Although agreement between methods was limited, the findings indicated that a treatment preference was established based on small differences in change scores. Median treatment differences were 0 and 1 for a preference to ibuprofen and capsaicin, respectively, supporting the use of 1 point as the MCID between treatments. However, variation in treatment difference scores were observed for individuals with the same treatment preference, particularly for ibuprofen, suggesting that further methods for determining treatment preference need to be explored.

5.4.4 Methodological comments

N-of-1 trials are useful for determining patient preference to treatment at an individual level. This has traditionally been their primary use and the setting in which they are most able to guide evidence based practice. However, they are increasingly being advocated for use in precision medicine (Lillie et al., 2011). Consideration of the present pilot n-of-1 series suggests they may be inefficient for the identification of possible predictors of response (hypothesis-generating) unless universally implemented as part of standard care. Each additional cycle contributes to the certainty of the treatment preference, but increasing numbers of participants are required in a series to identify possible predictors. The trial series are therefore resource-intense where the main goal is to identify possible predictors of response.

Efficiency could be increased by measuring non-constant potential predictors at the beginning of each treatment cycle. This may increase the power of the multilevel interaction models, but is dependent on the assumption that the treatment does not alter the baseline characteristic. This assumption may not be valid for characteristics, such as the quality of pain or severity of inflammation, which may be affected by treatment.

Instead, n-of-1 trial series may be more useful in confirmatory analyses of predictors previously identified in large-scale trials or IPD MAs. In this instance, the addition of multiple treatment cross-overs (as compared to an RCT) may be justified by establishing the preference with more certainty. In these cases, statistical power should be increased through the addition of more participants, rather than more cycles. As greater variation in responses was observed *between*, rather than *within*, participants the inclusion of additional treatment cycles would add little.

The optimal number of treatment cycles will likely vary between participants. Where a strong response is reported by a participant, the number of cycles required is low. Each additional cycle adds little to the precision. However, where the response is small or heterogeneous,

more cycles are likely needed. Individualising the trial design to the participant allows investigators to balance precision with efficiency. Where a clear response is established early on, it may not be appropriate to expose participants to periods using treatments they have already shown that they do not benefit from. The costs and duration of additional cycles may be spared, which may be justified for only minor losses in precision.

The duration of each treatment period needs particular consideration. Although the interpretation is limited by considerable missing data, diary data suggested that it may be appropriate to shorten treatment periods to three weeks. Moderate agreement was observed between diary scores at three and four weeks and preferential treatment responses were similar.

The frequency and nature of outcome measures should be considered for future n-of-1 trial series. Increasing the frequency of data collection would theoretically generate more data for analyses, but must be balanced against the increased workload expected from participants. The current work found that increasing the frequency of outcome collection was associated with a greater proportion of missing data. Using weekly diary scores may therefore not be appropriate. Alternative means of collecting outcome data at a greater frequency may be considered, such as the use of wearable technology or apps for data collection.

Finally, the nature of the outcome assessed should be reconsidered. Anecdotal experience suggests that participants did not feel that the NRS accurately captured their pain experience. A multidimensional pain tool, such as the Measure of Intermittent and Constant Osteoarthritis Pain questionnaire may be considered in its place. This should be supplemented with direct enquiry to participants regarding whether or not a treatment provided meaningful pain relief. The application of MCID to NRS pain scores can be likened to a surrogate marker for efficacy. It may be more appropriate to directly enquire if the treatment provided clinically meaningful pain relief than to apply MCID thresholds to NRS.

6 CHAPTER 6: Cross disease comparison – efficacy of topical NSAIDs and capsaicin in NP

6.1 Overview

As has been previously established, a subset of individuals with OA report neuropathic-like pain. However, little is known about the efficacy of topical NSAIDs and capsaicin in those suffering with neuropathic-like pain. In particular, although capsaicin has long been used for NP, it is still unknown whether this translates to an improved response for those with neuropathic-like pain, compared to those without, in OA.

Unfortunately, RCTs of topical NSAIDs and capsaicin have not assessed the presence or impact of neuropathic-like pain on treatment efficacy. It has therefore not been possible to determine if treatment efficacy varies by the presence of neuropathic-like pain in OA using IPD. A cross-disease comparison was therefore planned using NP. The following chapter describes the MAs conducted for topical NSAIDs and capsaicin in NP.

6.2 Methods

RCTs of topical NSAIDs or capsaicin in NP (e.g., PHN, PDN, and HIV) were identified using systematic literature searches, hand searches, and direct enquiries to pharmaceutical companies. All identified studies, regardless of the comparator and publication status, were deemed eligible for inclusion.

6.2.1 Identification of eligible trials

6.2.1.1 Systematic literature search

The methods described for the systematic literature searches in OA (Section 2.1.1.1) were largely unchanged except for the substitution of NP as the disease. The methods are briefly summarised below, including the amendments made to capture NP participants.

6.2.1.1.1 Search terms and strategies

Two search strategies were developed to identify RCTs assessing topical NSAIDs or capsaicin in NP. As for the systematic literature search in OA, the search strategies were created using four main search terms combined using “and”: (1) terms for RCTs; (2) terms for topical administration; (3) terms for NP; and (4) terms for the treatment (NSAID or capsaicin). The NICE guidelines for NP were referred to when aggregating search terms for NP (NICE, 2018). They were expanded upon through the addition of terms found through searching the internet. The search strategies used can be found in Appendix 1 – Systematic literature search strategies.

6.2.1.1.2 Databases and literature search output

Medline, Embase, AMED, CINAHL, Web of Science, and Cochrane Library were searched from date of inception to 30/03/2016.

Following the removal of duplicates, the results of the literature search were screened using the titles. Abstracts of those potentially eligible were sought and assessed. Finally, full texts were scrutinised and reasons for exclusion at this stage were noted.

6.2.1.2 Identification of other studies

The systematic literature search was supplemented by searching reference lists of included publications, reviews in the area, and hard copies of potentially relevant trials kept in the department. Reviews were identified through literature searches in Medline and Cochrane Library using MeSH headers for “neuralgia” and “administration, topical”. The results were then limited to “meta-analysis or systematic review”.

As for the OA trials, pharmaceutical suppliers of topical NSAIDs and capsaicin were contacted to ask for any unpublished trial data.

6.2.1.3 Study eligibility criteria

Trial eligibility was determined based on five domains: the types of (1) studies; (2) participants; (3) interventions used; (4) comparators used;

and (5) outcome assessments. All inclusion and exclusion criteria were identical to the search in OA, aside from the types of participants (Table 6-1).

Table 6-1 - Inclusion and exclusion criteria used for MAs of topical NSAIDs and capsaicin in NP

Domain	Inclusion criteria	Exclusion criteria
Study	RCT Minimum 1 week duration No language or geographic restrictions Open-label and blinded studies	Trials with the same drug and concentration in both arms
Neuropathic participants	Painful long-term NP, e.g. post-herpetic, diabetic, HIV-associated, idiopathic	Post-surgical NP
Intervention	Any topical NSAID or capsaicin Any formulation, dosage, or drug within group	Salicylates Interventions using US or magnetophoresis for application
Comparator	Controls: placebo, active comparators, continuing usual treatment, no treatment	
Outcome	Adequate assessment of pain	

HIV, human immunodeficiency virus; NP, neuropathic pain; NSAID, non-steroidal anti-inflammatory drug; US, ultrasound

Men and women with painful long-term neuropathic conditions, such as PHN, PDN, HIV-associated neuropathy, and idiopathic neuropathy, were included. Post-surgical NP was excluded after running the search strategy as participants did not exhibit pain that was purely neuropathic in nature (Bischoff et al., 2014).

Studies investigating only eligible participants and studies with only a subgroup of eligible participants were included, as long as IPD could be collected separately for the eligible participants.

6.2.2 Data collection

Data collection, validation, and management for study-level data extracted from published reports were conducted as described in section 2.1.3.

Author contact and collection of IPD were conducted as described in section 4.2.1. The same procedures and same level of rigour was applied in attempting to retrieve IPD for NP trials as was applied for the OA trials. In brief, the first or corresponding author of all eligible trials was contacted using a standardised email, personalised to include the author name and study details. Where no response was received, additional attempts to contact data custodians were made by sending two reminder emails, contacting via letter and telephone, contacting all other publication authors, contacting the institution where the research was conducted, and reaching out to the trial funder or sponsor. Unless contact details were unavailable, all additional approaches were implemented for the studies until a definitive response (accepting or declining collaboration) was received or data collection was closed.

6.2.3 Data analysis

Data analysis was conducted as for the OA trials, although this was dependent on data availability.

As for OA trials, the specific and overall treatment effect were examined in CMA. Primary analysis was for pain relief at or nearest four weeks. Cumulative MA was conducted to examine how the effect has changed over time. Funnel plots and formal tests for publication bias were conducted, where appropriate, as per section 3.2.1.1 (bullet 3). Sensitivity analyses were conducted for trials with adequate allocation concealment and large trials ($n > 100$ per arm). Finally, analyses were conducted by duration of follow-up.

NMA was not conducted as no common comparator was identified between topical NSAIDs and capsaicin.

IPD were sought for NP participants, but it was not possible to conduct an IPD MA.

6.3 Results

6.3.1 Study-level MAs

The following section describes the results of the literature search for topical NSAIDs and capsaicin in NP. It describes the specific treatment effect followed by the overall treatment effect and PCE (Zou et al., 2016). The findings at or nearest to four weeks (primary endpoint) are presented first. These are followed by sensitivity and subgroup analyses at four weeks and by analyses at other treatment time-points.

A CMA could not be conducted for topical NSAIDs and the focus of the chapter is the description of the studies and the results for capsaicin in NP. Due to the lack of data on topical NSAIDs, it was also not possible to conduct a NMA to compare topical NSAIDs to capsaicin in NP.

6.3.1.1 Description of studies

6.3.1.1.1 Topical NSAIDs

6.3.1.1.1.1 Literature search

The results of the systematic literature search are illustrated in Figure 6-1.

The systematic literature search identified one study that compared diclofenac lotion to placebo, however, it included non-eligible participants with post-surgical NP. Data could not be extracted separately for only the eligible participants and the study was not included in the analyses. A MA was not conducted for topical NSAIDs in NP.

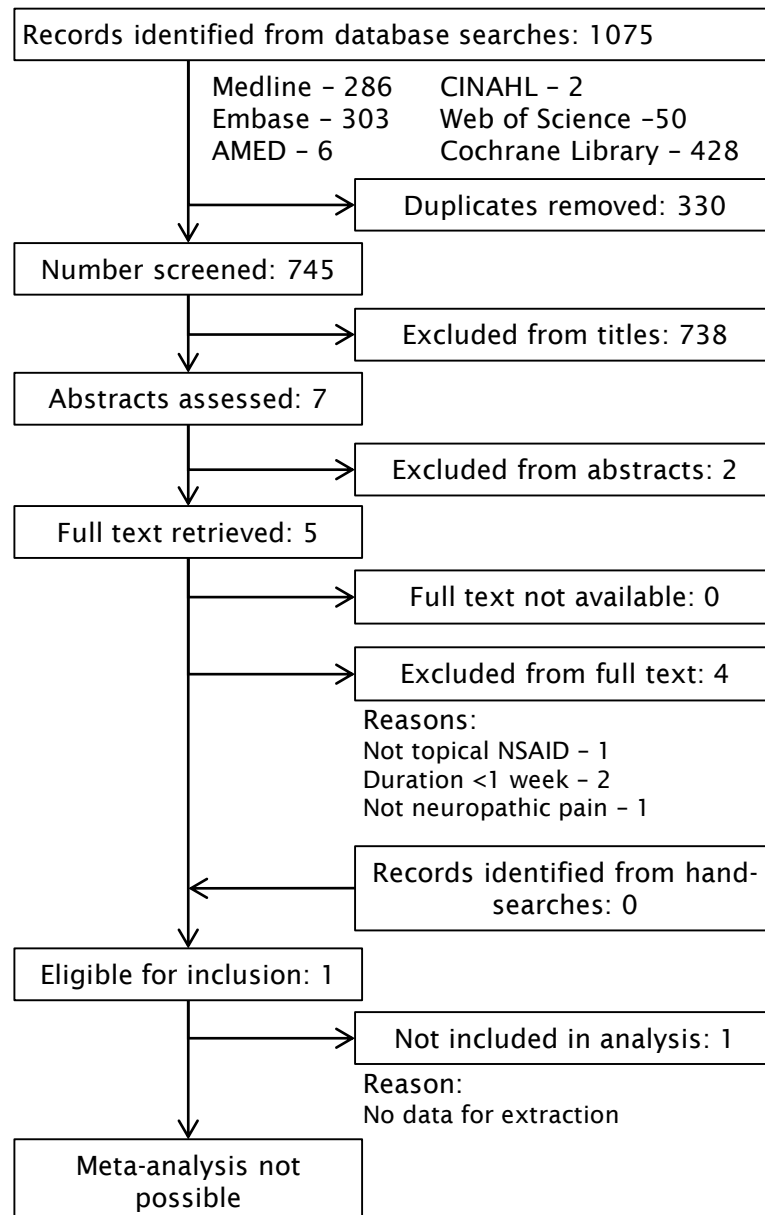


Figure 6-1 - Flow Diagram: Literature search results for topical NSAIDs in NP

6.3.1.1.2 Capsaicin

6.3.1.1.2.1 Literature search

The results of the systematic literature search are described in Figure 6-2.

The literature search identified 20 RCTs of capsaicin in NP. Ten studies did not compare capsaicin to placebo. Of these, eight compared the 8% to the 0.04% (control) capsaicin patch. The 0.04% capsaicin was used for blinding purposes, however, it was not regarded as a placebo for the current MA as its capsaicin concentration was greater than in many of the “active” capsaicin creams. Finally, three studies did not provide sufficient data for analysis (data presented graphically only, missing variance, or missing number analysed) and the remaining seven studies were included in the analysis of capsaicin versus placebo.

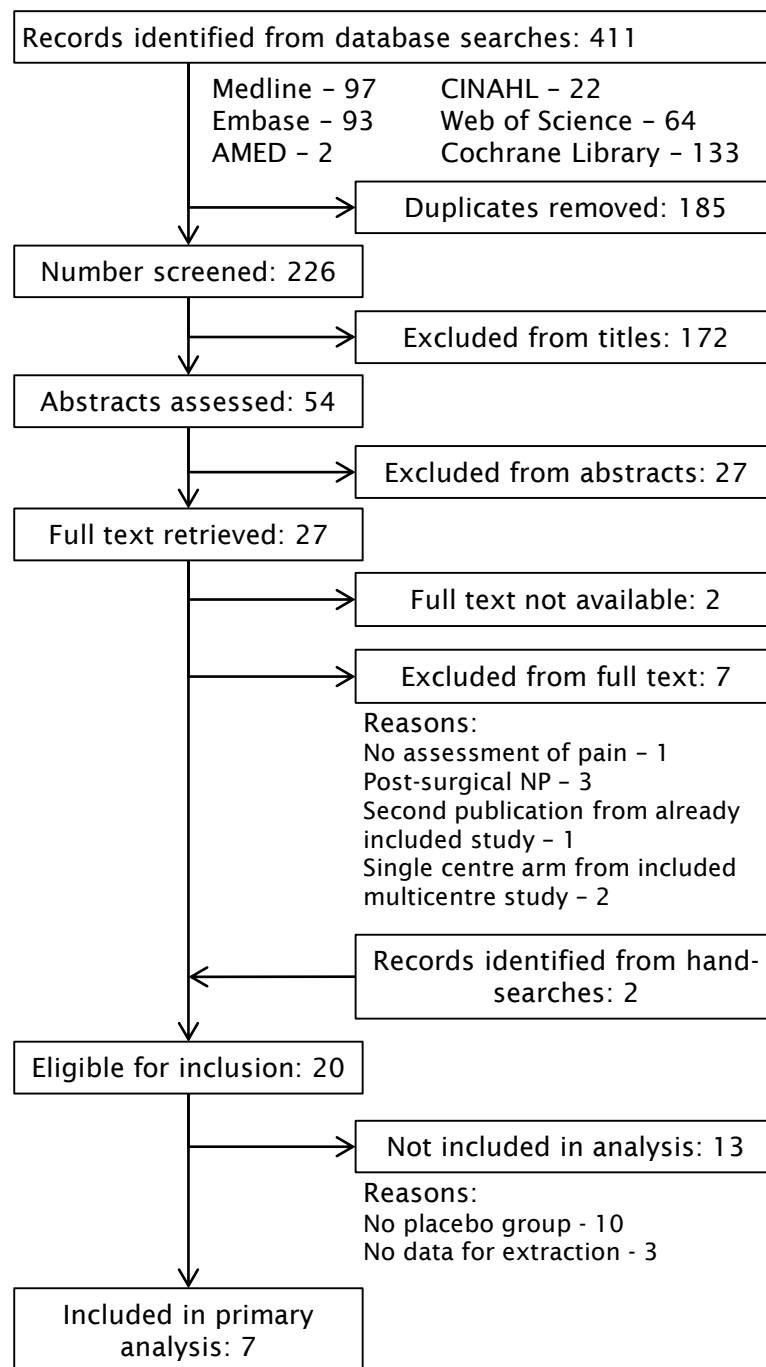


Figure 6-2 - Flow Diagram: Literature search results for capsaicin in NP

6.3.1.1.2.2 Description of the studies

All trials were published in English and are described in Table 6-2.

Although the seven placebo controlled trials randomised 480 participants at baseline, the primary analysis was based on 243 participants on capsaicin and 259 on placebo. This is because: (1) two studies were cross-over trials and presented data for the treatment periods combined (Kulkantrakorn et al., 2012, Teixeira et al., 2015), and (2) another study randomised by leg, whereby participants were randomly allocated to apply capsaicin to one leg and placebo to the other (Low et al., 1995). Only one study used ITT analysis (Kulkantrakorn et al., 2012).

The median number of participants analysed per study was 46 and this ranged from 21 to 234 participants per trial. Participants had NP due to PDN (three studies), PHN (two studies), HIV (one study), and other causes including PDN and inherited neuropathy (one study).

Capsaicin was compared to placebo in all seven studies, and details of the interventions are given in Table 6-3.

All but two studies (Kulkantrakorn et al., 2012, Teixeira et al., 2015) investigated 0.075% capsaicin cream applied four times per day. Plain vehicle was used as the placebo comparator in six studies. One study (Low et al., 1995) used methyl nicotinate in the first tube of placebo in order to mimic the erythema and discomfort associated with the application of capsaicin.

Table 6-2 - Description of the studies included in the analysis of capsaicin in NP

Trial Name	Year	Country	Funding Source	Design	Setting	Aetiology
Bernstein 1989	1989	USA	Mixed	Parallel	Unclear	PHN
Chad 1990	1990	USA	Unclear	Parallel	Community	PDN
CSG 1991	1991	USA, Canada	Commercial	Parallel	Community	PDN
Low 1995	1995	USA	Commercial	Parallel	Unclear	Idiopathic, Other, PDN
Paice 2000	2000	USA	Public	Parallel*	Community, Hospital	HIV
Kulkantrakorn 2012	2012	Thailand	Mixed	Crossover	Community	PDN
Teixeira 2015	2015	Brazil	Unclear	Crossover	Community	PHN

*Randomised participants by leg, whereby one leg was randomly allocated placebo and the other capsaicin

Year is year of publication

CSG, Capsaicin Study Group; HIV, human immunodeficiency virus; PDN, peripheral diabetic neuropathy; PHN, post-herpetic neuralgia

Table 6-3 - Description of the interventions and comparators in the studies included in the analysis of capsaicin in NP

Trial Name	Intervention	Concentration	Formulation	Frequency per day	Comparator(s)
Bernstein 1989	Capsaicin	0.075%	Cream	3-4	Placebo
Chad 1990	Capsaicin	0.075%	Cream	4	Placebo
CSG 1991	Capsaicin	0.075%	Cream	4	Placebo
Low 1995	Capsaicin	0.075%	Cream	4	Placebo + methyl nicotinate
Paice 2000	Capsaicin	0.075%	Cream	4	Placebo
Kulkantrakorn 2012	Capsaicin	0.025%	Gel	3	Placebo
Teixeira 2015	Liposomal capsaicin	0.025%	Cream	2-3	Placebo

6.3.1.1.2.3 Risk of bias assessment

The risk of bias of the analysed trials is presented in Figure 6-3. Overall, there was evidence of considerable risks of bias.

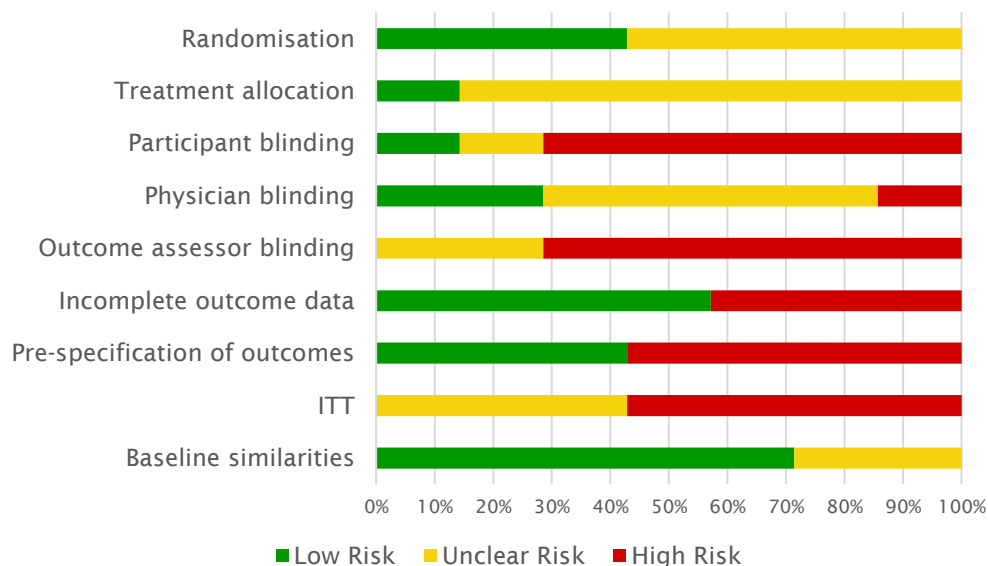


Figure 6-3 - Risk of bias assessment of the trials included in the analysis of capsaicin in NP

The domains associated with the highest risk of bias were blinding, pre-specification of outcomes, and ITT use. Despite all studies claiming to be double blinded, only one study used an agent in the placebo cream to mimic capsaicin and ensure adequate blinding (Low et al., 1995). Four studies excluded participants from the analysis post-randomisation (CSG, 1991, Chad et al., 1990, Bernstein et al., 1989, Paice et al., 2000). The remaining three studies did not provide sufficient information to determine the population analysed (Low et al., 1995, Teixeira et al., 2015) or to determine the adequacy of the modified definition of ITT (Kulkantrakorn et al., 2012). Only three studies fully reported all pre-specified outcomes (Kulkantrakorn et al., 2012, CSG, 1991, Chad et al., 1990). One study did not describe an outcome listed in the methods section (Low et al., 1995), whilst the three remaining studies incompletely reported the outcomes.

Three studies used a computer-generated randomisation procedure (Kulkantrakorn et al., 2012, CSG, 1991, Bernstein et al., 1989), whilst the remaining studies did not provide sufficient detail regarding the

randomisation sequence generation for the risk of bias to be determined. Similarly, only one study mentioned precautions taken to conceal allocation (Kulkantrakorn et al., 2012).

6.3.1.2 Specific treatment effect

The following section describes the specific treatment effects of topical NSAIDs and capsaicin in NP, i.e., the difference between treatment and placebo groups for pain relief.

6.3.1.2.1 Topical NSAIDs

No studies were available for the comparison of topical NSAID versus placebo and the specific treatment effect could not be examined.

6.3.1.2.2 Capsaicin

6.3.1.2.2.1 At or nearest to four weeks

Seven trials (502 treatment comparisons within 480 participants) were analysed for the primary outcome of capsaicin versus placebo at or nearest to four weeks (Figure 6-4).

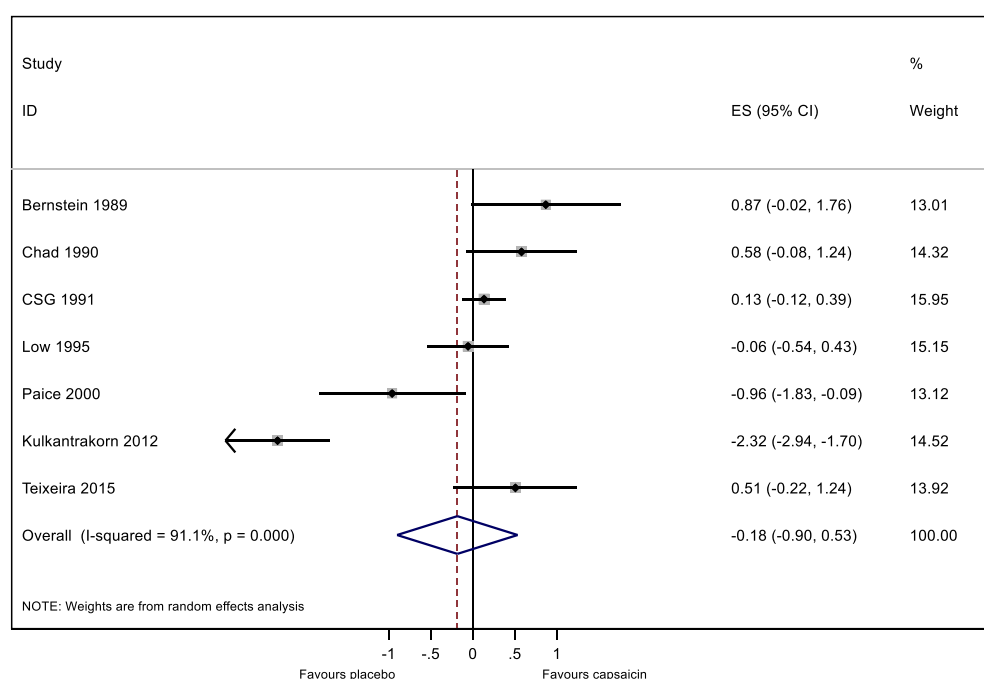


Figure 6-4 - Capsaicin versus placebo in NP. Outcome: Pain at or nearest to 4 weeks

Mean change from baseline pain scores were calculated for two studies (Teixeira et al., 2015, Kulkantrakorn et al., 2012). One study presented mean pain at time point (Paice et al., 2000), one presented the percent change in pain score from baseline (CSG, 1991), and three presented the number/percent of participants with improvement in pain (Bernstein et al., 1989, Chad et al., 1990, Low et al., 1995). In order to minimise the loss of information, all outcome types were combined for the primary analysis.

Very low quality evidence indicated that capsaicin was not superior to placebo for pain relief at or nearest to four weeks (ES -0.18, 95%CI - 0.90 to 0.53) (Appendix 19 – GRADE assessment: Capsaicin in NP). This is equivalent to an ES of -0.5 cm (95%CI -2.3 to 1.3) on 0-10 cm VAS.

Two studies found that participants on placebo had significantly less pain than those on capsaicin. In the study by Kulkantrakorn et al. (2012), participants on capsaicin experienced increased levels of pain at four weeks compared to baseline. Excluding the trial from the MA in a sensitivity analysis raised the ES to favour capsaicin, but the difference remained non-significant (ES 0.18, 95%CI -0.19 to 0.54). Additionally, in the study by Paice et al. (2000), placebo was superior to capsaicin at week one. However, the difference between capsaicin and placebo was non-significant at all other durations of follow-up in the trial. Only the data for week one was presented fully in the publication and the pain scores at all other time points were only presented graphically and could therefore not be extracted. A sensitivity analysis excluding this trial did not alter the conclusion and capsaicin remained no better than placebo for NP (ES -0.07, 95%CI - 0.85 to 0.72). Substantial and significant heterogeneity in the estimate was evident ($I^2=91.1\%$, $p<0.001$).

6.3.1.2.2.1.1 Cumulative meta-analysis

Studies on capsaicin in NP have been published from 1989 up to 2015. The results of the cumulative MA are presented in Figure 6-5.

The ES was significantly in favour of capsaicin only at the end of 1990. With the addition of new evidence, the ES for capsaicin has declined at a steady rate and shown no separation from placebo. Until the publication of the most recent study using liposomal capsaicin, the ES of capsaicin had decreased with the addition of each new study.

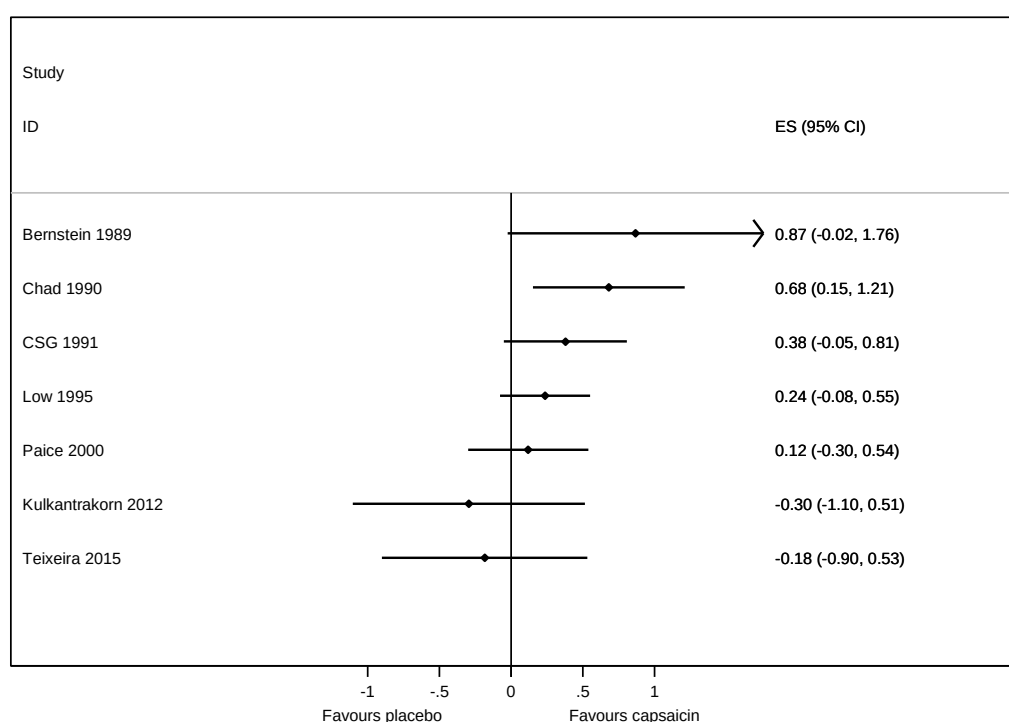


Figure 6-5 - Cumulative MA of capsaicin versus placebo in NP. Outcome: Pain at or nearest to four weeks

6.3.1.2.2.1.2 Publication bias and small study effects

The funnel plot for capsaicin versus placebo in NP is presented in Figure 6-6. There is some asymmetry on visual inspection of the funnel plot, however further tests were not conducted due to the small number of studies available (Higgins JPT, 2011).

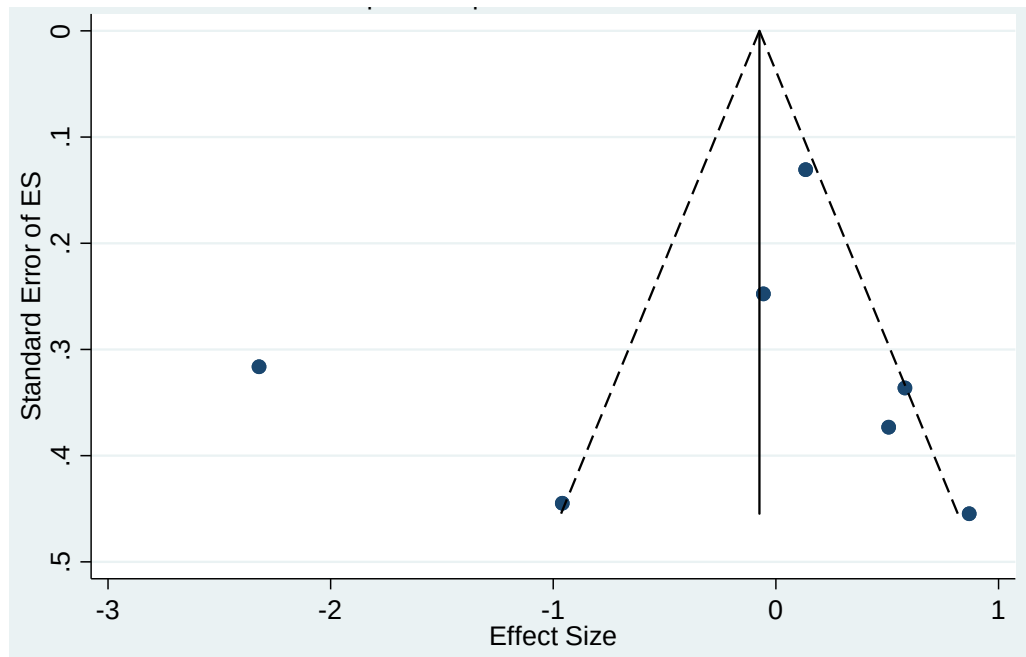


Figure 6-6 - Funnel plot with pseudo 95% confidence limits for capsaicin in NP

6.3.1.2.2.1.3 Sensitivity analyses

The results of the sensitivity analyses for high quality trials (using allocation concealment and n>100 per arm as indicators) are presented in Table 6-4.

Only one study (CSG, 1991) had more than 100 participants per treatment arm and it found that capsaicin was not superior to placebo. Similarly, only one study (Kulkantrakorn et al., 2012) had a low risk of bias for allocation concealment. In this study, participants on capsaicin had *increased* levels of pain at four weeks compared to baseline.

Table 6-4 - Sensitivity analyses by study quality for capsaicin in NP. Outcome: Pain at or nearest to four weeks

Studies included	Trials	n	ES (95%CI)	I ² (p-value)
All studies	7	502	-0.18 (-0.90 to 0.53)	91.1% (p<0.001)
Allocation concealment	1	66	-2.32 (-2.94 to -1.70)	-
n>100 in treatment arm	1	234	0.13 (-0.12 to 0.39)	-

Presented as Hedges' effect size (ES) and associated 95% confidence interval (95%CI). I² (the variation in ES attributable to heterogeneity) and associated p-value are presented n, number of participants

6.3.1.2.2.1.4 Subgroup analysis by duration of treatment

The effect of capsaicin was examined at each duration of follow-up available from the included studies (Figure 6-7). As per Table 3-1 durations over four weeks (six and eight weeks) were grouped together.

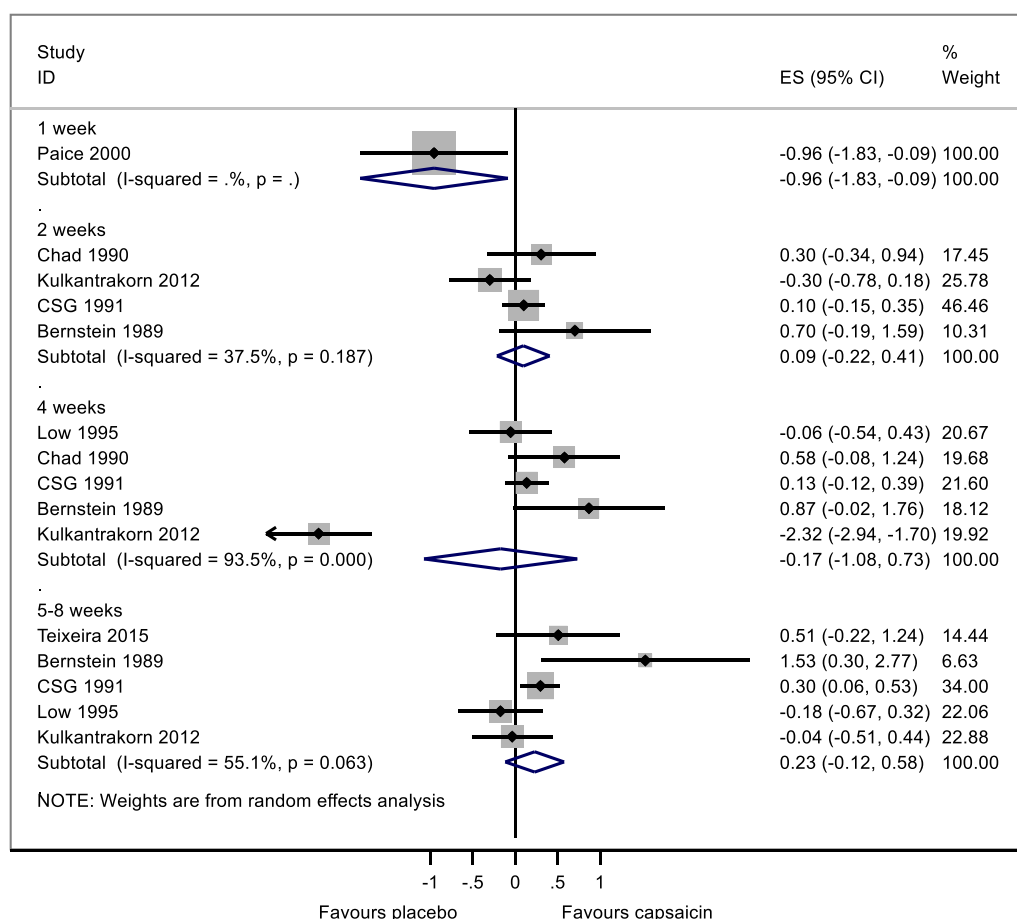


Figure 6-7 - Capsaicin versus placebo in NP. Outcome: Pain by weeks of follow-up

Capsaicin was significantly inferior to placebo in the first week of treatment. Following this, there were no differences between capsaicin and placebo for pain relief at any other duration of treatment. The largest ES in favour of capsaicin was evident at 5-8 weeks of treatment (ES 0.23, 95%CI -0.12 to 0.58), however, it remained non-significant.

6.3.1.3 Overall treatment effect and PCE

6.3.1.3.1 Topical NSAIDs

No studies were available for the comparison of topical NSAID versus placebo and a CMA of the overall treatment effect and PCE was not conducted.

6.3.1.3.2 Capsaicin

Only two studies presented mean baseline and endpoint pain data. In one (Kulkantrakorn et al., 2012), the participants in the capsaicin arm worsened from baseline at four weeks thereby preventing the calculation of the PCE (Zou et al., 2016). The overall treatment effect and PCE at four weeks was calculated for the remaining study (Teixeira et al., 2015). The overall ES (0.65, 95%CI 0.10 to 1.20) suggested a moderate level of pain reduction with only a minor contribution by contextual effects (PCE = 21%). No further analyses were conducted for the overall effect and PCE of capsaicin in NP.

6.3.1.4 Summary of findings

Comparison	SMD treatment effect	0-10 cm VAS treatment effect	n* (trials)	Confidence in effect estimate
Topical NSAID vs placebo	-	-	No data	-
Capsaicin vs placebo	On average, pain scores in the capsaicin group were 0.18 SDs higher (0.90 higher to 0.53 lower) than in the placebo group	On average, pain scores in the capsaicin group were 0.5 cm higher (2.3 higher to 1.3 lower) than in the placebo group	502 (7)	⊕○○○ Very low for risk of bias, inconsistency, imprecision, publication bias
Topical NSAID vs capsaicin	-	-	No data	-
Topical NSAIDs overall ES	-	-	No data	-
Capsaicin overall ES	On average, those in the capsaicin group experienced 0.65 SDs (0.10 to 1.20) pain relief compared to baseline	On average, those in the capsaicin group experienced 1.6 cm (0.3 to 3) pain relief compared to baseline	28 (1)	⊕○○○ Very low for study type, risk of bias, indirectness, imprecision, publication bias

*Due to the inclusion of crossover trials where treatment periods were combined, number of participants (n) describes the number of *comparisons* the randomised participants generated (note that a single participant may contribute two comparisons)
n, number of participants; SD, standard deviation; SMD, standardised mean difference; VAS, visual analogue scale

6.3.2 IPD MA

6.3.2.1 Description of studies

6.3.2.1.1 Topical NSAIDs

One eligible trial was identified for inclusion to the IPD MA. The author was contacted on multiple occasions, however, they did not respond. No IPD were received for topical NSAIDs in NP (Figure 6-8).

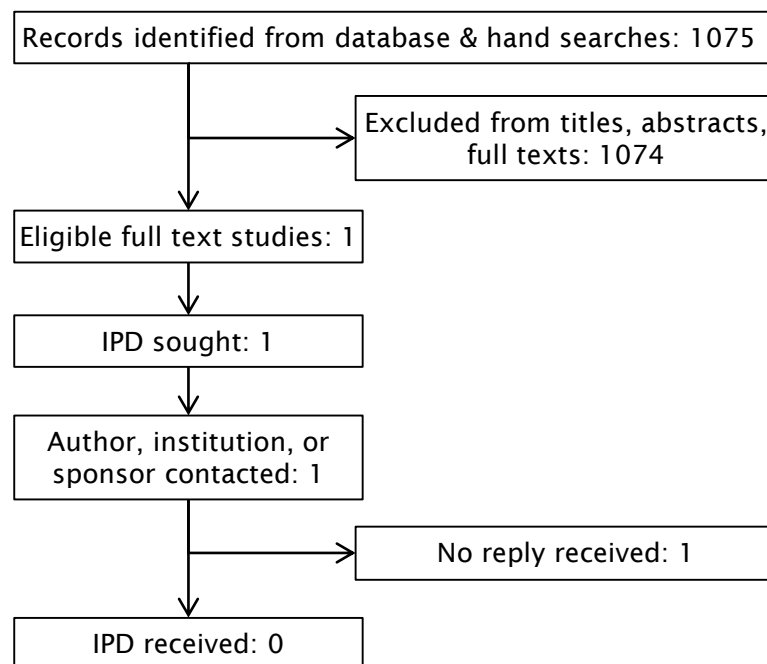


Figure 6-8- Flow diagram of trial identification, contact, and data acquisition for IPD MA of topical NSAIDs in NP

6.3.2.1.2 Capsaicin

IPD were sought from 21 studies; 19 identified from the CMA and two abstract-only publications (Figure 6-9). One study from the CMA randomised treatment by *leg*, rather than individual, and was not eligible for inclusion to the IPD MA. No reply was received for four of the studies. Of the replies received for the remaining 17 studies, 16 did not collaborate. Seven studies were held by a pharmaceutical company that declined collaboration on the terms that they do not share medical information with non-US based healthcare professionals. A further study author was not interested due to time-constraints. For

four of the studies, the person responding did not hold the data and was not able to locate it. One trialist responded that the original raw data were no longer available. Finally, four expressed interest in participating. Of these, one trial author did not respond to follow-up emails after initially expressing interest. The data for two studies were held by a pharmaceutical company that was interested in collaborating. The terms were discussed over telephone and an application for the data was made via clinicalstudydatarequest.com. However, during this process, the rights to share the data were sold to another pharmaceutical company. The contact provided for this pharmaceutical company failed to respond to attempts to obtain the IPD.

This pharmaceutical company, in addition to four others, had also been contacted in parallel as a potential pharmaceutical company that may hold IPD for capsaicin. No other pharmaceutical company responded to the contact attempts.

The IPD were obtained only for one study (Kulkantrakorn et al., 2012) and an IPD MA could not be conducted in NP.

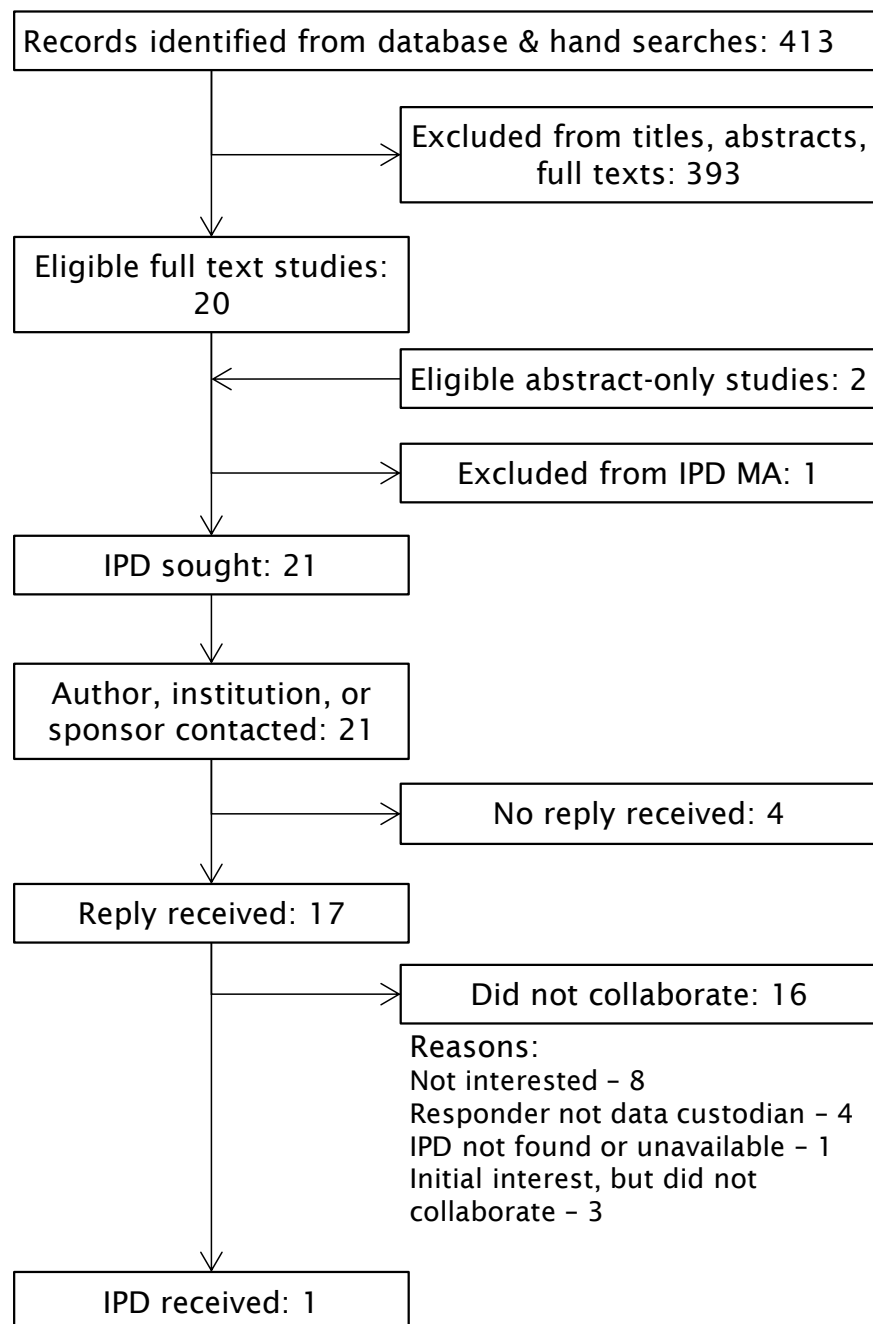


Figure 6-9 - Flow diagram of trial identification, contact, and data acquisition for IPD MA of capsaicin in NP

6.4 Discussion

6.4.1 Key findings and interpretation

Very poor evidence suggests that capsaicin cream is not superior to placebo for pain relief in NP. No conclusions could be drawn regarding the efficacy of topical NSAIDs in NP. Additionally, it was not possible to examine for predictors of response to topical NSAIDs or capsaicin in NP.

Although no significant difference was found between capsaicin and placebo, the specific effect estimate favoured placebo rather than active treatment. This effect was mainly driven by the findings of Kulkantrakorn et al. (2012), who found that at four weeks participants on capsaicin reported an increase in pain compared to baseline. In contrast, those in the placebo group had a slight decrease in pain compared to baseline. The resulting extreme ES in favour of placebo at four weeks consequently lowered the pooled effect to favour placebo. In contrast, the majority of other studies found a small, albeit non-significant, effect in favour of capsaicin. Although the study by (Kulkantrakorn et al., 2012) was considered the only study to implement appropriate methods for allocation concealment, its classification as high quality evidence may not be appropriate. Due to severe flooding at the research site, the study reported approximately 64% withdrawal rates. Although the publication states that ITT analysis was used, no methods for how missing data were imputed were described. Although drop-outs appeared balanced between the treatment arms, such rates of withdrawal may severely bias the results. In addition, typographical errors in the manuscript questions the quality of reporting and the peer review process. For example, although the figure indicated that VAS pain scores were approximately 50mm at four weeks in the capsaicin group, the same data presented in the table indicated that the pain scores were 4.91. Data at one and two weeks were in keeping with the figure and it was therefore assumed that the data at four weeks was presented with an error in decimal point (i.e., it should be 49.1 instead of 4.91). Efforts were made to confirm this assumption using the raw data provided by the

study authors, but the substantial missing data and unknown imputation assumptions meant that the publication findings could not be replicated. The large potential for bias and uncertainty regarding the data extracted indicates that these findings may not be reliable. The true effect of capsaicin is more likely reflected by the results from the studies with smaller effects favouring capsaicin. However, exclusion of the study in sensitivity analysis did not change the conclusion that there were no significant differences between capsaicin cream and placebo for NP.

The finding that capsaicin cream is not superior to placebo in NP is in line with the Cochrane review of low-dose capsaicin in NP (Derry and Moore, 2012). However, it is in contrast with much of the literature and recommendations that have concluded that capsaicin cream is effective for NP (Zhang and Li Wan Po, 1994, NICE, 2018, Mason et al., 2004, Ricard et al., 2013). This discrepancy is likely attributable to differences in the trials analysed. Duplicate publications containing the data for single centres from the multicentre trial performed by the Capsaicin Study Group in the early 1990's were excluded from the present MA (Tandan et al., 1992, Scheffler et al., 1991). No cross-reference was available in the publications and these publications were erroneously included in previous MAs (Zhang and Li Wan Po, 1994, NICE, 2018, Finnerup et al., 2015). Inclusion of covert duplicate publications, which are more likely to report a large treatment effect, is likely to have led previous authors to overestimate treatment effects (Tramèr et al., 1997). Differences in findings may further be explained by the exclusion of post-surgical pain from the present MA. These trials favoured capsaicin cream (non-significant) and their exclusion may have lowered the apparent treatment effect. However, participants with post-surgical pain do not exhibit purely NP and their exclusion from the present MA ensures that only NP was examined (Bischoff et al., 2014). Finally, the MA is updated to include new evidence, such as Teixeira et al. (2015), that did not find a significant difference between capsaicin cream and placebo.

The present MA provides the first evidence for the overall treatment effect and PCE of capsaicin in NP. The overall effect of capsaicin was

larger for OA than NP and this is in line with the commonly-held belief that NP is difficult to treat (Freynhagen and Bennett, 2009).

It was not possible to conduct an IPD MA for topical NSAIDs or capsaicin in NP, and the present work is therefore unable to further our exploration of individual predictors of response to the two treatments.

6.4.2 Caveats

The review is subject to several limitations. First, it was not possible to examine for predictors of response to topical NSAIDs or capsaicin in NP. Of the 23 trials that were identified, IPD were only received for one small trial. No hypotheses can be generated regarding predictors of response for people with neuropathic-like pain in OA.

Second, the confidence in the efficacy of capsaicin cream in NP is very low secondary to inconsistency, imprecision, publication bias, and high risks of bias across multiple domains. In particular, the majority of trials failed to adequately blind participants. In addition, the overall ES and PCE of capsaicin could only be examined in one, small study. The true efficacy of capsaicin is likely to be substantially different from the observed estimate.

Third, it may not have been appropriate to pool disparate presentations of outcome data. In order to minimise systematic losses of data, mean change from baseline, endpoint scores, percent change from baseline, and the proportion that improved were converted into SMD and pooled in MA. Differences in the effect estimates across trials may therefore be reflective of the method of outcome assessment rather than true differences in effect.

Fourth, subgroup analyses were conducted *post hoc* in limited numbers of trials/participants. As such, the analyses do not necessarily exclude clinically significant benefit in those subgroups. Inversely, spurious significant subgroup effects may have been found simply by chance. As these estimates were conducted at the study-level, the subgroup effects may be subject to ecological bias.

Finally, the efficacy of the high-concentration capsaicin patch was not examined in the present MA as study controls used capsaicin concentrations greater than the “active” concentration in the capsaicin creams. Although 0.04% capsaicin was used in the control patches for blinding purposes, it does not entail that the control patch is therapeutically inert. Current evidence indicates that the high-concentration capsaicin patch is effective in NP (Ricard et al., 2013, Finnerup et al., 2015, Derry et al., 2017).

7 DISCUSSION

7.1 Key findings and interpretation

There is moderate level evidence to suggest that topical NSAIDs may be effective for pain relief in OA. There is very low level evidence that capsaicin, when used at 0.025% concentration, may be effective for OA. Absolute levels of pain relief from the treatments, compared to baseline, are likely to be clinically meaningful.

On average, topical NSAIDs and capsaicin appear equally effective. However, response to treatment varies greatly between people. Reliance on study-level average effects alone, without consideration of individual variations in response, may be misleading. No baseline characteristics robustly predicted a preference to either treatment, and consideration of the treatments in the clinical setting should be done on an individual basis. Physicians and patients may guide decision-making through direct comparison of the treatments, in an informal n-of-1 trial, if appropriate.

Examination of potential predictors of response throughout this thesis were exploratory and subgroup findings should therefore be interpreted with caution. There is some indication that women respond better to topical NSAIDs, however, the difference in effect between the sexes is unlikely to be clinically detectable and may not be attributable to sex differences alone. It may be that other unmeasured predictors mediate the observed interaction. The evidence suggests that women report greater overall treatment effects regardless of treatment, but this is not independent of baseline pain severity. Increased pain at baseline significantly associated with greater overall treatment effects. This is thought to reflect regression to the mean rather than any biological mechanism. Finally, inconsistent findings suggest that individuals with greater levels of inflammation in the knee may benefit more from topical NSAIDs than capsaicin. Not only do topical NSAIDs and capsaicin have different hypothesised mechanisms of action which affect their specific effect, but they also have different contextual effects. Whilst capsaicin cream feels warm on application, NSAID gels

may give a cooling sensation due to the addition of ethanol. Treatment preference may reflect an interaction between the biologically disparate mechanisms of pain production and the treatment's specific mechanism of action. Alternatively, it may represent an interaction between the participant's phenotype and the contextual effect of the treatment.

The study-level findings for capsaicin in NP may be extrapolated to OA with caution. The lack of efficacy of capsaicin in true NP may indicate that participants with neuropathic-like pain in OA either a) would not benefit from capsaicin cream, or b) require greater concentrations of capsaicin. Previous work has indicated that high concentration capsaicin is effective in NP (Derry et al., 2017, Finnerup et al., 2015), indicating that individuals with NP may require a greater exposure to capsaicin to gain clinically significant pain relief. This would suggest that individuals with neuropathic-like pain in OA may benefit from higher concentrations of capsaicin.

The combination of evidence synthesis methods allows greater insight into the efficacy of a treatment. MAs of available RCT evidence are efficient and able to provide meaningful findings that may aid shared decision making. However, they are limited by the availability of data and may not be able to capture the variation observed in clinical practice. These methods may therefore need to be supplemented with evidence from a primary trial designed to examine the relative efficacy and important predictors of response to treatment. The primary synthesis of data, rather than the reliance on RCT data, allows the examination of baseline characteristics that are of hypothesised importance, such as SH or neuropathic-like pain.

7.2 Future work

There is a relative paucity of evidence for the benefit of capsaicin in OA. Rigorously conducted, large, and adequately blinded trials for capsaicin cream in OA are required. One such trial has been conducted and published (Schnitzer et al., 2012), although it was not included in the present MA as the control capsaicin concentration was similar to

the “active” concentration used for other studies. This trial illustrates that it is possible to conduct high quality trials for capsaicin in OA and these should be done. In addition, one should consider investigating the use of the 8% capsaicin patch in OA. Future trials in OA should record pain characteristics at baseline (e.g., presence of neuropathic-like pain) as these may provide valuable information for precision medicine.

Potential predictors of response to topical NSAIDs and capsaicin were identified in the present work. Targeted investigation of the association between US SH, markers of central sensitisation, and treatment with topical NSAID and capsaicin may be warranted. Future trials should be adequately blinded and include a placebo comparison, in order to allow separate examination of the specific and contextual treatment effects.

There is a need to increase the awareness and interest in IPD MAs. They are a valuable resource for the methodologically-robust examination of predictors of response, but are profoundly limited by data availability. Steps to increase the provision of IPD are required, which, at their core, involve a fundamental change in the attitude of researchers. Policies which require data custodians to make IPD available may prove invaluable for future research. RCTs do not aim to examine for predictors of response, and so do not commonly assess characteristics that may modify treatment. Recommendations for a list of standardised characteristics, relevant to precision medicine, that are required to be measured at baseline in all RCTs should be developed. The measurement of standardised, important baseline characteristics across RCTs would facilitate the examination of important predictors for future IPD MAs.

Methodological guidelines for the conduct and presentation of IPD MAs should be developed and made readily accessible to non-statisticians. There is a multitude of methodological publications available, but the continued use of inappropriate statistical methods (Hua et al., 2017) indicates the need for clearer recommendations. Methods for the extension of IPD MAs to examine predictors of the overall treatment

effect and to include repeated measures need to be refined. Finally, GRADE guidelines specific to IPD MAs need to be developed.

N-of-1 trial series allow the examination of potential predictors of response. However, the trial duration and relative resource-intensity may outweigh the increased precision gained from multiple treatment cycles. The examination of predictors of response using an n-of-1 trial series should be compared to a traditional cross-over trial design to determine whether inclusion of additional cycles is justified. If justified, there is a need to develop and disseminate methodology for sample size calculations in n-of-1 trial series investigating treatment-by-covariate interactions.

The use of clinically meaningful differences in the investigation of predictors of response should be examined. The effect of applying the MCID to IPD MAs should be compared to the approaches based on continuous data adopted in the present work. Primary trials should consider including direct assessments of whether a treatment generated clinically meaningful pain reductions, instead of relying purely on the back-translation of continuous outcome data using MCID thresholds.

REFERENCES

- ABRAMSON, S. 2003. The role of NSAIDs in the treatment of osteoarthritis. *In*: BRANDT, K., DOHERTY, M. & LOHMANDER, L. S. (eds.) *Osteoarthritis*. 2nd ed ed. Oxford: Oxford University Press.
- AHMED, I., SUTTON, A. J. & RILEY, R. D. 2012. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*, 344, d7762.
- AIGNER, T., SCHMITZ, N. & SALTER, D. M. 2015. Pathogenesis and pathology of osteoarthritis. *In*: HOCHBERG, M. C., SILMAN, A. J., SMOLEN, J. S., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed. ed. Philadelphia: Elsevier Mosby.
- AKIN-AKINYOSOYE, K., FROWD, N., MARSHALL, L., STOCKS, J., FERNANDES, G. S., VALDES, A., MCWILLIAMS, D. F., ZHANG, W., DOHERTY, M. & FERGUSON, E. 2018. Traits associated with central pain augmentation in the Knee Pain In the Community (KPIC) cohort. *Pain*, 159, 1035.
- ALLEN, K. D., COFFMAN, C. J., GOLIGHTLY, Y. M., STECHUCHAK, K. M. & KEEFE, F. J. 2009. Daily pain variations among patients with hand, hip, and knee osteoarthritis. *Osteoarthritis and Cartilage*, 17, 1275-1282.
- ALTMAN, R., ALARCON, G., APPELROUTH, D., BLOCH, D., BORENSTEIN, D., BRANDT, K., BROWN, C., COOKE, T., DANIEL, W. & FELDMAN, D. 1991. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis & Rheumatism*, 34, 505-514.
- ALTMAN, R., ALARCON, G., APPELROUTH, D., BLOCH, D., BORENSTEIN, D., BRANDT, K., BROWN, C., COOKE, T., DANIEL, W. & GRAY, R. 1990. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis & Rheumatism*, 33, 1601-1610.
- ALTMAN, R., ASCH, E., BLOCH, D., BOLE, G., BORENSTEIN, D., BRANDT, K., CHRISTY, W., COOKE, T., GREENWALD, R., HOCHBERG, M., HOWELL, D., KAPLAN, D., KOOPMAN, W., LONGLEY III, S., MANKIN, H., MCSHANE, D., MEDSGER JR, T., MEENAN, R., MIKKELSEN, W., MOSKOWITZ, R., MURPHY, W., ROTHCHILD, R., SEGAL, M., SOKOLOFF, L. & WOLFE, F. 1986. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis & Rheumatism*, 29, 1039-1049.
- ALTMAN, R., BOSCH, B., BRUNE, K., PATRIGNANI, P. & YOUNG, C. 2015. Advances in NSAID Development: Evolution of Diclofenac Products Using Pharmaceutical Technology. *Drugs*, 75, 859-877.

- ALTMAN, R., BRANDT, K., HOCHBERG, M., MOSKOWITZ, R., BELLAMY, N., BLOCH, D. A., BUCKWALTER, J., DOUGADOS, M., EHRLICH, G., LEQUESNE, M., LOHMANDER, S., MURPHY JR, W. A., ROSARIO-JANSEN, T., SCHWARTZ, B. & TRIPPEL, S. 1996. Design and conduct of clinical trials in patients with osteoarthritis: Recommendations from a task force of the Osteoarthritis Research Society: Results from a workshop. *Osteoarthritis and Cartilage*, 4, 217-243.
- ALTMAN, R. D. 2015. Clinical features of osteoarthritis. *In*: HOCHBERG, M. C., SILMAN, A. J., SMOLEN, J., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed ed. Philadelphia: Elsevier Mosby.
- ALTMAN, R. D., AVEN, A., HOLMBURG, C. E., PFEIFER, L. M., SACK, M. & YOUNG, G. T. 1994. Capsaicin cream 0.025% as monotherapy for osteoarthritis: A double-blind study. *Seminars in Arthritis and Rheumatism*, 23, 25-33.
- ALTMAN, R. D. & BARTHEL, H. R. 2011. Topical Therapies for Osteoarthritis. *Drugs*, 71, 1259-1279.
- ALTMAN, R. D., DREISER, R. L., FISHER, C. L., CHASE, W. F., DREHER, D. S. & ZACHER, J. 2009. Diclofenac Sodium Gel in Patients with Primary Hand Osteoarthritis: A Randomized, Double-blind, Placebo-controlled Trial. *Journal of Rheumatology*, 36, 1991-1999.
- ALTMAN, R. D. & HOCHBERG, M. C. 2015. Management of Osteoarthritis. *In*: HOCHBERG, M. C., SILMAN, A. J., SMOLEN, J., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed. Philadelphia: Elsevier Mosby.
- ANAND P, B. K. 2011. Topical capsaicin for pain management: therapeutic potential and mechanisms of action of the new high-concentration capsaicin 8% patch. *British Journal of Anaesthesia*, 107, 490-502.
- ARENDT-NIELSEN, L. 2017. Pain sensitisation in osteoarthritis. *Clin Exp Rheumatol*, 35, S68-S74.
- ARENDT-NIELSEN, L., EGSGAARD, L. L. & PETERSEN, K. K. 2016. Evidence for a central mode of action for etoricoxib (COX-2 inhibitor) in patients with painful knee osteoarthritis. *Pain*, 157, 1634-1644.
- ATTAL, N., CRUCCU, G., BARON, R., HAANPÄÄ, M., HANSSON, P., JENSEN, T. S. & NURMIKKO, T. 2010. EFNS guidelines on the pharmacological treatment of neuropathic pain: 2010 revision. *European Journal of Neurology*, 17, 1113-1188.
- AVERBUCH, M. & KATZPER, M. 2000. A search for sex differences in response to analgesia. *Archives of internal medicine*, 160, 3424-3428.

- BAER, P. A., THOMAS, L. M. & SHAINHOUSE, Z. 2005. Treatment of osteoarthritis of the knee with a topical diclofenac solution: a randomised controlled, 6-week trial [ISRCTN53366886]. *BMC Musculoskeletal Disorders*, 6, 44.
- BALSHAM, H., HELFAND, M., SCHÜNEMANN, H. J., OXMAN, A. D., KUNZ, R., BROZEK, J., VIST, G. E., FALCK-YTTER, Y., MEERPOHL, J. & NORRIS, S. 2011. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64, 401-406.
- BARAF, H. S. B., GLOTH, F. M., BARTHEL, H. R., GOLD, M. S. & ALTMAN, R. D. 2011. 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. *Drugs and Aging*, 28, 27-40.
- BARAF, H. S. B., GOLD, M. S., CLARK, M. B. & ALTMAN, R. D. 2010. Safety and Efficacy of Topical Diclofenac Sodium 1% Gel in Knee Osteoarthritis: A Randomized Controlled Trial. *Physician and Sportsmedicine*, 38, 19-28.
- BARDEN, J., EDWARDS, J. E., MOORE, R. A. & MCQUAY, H. J. 2002. Ibuprofen 400 mg is effective in women, and women are well represented in trials. *BMC Anesthesiology*, 2, 6.
- BARNETT, A. G., VAN DER POLS, J. C. & DOBSON, A. J. 2004. Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, 34, 215-220.
- BARON, R., BINDER, A. & WASNER, G. 2010. Neuropathic pain: diagnosis, pathophysiological mechanisms, and treatment. *The Lancet Neurology*, 9, 807-819.
- BARR, A. J., CAMPBELL, T. M., HOPKINSON, D., KINGSBURY, S. R., BOWES, M. A. & CONAGHAN, P. G. 2015. A systematic review of the relationship between subchondral bone features, pain and structural pathology in peripheral joint osteoarthritis. *Arthritis Res Ther*, 17, 228.
- BARTHEL, H. R., HASSELWOOD, D., LONGLEY, S., GOLD, M. S. & ALTMAN, R. D. 2009. Randomized controlled trial of diclofenac sodium gel in knee osteoarthritis. *Seminars in arthritis and rheumatism*, 39, 203-212.
- BARTHEL, H. R., PENISTON, J. H., CLARK, M. B., GOLD, M. S. & ALTMAN, R. D. 2010. Correlation of pain relief with physical function in hand osteoarthritis: randomized controlled trial post hoc analysis. *Arthritis Research & Therapy*, 12, R7.
- BARTLEY, E. J. & FILLINGIM, R. B. 2013. Sex differences in pain: a brief review of clinical and experimental findings. *British journal of anaesthesia*, 111, 52-58.

- BATTAGGIA, A., LORA APRILE, P., CRICELLI, I., FORNASARI, D., FANELLI, A., CRICELLI, C. & LAPI, F. 2016. Paracetamol: a probably still safe drug. *Annals of the Rheumatic Diseases*, 75, e57-e57.
- BELLAMY, N. 2014. Principles of clinical outcome assessment. *In*: HOCHBERG, M. C., SILMAN, A. J., SMOLEN, J., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed. Philadelphia: Elsevier.
- BELLAMY, N., BUCHANAN, W. W., GOLDSMITH, C. H., CAMPBELL, J. & STITT, L. W. 1988. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *Journal of Rheumatology*, 15, 1833-1840.
- BELLAMY, N., CAMPBELL, J., HARAOU, B., BUCHBINDER, R., HOBBY, K., ROTH, J. H. & MACDERMID, J. C. 2002. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis and Cartilage*, 10, 855-862.
- BERENBAUM, F. 2013. Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). *Osteoarthritis and Cartilage*, 21, 16-21.
- BERNSTEIN, J. E., KORMAN, N. J., BICKERS, D. R., DAHL, M. V. & MILLIKAN, L. E. 1989. Topical capsaicin treatment of chronic postherpetic neuralgia. *Journal of the American Academy of Dermatology*, 21, 265-270.
- BESWICK, A. D., WYLDE, V., GOOBERMAN-HILL, R., BLOM, A. & DIEPPE, P. 2012. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open*, 2.
- BIERMA-ZEINSTRA, S. M. & VERHAGEN, A. P. 2011. Osteoarthritis subpopulations and implications for clinical trial design. *Arthritis Research & Therapy*, 13, 213.
- BIESTERBOS, J. W. H., DUDZINA, T., DELMAAR, C. J. E., BAKKER, M. I., RUSSEL, F. G. M., VON GOETZ, N., SCHEEPERS, P. T. J. & ROELEVELD, N. 2013. Usage patterns of personal care products: Important factors for exposure assessment. *Food and Chemical Toxicology*, 55, 8-17.
- BINGHAM III, C. O., SMUGAR, S. S., WANG, H., PELOSO, P. M. & GAMMAITONI, A. 2011. Predictors of response to cyclo-oxygenase - 2 inhibitors in osteoarthritis: pooled results from two identical trials comparing etoricoxib, celecoxib, and placebo. *Pain Medicine*, 12, 352-361.

- BISCHOFF, H. A., ROOS, E. M. & LIANG, M. H. 2003. Outcome assessment in osteoarthritis: a guide for research and clinical practice. In: BRANDT, K. D., DOHERTY, M. & LOHMANDER, L. S. (eds.) *Osteoarthritis*. 2nd ed. Oxford: Oxford University Press.
- BISCHOFF, J. M., RINGSTED, T. K., PETERSEN, M., SOMMER, C., UCEYLER, N. & WERNER, M. U. 2014. A capsaicin (8%) patch in the treatment of severe persistent inguinal postherniorrhaphy pain: a randomized, double-blind, placebo-controlled trial. *PLoS ONE*, 9, e109144.
- BISWAL, S., MEDHI, B. & PANDHI, P. 2006. Longterm efficacy of topical nonsteroidal antiinflammatory drugs in knee osteoarthritis: Metaanalysis of randomized placebo controlled clinical trials. *Journal of Rheumatology*, 33, 1841-1844.
- BJELLAND, I., DAHL, A. A., HAUG, T. T. & NECKELMANN, D. 2002. The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, 52, 69-77.
- BJORDAL, J. M., KLOVNING, A., LJUNGGREN, A. E. & SLØRDAL, L. 2007. Short - term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: A meta - analysis of randomised placebo - controlled trials. *European Journal of Pain*, 11, 125-138.
- BLACKBURN, S., RESEARCH USER, G., RHODES, C., HIGGINBOTTOM, A. & DZIEDZIC, K. 2016. The OARSI standardised definition of osteoarthritis: A lay version. *Osteoarthritis and Cartilage*, 24, Supplement 1, S192.
- BLAGOJEVIC, M., JINKS, C., JEFFERY, A. & JORDAN, K. P. 2010. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. *Osteoarthritis and Cartilage*, 18, 24-33.
- BNF. 2018. *British National Formulary (online)* [Online]. London: BMJ Group and Royal Pharmaceutical Society of Great Britain. Available: <https://bnf.nice.org.uk/> [Accessed 09/01/2019].
- BOOKMAN, A., WILLIAMS, K. & SHAINHOUSE, J. 2004. Effect of a topical diclofenac solution for relieving symptoms of primary osteoarthritis of the knee: a randomized controlled trial. *Canadian Medical Association Journal*, 171, 333-8.
- BORENSTEIN, M., HEDGES, L., HIGGINS, J. & ROTHSTEIN, H. 2011. *Introduction to Meta-Analysis*, Chichester, John Wiley & Sons Ltd.
- BRIGNARDELLO-PETERSEN, R., BONNER, A., ALEXANDER, P. E., SIEMIENIUK, R. A., FURUKAWA, T. A., ROCHWERG, B., HAZLEWOOD, G. S., ALHAZZANI, W., MUSTAFA, R. A. & MURAD, M. H. 2018. Advances in the GRADE approach to rate the

- certainty in estimates from a network meta-analysis. *Journal of Clinical Epidemiology*, 93, 36-44.
- BROADBENT, E., PETRIE, K. J., MAIN, J. & WEINMAN, J. 2006. The brief illness perception questionnaire. *Journal of Psychosomatic Research*, 60, 631-637.
- BROOKES, S. T., BIDDLE, L., PATERSON, C., WOOLHEAD, G. & DIEPPE, P. 2007. "Me's me and you's you": Exploring patients' perspectives of single patient (n-of-1) trials in the UK. *Trials*, 8, 10.
- BROOKES, S. T., WHITLEY, E., PETERS, T. J., MULHERAN, P. A., EGGER, M. & DAVEY SMITH, G. 2001. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health technology assessment*, 5, 1-56.
- BRÜHLMANN, P. & MICHEL, B. A. 2003. Topical diclofenac patch in patients with knee osteoarthritis: a randomized, double-blind, controlled clinical trial. *Clinical and experimental rheumatology*, 21, 193-8.
- BRUYÈRE, O., COOPER, C., ARDEN, N., BRANCO, J., BRANDI, M. L., HERRERO-BEAUMONT, G., BERENBAUM, F., DENNISON, E., DEVOGELAER, J.-P., HOCHBERG, M., KANIS, J., LASLOP, A., MCALINDON, T., REITER, S., RICHETTE, P., RIZZOLI, R. & REGINSTER, J.-Y. 2015. Can We Identify Patients with High Risk of Osteoarthritis Progression Who Will Respond to Treatment? A Focus on Epidemiology and Phenotype of Osteoarthritis. *Drugs & Aging*, 32, 179-187.
- BURIAN M, G. G. 2005. COX-dependent mechanisms involved in the antinociceptive action of NSAIDs at central and peripheral sites. *Pharmacology & Therapeutics*, 139-154.
- BURKE, D. L., ENSOR, J. & RILEY, R. D. 2017. Meta - analysis using individual participant data: one - stage and two - stage approaches, and why they may differ. *Statistics in medicine*, 36, 855-875.
- BUTUROVIC, L. 2018. A machine learning method for subgroup analysis of randomized controlled trials. *bioRxiv*, 338996.
- CAMERON, M., GAGNIER, J. J., LITTLE, C. V., PARSONS, T. J., BLÜMLE, A., CHRUBASIK, S., LI, Y. Q., MIZOKAMI, A., IZUMI, K. & NARIMOTO, K. 2009. Evidence of effectiveness of herbal medicinal products in the treatment of arthritis. *Phytother Res*, 23, 1497-1515.
- CARNES, D., ANWER, Y., UNDERWOOD, M., HARDING, G. & PARSONS, S. 2008. Influences on older people's decision making regarding choice of topical or oral NSAIDs for knee pain: qualitative study. *BMJ*, 336, 142-145.
- CHAD, D. A., ARONIN, N., LUNDSTROM, R., MCKEON, P., ROSS, D., MOLITCH, M., SCHIPPER, H. M., STALL, G., DYESS, E. & TARSY, D.

1990. Does capsaicin relieve the pain of diabetic neuropathy? *Pain*, 42, 387-388.
- CHEN, A., GUPTE, C., AKHTAR, K., SMITH, P. & COBB, J. 2012. The global economic cost of osteoarthritis: how the UK compares. *Arthritis*, 2012, e698709.
- COHEN, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- COLLINS, F. S. & VARMUS, H. 2015. A new initiative on precision medicine. *New England Journal of Medicine*, 372, 793-795.
- COLLINS, N. J., MISRA, D., FELSON, D. T., CROSSLEY, K. M. & ROOS, E. M. 2011. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care & Research*, 63, S208-S228.
- CONAGHAN, P. G., DICKSON, J., BOLTEN, W., CEVC, G. & ROTHER, M. 2013. A multicentre, randomized, placebo- and active-controlled trial comparing the efficacy and safety of topical ketoprofen in Transfersome gel (IDEA-033) with ketoprofen-free vehicle (TDT 064) and oral celecoxib for knee pain associated with osteoarthritis. *Rheumatology*, 52, 1303-1312.
- COPSEY, B., THOMPSON, J. Y., VADHER, K., ALI, U., DUTTON, S. J., FITZPATRICK, R., LAMB, S. E. & COOK, J. A. 2018. Problems persist in reporting of methods and results for the WOMAC measure in hip and knee osteoarthritis trials. *Quality of Life Research*, 1-9.
- CRUCCU, G., ANANDC, P., ATTALD, N., GARCIA-LARREAA, L., HAANPÄÄ, M., JØRUMA, E., SERRAA, J. & JENSENA, T. 2004. EFNS guidelines on assessment of neuropathic pain and treatment. *European Journal of Neurology*, 11, 1-10.
- CSG 1991. Treatment of painful diabetic neuropathy with topical capsaicin. A multicenter, double-blind, vehicle-controlled study. The Capsaicin Study Group. *Archives of Internal Medicine*, 151, 2225-9.
- D'AGOSTINO, M. A., CONAGHAN, P., LE BARS, M., BARON, G., GRASSI, W., MARTIN-MOLA, E., WAKEFIELD, R., BRASSEUR, J.-L., SO, A., BACKHAUS, M., MALAISE, M., BURMESTER, G., SCHMIDELY, N., RAVAUD, P., DOUGADOS, M. & EMERY, P. 2005. EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1:

Prevalence of inflammation in osteoarthritis. *Annals of the Rheumatic Diseases*, 64, 1703-1709.

- DEAL, C. L., SCHNITZER, T. J., LIPSTEIN, E., SEIBOLD, J. R., STEVENS, R. M., LEVY, M. D., ALBERT, D. & RENOLD, F. 1991. Treatment of arthritis with topical capsaicin: a double-blind trial. *Clinical Therapeutics*, 13, 383-95.
- DEBRAY, T. P. A., MOONS, K. G. M., VALKENHOEF, G., EFTHIMIOU, O., HUMMEL, N., GROENWOLD, R. H. H. & REITSMA, J. B. 2015. Get real in individual participant data (IPD) meta - analysis: a review of the methodology. *Research Synthesis Methods*, 6, 293-309.
- DENG, Z.-H., ZENG, C., YANG, Y., LI, Y.-S., WEI, J., YANG, T., LI, H. & LEI, G.-H. 2016. Topical diclofenac therapy for osteoarthritis: a meta-analysis of randomized controlled trials. *Clinical Rheumatology*, 35, 1253-1261.
- DEO, R. C. 2015. Machine learning in medicine. *Circulation*, 132, 1920-1930.
- DEQUEKER, J. & LUYTEN, F. 2008. The history of osteoarthritis-osteoarthrosis. *Annals of the Rheumatic Diseases*, 67, 5-10.
- DERRY, S., CONAGHAN, P., DA SILVA, J. A. P., WIFFEN, P. J. & MOORE, R. A. 2016. Topical NSAIDs for chronic musculoskeletal pain in adults. *Cochrane Database of Systematic Reviews*, 4.
- DERRY, S., MATTHEWS, P. R., WIFFEN, P. J. & MOORE, R. A. 2014. Salicylate-containing rubefacients for acute and chronic musculoskeletal pain in adults. *Cochrane Database of Systematic Reviews*, 11.
- DERRY, S. & MOORE, R. A. 2012. Topical capsaicin (low concentration) for chronic neuropathic pain in adults. *Cochrane Database of Systematic Reviews*, 9.
- DERRY S, M. R., RABBIE R 2012. Topical NSAIDs for chronic musculoskeletal pain in adults (Review). *Cochrane Database of Systematic Reviews*, 9.
- DERRY S, R. A., COLE P, TAN T, MOORE RA 2013. Topical capsaicin (high concentration) for chronic neuropathic pain in adults. *Cochrane Database of Systematic Reviews*, 2.
- DERRY, S., RICE, A. S., COLE, P., TAN, T. & MOORE, R. A. 2017. Topical capsaicin (high concentration) for chronic neuropathic pain in adults. *Cochrane Database of Systematic Reviews*, 1.
- DETORA, L. M., KRUPA, D., BOLOGNESE, J., SPERLING, R. S. & EHRICH, E. W. 2001. Rofecoxib shows consistent efficacy in osteoarthritis clinical trials, regardless of specific patient demographic and disease factors. *Journal of Rheumatology*, 28, 2494-2503.

- DEVEZA, L. A., MELO, L., YAMATO, T. P., MILLS, K., RAVI, V. & HUNTER, D. J. 2017. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. *Osteoarthritis Cartilage*, 25, 1926-1941.
- DIAS, S., WELTON, N. J., SUTTON, A. J. & ADES, A. E. 2011, updated 2016. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. Available: <http://www.nicedsu.org.uk> [Accessed 21/09/2017].
- DREISER, R. L. & TISNE-CAMUS, M. 1993. DHEP plasters as a topical treatment of knee osteoarthritis - A double-blind placebo-controlled study. *Drugs under Experimental and Clinical Research*, 19, 117-123.
- DUVAL, S. & TWEEDIE, R. 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89-98.
- DWORKIN, R. H., O'CONNOR, A. B., AUDETTE, J., BARON, R., GOURLAY, G. K., HAANPÄÄ, M. L., KENT, J. L., KRANE, E. J., LEBEL, A. A., LEVY, R. M., MACKEY, S. C., MAYER, J., MIASKOWSKI, C., RAJA, S. N., RICE, A. S. C., SCHMADER, K. E., STACEY, B., STANOS, S., TREEDE, R.-D., TURK, D. C., WALCO, G. A. & WELLS, C. D. 2010. Recommendations for the Pharmacological Management of Neuropathic Pain: An Overview and Literature Update. *Mayo Clinic Proceedings*, 85, S3-S14.
- ECKSTEIN, F., GUERMAZI, A., GOLD, G., DURYEA, J., HELLIO LE GRAVERAND, M. P., WIRTH, W. & MILLER, C. G. 2014. Imaging of cartilage and bone: promises and pitfalls in clinical trials of osteoarthritis. *Osteoarthritis and Cartilage*, 22, 1516-1532.
- EDWARDS, R. R., DOLMAN, A. J., MARTEL, M. O., FINAN, P. H., LAZARIDOU, A., CORNELIUS, M. & WASAN, A. D. 2016. Variability in conditioned pain modulation predicts response to NSAID treatment in patients with knee osteoarthritis. *BMC Musculoskeletal Disorders*, 17, 1-9.
- EGGER, M., SMITH, G. D., SCHNEIDER, M. & MINDER, C. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629-634.
- ELDRIDGE, S. M., LANCASTER, G. A., CAMPBELL, M. J., THABANE, L., HOPEWELL, S., COLEMAN, C. L. & BOND, C. M. 2016. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PLoS one*, 11, e0150205.
- EMA. 2010. *Guideline on clinical investigation of medicinal products used in the treatment of osteoarthritis: CPMP/EWP/784/97 Rev. 1* [Online]. London: European Medicines Agency. Available: <http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulatio>

[n/general/general_content_001135.jsp&mid=WC0b01ac0580034cf4](http://www.medicines.org.uk/emc/medicine/19985) [Accessed 22/05/2018 2018].

- EMC. 2017. *Voltarol 1.16% Emulgel, gel - Summary of Product Characteristics* [Online]. Available: <http://www.medicines.org.uk/emc/medicine/19985> [Accessed 28/11/2017].
- ERGUN, H., KULCU, D., KUTLAY, S., BODUR, H. & TULUNAY, F. C. 2007. Efficacy and safety of topical nimesulide in the treatment of knee osteoarthritis. *Journal of Clinical Rheumatology*, 13, 251-5.
- EVANGELOU, E., TSIANOS, G. & IOANNIDIS, J. P. A. 2008. Doctors' versus patients' global assessments of treatment effectiveness: empirical survey of diverse treatments in clinical trials. *BMJ*, 336, 1287-1290.
- FELSON, D. T. 2013. Osteoarthritis as a disease of mechanics. *Osteoarthritis and Cartilage*, 21, 10-15.
- FELSON, D. T., LAWRENCE, R. C., DIEPPE, P. A., HIRSCH, R., HELMICK, C. G., JORDAN, J. M., KINGTON, R. S., LANE, N. E., NEVITT, M. C., ZHANG, Y., SOWERS, M., MCALINDON, T., SPECTOR, T. D., POOLE, A. R., YANOVSKI, S. Z., ATESHIAN, G., SHARMA, L., BUCKWALTER, J. A., BRANDT, K. D. & FRIES, J. F. 2000. Osteoarthritis: New Insights. Part 1: The Disease and Its Risk Factors. *Annals of Internal Medicine*, 133, 635-646.
- FELSON, D. T. & ZHANG, Y. 2015. Local and systemic risk factors for incidence and progression of osteoarthritis. In: HOCHBERG, M. C., SILMAN, A. J., SMOLEN, J., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed ed. Philadelphia: Elsevier Mosby.
- FERNANDES, G. S., SARMANOVA, A., WARNER, S., HARVEY, H., AKIN-AKINYOSOYE, K., RICHARDSON, H., FROWD, N., MARSHALL, L., STOCKS, J. & HALL, M. 2017. Knee pain and related health in the community study (KPIC): a cohort study protocol. *BMC musculoskeletal disorders*, 18, 404.
- FERNANDES, G. S., VALDES, A. M., WALSH, D. A., ZHANG, W. & DOHERTY, M. 2018. Neuropathic-like knee pain and associated risk factors: a cross-sectional study in a UK community sample. *Arthritis Research & Therapy*, 20, 215.
- FERNANDES, L., HAGEN, K. B., BIJLSMA, J. W., ANDREASSEN, O., CHRISTENSEN, P., CONAGHAN, P. G., DOHERTY, M., GEENEN, R., HAMMOND, A. & KJEKEN, I. 2013. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Annals of the Rheumatic Diseases*, 72, 1125-1135.
- FINGLETON, C., SMART, K., MOLONEY, N., FULLEN, B. M. & DOODY, C. 2015. Pain sensitization in people with knee osteoarthritis: a

systematic review and meta-analysis. *Osteoarthritis and Cartilage*, 23, 1043-1056.

FINNERUP, N. B., ATTAL, N., HAROUTOUNIAN, S., MCNICOL, E., BARON, R., DWORKIN, R. H., GILRON, I., HAANPÄÄ, M., HANSSON, P., JENSEN, T. S., KAMERMAN, P. R., LUND, K., MOORE, A., RAJA, S. N., RICE, A. S. C., ROWBOTHAM, M., SENA, E., SIDDALL, P., SMITH, B. H. & WALLACE, M. 2015. Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis. *The Lancet Neurology*, 14, 162-173.

FISH, D., KRETZMANN, H., BRANTINGHAM, J. W., GLOBE, G., KORPORAAL, C. & MOEN, J. R. 2008. A randomized clinical trial to determine the effect of combining a topical capsaicin cream and knee-joint mobilization in the treatment of osteoarthritis of the knee. *Journal of the American Chiropractic Association*, 45, 8-23.

FISHER, D. 2017. IPDMETAN: Stata module for performing two-stage IPD meta-analysis.

FRENCH, H. P., SMART, K. M. & DOYLE, F. 2017. Prevalence of neuropathic pain in knee or hip osteoarthritis: a systematic review and meta-analysis. *Seminars in Arthritis and Rheumatism*, 47, 1-8.

FREYNHAGEN, R., BARON, R., GOCKEL, U. & TÖLLE, T. R. 2006. Pain DETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Current medical research and opinion*, 22, 1911-1920.

FREYNHAGEN, R. & BENNETT, M. I. 2009. Diagnosis and management of neuropathic pain. *BMJ*, 339.

FU, R., VANDERMEER, B. W., SHAMLIYAN, T. A., O'NEIL, M. E., YAZDI, F., FOX, S. H. & MORTON, S. C. 2013. Handling continuous outcomes in quantitative synthesis. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Agency for Healthcare Research and Quality.

GABLER, N. B., DUAN, N., VOHRA, S. & KRAVITZ, R. L. 2011. N-of-1 trials in the medical literature: a systematic review. *Medical care*, 49, 761-768.

GANDEK, B. 2015. Measurement properties of the Western Ontario and McMaster Universities Osteoarthritis Index: a systematic review. *Arthritis Care & Research*, 67, 216-229.

GDPR. 2016. *General Data Protection Regulation (EU) 2016/679 - cor. OJ L 127, 23.5.2018* [Online]. Available: <https://gdpr-info.eu/> [Accessed 11/12/2018 2018].

GEENEN, R., OVERMAN, C. L., CHRISTENSEN, R., ÅSENLÖF, P., CAPELA, S., HUISINGA, K. L., HUSEBØ, M. E. P., KÖKE, A. J., PASKINS, Z. &

- PITSILLIDOU, I. A. 2018. EULAR recommendations for the health professional's approach to pain management in inflammatory arthritis and osteoarthritis. *Annals of the rheumatic diseases*, 77, 797-807.
- GIACOMONI, P. U., MAMMONE, T. & TERI, M. 2009. Gender-linked differences in human skin. *Journal of Dermatological Science*, 55, 144-149.
- GILROY, D. W., COLVILLE-NASH, P., WILLIS, D., CHIVERS, J., PAUL-CLARK, M. & WILLOUGHBY, D. 1999. Inducible cyclooxygenase may have anti-inflammatory properties. *Nature medicine*, 5, 698-701.
- GLASS, G. V. 1976. Primary, Secondary, and Meta-Analysis of Research 1. *Educational Researcher*, 5, 3-8.
- GLYMOUR, M. M., WEUVE, J., BERKMAN, L. F., KAWACHI, I. & ROBINS, J. M. 2005. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *American journal of epidemiology*, 162, 267-278.
- GRACE, D., ROGERS, J., SKEITH, K. & ANDERSON, K. 1999. Topical diclofenac versus placebo: A double blind, randomized clinical trial in patients with osteoarthritis of the knee. *Journal of Rheumatology*, 26, 2659-2663.
- GREENSPAN, J. D., CRAFT, R. M., LERESCHE, L., ARENDT-NIELSEN, L., BERKLEY, K. J., FILLINGIM, R. B., GOLD, M. S., HOLDCROFT, A., LAUTENBACHER, S. & MAYER, E. A. 2007. Studying sex and gender differences in pain and analgesia: a consensus report. *Pain*, 132, S26-S45.
- GUERMAZI, A., ROEMER, F. W., FELSON, D. T. & BRANDT, K. D. 2013. Unresolved Questions in Rheumatology: Motion for Debate: Osteoarthritis Clinical Trials Have Not Identified Efficacious Therapies Because Traditional Imaging Outcome Measures Are Inadequate. *Arthritis & Rheumatism*, 65, 2748-2758.
- GUSTORFF, B., POOLE, C., KLOIMSTEIN, H., HACKER, N. & LIKAR, R. 2013. Treatment of neuropathic pain with the capsaicin 8% patch: Quantitative sensory testing (QST) in a prospective observational study identifies potential predictors of response to capsaicin 8% patch treatment. *Scandinavian journal of pain*, 4, 138-145.
- GUYATT, G., OXMAN, A. D., AKL, E. A., KUNZ, R., VIST, G., BROZEK, J., NORRIS, S., FALCK-YTTER, Y., GLASZIOU, P. & JAESCHKE, R. 2011a. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64, 383-394.
- GUYATT, G., SACKETT, D., TAYLOR, D. W., GHONG, J., ROBERTS, R. & PUGSLEY, S. 1986. Determining Optimal Therapy —

Randomized Trials in Individual Patients. *New England Journal of Medicine*, 314, 889-892.

GUYATT, G. H., OXMAN, A. D., KUNZ, R., BROZEK, J., ALONSO-COELLO, P., RIND, D., DEVEREAUX, P., MONTORI, V. M., FREYSCHUSS, B. & VIST, G. 2011b. GRADE guidelines 6. Rating the quality of evidence—imprecision. *Journal of Clinical Epidemiology*, 64, 1283-1293.

GUYATT, G. H., OXMAN, A. D., KUNZ, R., WOODCOCK, J., BROZEK, J., HELFAND, M., ALONSO-COELLO, P., FALCK-YTTER, Y., JAESCHKE, R. & VIST, G. 2011c. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *Journal of Clinical Epidemiology*, 64, 1303-1310.

GUYATT, G. H., OXMAN, A. D., KUNZ, R., WOODCOCK, J., BROZEK, J., HELFAND, M., ALONSO-COELLO, P., GLASZIOU, P., JAESCHKE, R. & AKL, E. A. 2011d. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *Journal of Clinical Epidemiology*, 64, 1294-1302.

GUYATT, G. H., OXMAN, A. D., MONTORI, V., VIST, G., KUNZ, R., BROZEK, J., ALONSO-COELLO, P., DJULBEGOVIC, B., ATKINS, D. & FALCK-YTTER, Y. 2011e. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *Journal of Clinical Epidemiology*, 64, 1277-1282.

GUYATT, G. H., OXMAN, A. D., SULTAN, S., GLASZIOU, P., AKL, E. A., ALONSO-COELLO, P., ATKINS, D., KUNZ, R., BROZEK, J. & MONTORI, V. 2011f. GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, 64, 1311-1316.

GUYATT, G. H., OXMAN, A. D., VIST, G., KUNZ, R., BROZEK, J., ALONSO-COELLO, P., MONTORI, V., AKL, E. A., DJULBEGOVIC, B. & FALCK-YTTER, Y. 2011g. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *Journal of Clinical Epidemiology*, 64, 407-415.

GUYATT, G. H., THORLUND, K., OXMAN, A. D., WALTER, S. D., PATRICK, D., FURUKAWA, T. A., JOHNSTON, B. C., KARANICOLAS, P., AKL, E. A. & VIST, G. 2013. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles—continuous outcomes. *Journal of Clinical Epidemiology*, 66, 173-183.

HAANPÄÄ, M. & TREEDE, R.-D. 2012. Capsaicin for neuropathic pain: linking traditional medicine and molecular biology. *European neurology*, 68, 264-275.

HAGEN, M. & BAKER, M. 2017. Skin penetration and tissue permeation after topical administration of diclofenac. *Current medical research and opinion*, 33, 1623-1634.

- HAROUTIUNIAN, S., DRENNAN, D. A. & LIPMAN, A. G. 2010. Topical NSAID Therapy for Musculoskeletal Pain. *Pain Medicine*, 11, 535-549.
- HARTLING, L., OSPINA, M., LIANG, Y., DRYDEN, D. M., HOOTON, N., SEIDA, J. K. & KLASSEN, T. P. 2009. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *Bmj*, 339, b4012.
- HASLER-NGUYEN, N. & FOTOPOULOS, G. 2012. Effect of rubbing on the in vitro skin permeation of diclofenac-diethylamine 1.16% gel. *BMC research notes*, 5, 321.
- HASSAN H, W. D. 2014. Central pain processing in osteoarthritis: implications for treatment. *Pain Management*, 4.
- HÄUSER, W., HANSEN, E. & ENCK, P. 2012. Nocebo phenomena in medicine: their relevance in everyday clinical practice. *Deutsches Ärzteblatt International*, 109, 459.
- HAWKER, G., STEWART, L., FRENCH, M., CIBERE, J., JORDAN, J., MARCH, L., SUAREZ-ALMAZOR, M. & GOOBERMAN-HILL, R. 2008a. Understanding the pain experience in hip and knee osteoarthritis—an OARSI/OMERACT initiative. *Osteoarthritis and Cartilage*, 16, 415-422.
- HAWKER, G. A., DAVIS, A. M., FRENCH, M. R., CIBERE, J., JORDAN, J. M., MARCH, L., SUAREZ-ALMAZOR, M., KATZ, J. N. & DIEPPE, P. 2008b. Development and preliminary psychometric testing of a new OA pain measure – an OARSI/OMERACT initiative. *Osteoarthritis and Cartilage*, 16, 409-414.
- HAWKER, G. A., MIAN, S., KENDZERSKA, T. & FRENCH, M. 2011. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care & Research*, 63, S240-S252.
- HAYASHI, D., ROEMER, F. W. & GUERMAZI, A. 2016. Imaging for osteoarthritis. *Annals of Physical and Rehabilitation Medicine*, 59, 161-169.
- HEDGES, L. V. 1982. Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational and Behavioral Statistics*, 7, 119-137.
- HEDGES, L. V., GUREVITCH, J. & CURTIS, P. S. 1999. The meta - analysis of response ratios in experimental ecology. *Ecology*, 80, 1150-1156.

- HIGGINS JPT, G. S. E. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1 [updated March 2011]. *The Cochrane Collaboration*.
- HIGGINS, J. P. T. & WELTON, N. J. 2015. Network meta-analysis: a norm for comparative effectiveness? *The Lancet*, 386, 628-630.
- HINGORANI, A. D., WINDT, D. A. V. D., RILEY, R. D., ABRAMS, K., MOONS, K. G. M., STEYERBERG, E. W., SCHROTER, S., SAUERBREI, W., ALTMAN, D. G. & HEMINGWAY, H. 2013. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ : British Medical Journal*, 346.
- HMAMOUCI, I., ALLALI, F., TAHIRI, L., KHAZZANI, H., EL MANSOURI, L., ALLA, S. A. O., ABOUQAL, R. & HAJJAJ-HASSOUNI, N. 2012. Clinically important improvement in the WOMAC and predictor factors for response to non-specific non-steroidal anti-inflammatory drugs in osteoarthritic patients: a prospective study. *BMC research notes*, 5, 58.
- HOCHBERG, M. C., ALTMAN, R. D., APRIL, K. T., BENKHALTI, M., GUYATT, G., MCGOWAN, J., TOWHEED, T., WELCH, V., WELLS, G. & TUGWELL, P. 2012. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care & Research*, 64, 465-474.
- HOCHMAN, J., DAVIS, A., ELKAYAM, J., GAGLIESE, L. & HAWKER, G. 2013. Neuropathic pain symptoms on the modified painDETECT correlate with signs of central sensitization in knee osteoarthritis. *Osteoarthritis and Cartilage*, 21, 1236-1242.
- HOCHMAN, J., GAGLIESE, L., DAVIS, A. & HAWKER, G. 2011. Neuropathic pain symptoms in a community knee OA cohort. *Osteoarthritis and Cartilage*, 19, 647-654.
- HOGBEN, L. & SIM, M. 1953. The self-controlled and self-recorded clinical trial for low-grade morbidity. *British Journal of Preventive & Social Medicine*, 7, 163.
- HONG, H., CARLIN, B. P., SHAMLIYAN, T. A., WYMAN, J. F., RAMAKRISHNAN, R., SAINFORT, F. & KANE, R. L. 2013. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Medical Decision Making*, 33, 702-714.
- HOZO, S. P., DJULBEGOVIC, B. & HOZO, I. 2005. Estimating the mean and variance from the median, range, and the size of a sample. *BMC medical research methodology*, 5, 13.
- HUA, H., BURKE, D. L., CROWTHER, M. J., ENSOR, J., TUDUR SMITH, C. & RILEY, R. D. 2017. One - stage individual participant data meta - analysis models: estimation of treatment - covariate interactions

must avoid ecological bias by separating out within - trial and across - trial information. *Statistics in medicine*, 36, 772-789.

HUNTER, D. J., GUERMAZI, A., ROEMER, F., ZHANG, Y. & NEOGI, T. 2013. Structural correlates of pain in joints with osteoarthritis. *Osteoarthritis and Cartilage*, 21, 1170-1178.

HUNTER, D. J., SCHOFIELD, D. & CALLANDER, E. 2014. The individual and socioeconomic impact of osteoarthritis. *Nature Reviews Rheumatology*, 10, 437-441.

HUTCHINGS, A., CALLOWAY, M., CHOY, E., HOOPER, M., HUNTER, D. J., JORDAN, J. M., ZHANG, Y., BASER, O., LONG, S. & PALMER, L. 2007. The Longitudinal Examination of Arthritis Pain (LEAP) study: relationships between weekly fluctuations in patient-rated joint pain and other health outcomes. *Journal of Rheumatology*, 34, 2291-2300.

HUTTON, B., SALANTI, G., CALDWELL, D. M., CHAIMANI, A., SCHMID, C. H., CAMERON, C., IOANNIDIS, J. P., STRAUS, S., THORLUND, K. & JANSEN, J. P. 2015. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine*, 162, 777-784.

IASP. 2017. *IASP Taxonomy: Pain terms* [Online]. Washington: International Association for the Study of Pain. Available: <http://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698&navItemNumber=576> [Accessed 11/01/2019].

JABBARI, M., HASHEMPUR, M., RAZAVI, S., SHAHRAKI, H., KAMALINEJAD, M. & EMTIAZY, M. 2016. Efficacy and short-term safety of topical Dwarf Elder (*Sambucus ebulus* L.) versus diclofenac for knee osteoarthritis: a randomized, double-blind, active-controlled trial. *Journal of Ethnopharmacology*, 188, 80-86.

JANSEN, J. P., CRAWFORD, B., BERGMAN, G. & STAM, W. 2008. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. *Value in Health*, 11, 956-964.

JIANG, L., RONG, J., WANG, Y., HU, F., BAO, C., LI, X. & ZHAO, Y. 2011. The relationship between body mass index and hip osteoarthritis: a systematic review and meta-analysis. *Joint Bone Spine*, 78, 150-155.

JIANG, L., XIE, X., WANG, Y., WANG, Y., LU, Y., TIAN, T., CHU, M. & SHEN, Y. 2016. Body mass index and hand osteoarthritis susceptibility: an updated meta - analysis. *International Journal of Rheumatic Diseases*, 19, 1244-1254.

JIN, X., BEGUERIE, J. R., ZHANG, W., BLIZZARD, L., OTAHAL, P., JONES, G. & DING, C. 2015. Circulating C reactive protein in osteoarthritis:

a systematic review and meta-analysis. *Annals of the Rheumatic Diseases*, 74, 703-710.

JORDAN, K., ARDEN, N., DOHERTY, M., BANNWARTH, B., BIJLSMA, J., DIEPPE, P., GUNTHER, K., HAUSELMANN, H., HERRERO-BEAUMONT, G. & KAKLAMANIS, P. 2003. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Annals of the Rheumatic Diseases*, 62, 1145-1155.

JUHL, C., LUND, H., ROOS, E. M., ZHANG, W. & CHRISTENSEN, R. 2012. A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. *Arthritis*, 2012, 136245.

JÜNI, P., REICHENBACH, S. & DIEPPE, P. 2006. Osteoarthritis: rational approach to treating the individual. *Best Practice & Research Clinical Rheumatology*, 20, 721-740.

KAGEYAMA, T. 1984. [Pharmaceutical Evaluation of KPG-200 (Ketoprofen Ointment) on Arthrosis Deformans: Double Blind Controlled Study Compared with KPG-200 Base]. *Yakuri to Chiryo (Japanese Pharmacology and Therapeutics)*, 12, 2297-317.

KAGEYAMA, T. 1987. A double blind placebo controlled multicenter study of piroxicam 0.5% gel in osteoarthritis of the knee. *European Journal of Rheumatology and Inflammation*, 8, 114-115.

KATZ, N. P., MOU, J., PAILLARD, F. C., TURNBULL, B., TRUDEAU, J. & STOKER, M. 2015a. Predictors of Response in Patients With Postherpetic Neuralgia and HIV-Associated Neuropathy Treated With the 8% Capsaicin Patch (Qutenza). *Clinical Journal of Pain*, 31, 859-866.

KATZ, N. P., PAILLARD, F. C. & EKMAN, E. 2015b. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *Journal of Orthopaedic Surgery and Research*, 10, 24.

KEERIE, C., TUCK, C., MILNE, G., ELDRIDGE, S., WRIGHT, N. & LEWIS, S. C. 2018. Data sharing in clinical trials—practical guidance on anonymising trial datasets. *Trials*, 19, 25.

KELLGREN, J. & LAWRENCE, J. 1957. Radiological assessment of osteoarthritis. *Annals of the Rheumatic Diseases*, 16, 494.

KIRKWOOD, B. & STERNE, J. 2003. *Essential Medical Statistics*, Massachusetts, Blackwell Science.

- KLOPPENBURG, M., KROON, F. P., BLANCO, F. J., DOHERTY, M., DZIEDZIC, K. S., GREIBROKK, E., HAUGEN, I. K., HERRERO-BEAMONT, G., JONSSON, H. & KJEKEN, I. 2018. 2018 update of the EULAR recommendations for the management of hand osteoarthritis. *Annals of the Rheumatic Diseases*, annrheumdis-2018-213826.
- KLOPPENBURG, M., MAHEU, E., KRAUS, V. B., CICUTTINI, F., DOHERTY, M., DREISER, R. L., HENROTIN, Y., JIANG, G. L., MANDL, L., MARTEL-PELLETIER, J., NELSON, A. E., NEOGI, T., PELLETIER, J. P., PUNZI, L., RAMONDA, R., SIMON, L. S. & WANG, S. 2015. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials for hand osteoarthritis. *Osteoarthritis and Cartilage*, 23, 772-786.
- KNEER, W., ROTHER, M., MAZGAREANU, S., SEIDEL, E. J. & GROUP, E. I.-S. 2013. A 12-week randomized study of topical therapy with three dosages of ketoprofen in Transfersome® gel (IDEA-033) compared with the ketoprofen-free vehicle (TDT 064), in patients with osteoarthritis of the knee. *Journal of pain research*, 6, 743.
- KOKI, A., KHAN, N. K., WOERNER, B. M., DANNENBERG, A., OLSON, L., SEIBERT, K., EDWARDS, D., HARDY, M., ISAKSON, P. & MASFERRER, J. L. 2002. Cyclooxygenase-2 in human pathological disease. *Eicosanoids and Other Bioactive Lipids in Cancer, Inflammation, and Radiation Injury*, 5. Springer.
- KONTOPANTELOS, E. & REEVES, D. 2013. A short guide and a forest plot command (ipdforest) for one-stage meta-analysis. *Stata Journal*, 13, 574-587.
- KOOPMAN, L., VAN DER HEIJDEN, G. J., HOES, A. W., GROBBEE, D. E. & ROVERS, M. M. 2008. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *International Journal of Technology Assessment in Health Care*, 24, 358-361.
- KOSUWON, W., SIRICHATIWAPEE, W., WISANUYOTIN, T., JEERAVIPOOLVARN, P. & LAUPATTARAKASEM, W. 2010. Efficacy of symptomatic control of knee osteoarthritis with 0.0125% of capsaicin versus placebo. *Journal of the Medical Association of Thailand*, 93, 1188-95.
- KRAUS, V. B., BLANCO, F. J., ENGLUND, M., KARSDAL, M. A. & LOHMANDER, L. S. 2015. Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use. *Osteoarthritis and Cartilage*, 23, 1233-1241.
- KRAVITZ, R. & DUAN, N. 2014. Design and Implementation of N-of-1 Trials: a user's guide. *Agency for healthcare research and quality, US Department of Health and Human Services*.

- KULKANTRAKORN, K., LORSUWANSIRI, C. & MEESAWATSOM, P. 2012. 0.025% capsaicin gel for the treatment of painful diabetic neuropathy: a randomized, double blind, cross-over, placebo controlled trial. *Pain Practice*, 13, 497-503.
- KURIEN, T., ARENDT-NIELSEN, L., PETERSEN, K. K., GRAVEN-NIELSEN, T. & SCAMMELL, B. E. 2018. Preoperative Neuropathic Pain-like Symptoms and Central Pain Mechanisms in Knee Osteoarthritis Predicts Poor Outcome 6 Months After Total Knee Replacement Surgery. *The Journal of Pain*, 19, 1329-1341.
- LANE, N. E., HOCHBERG, M. C., NEVITT, M. C., SIMON, L. S., NELSON, A. E., DOHERTY, M., HENROTIN, Y. & FLECHSENHAR, K. 2015. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials for hip osteoarthritis. *Osteoarthritis and Cartilage*, 23, 761-771.
- LANE, S. P. & HENNES, E. P. 2018. Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships*, 35, 7-31.
- LARSSON, A., PALSTAM, A., BJERSING, J., LÖFGREN, M., ERNBERG, M., KOSEK, E., GERDLE, B. & MANNERKORPI, K. 2018. Controlled, cross-sectional, multi-center study of physical capacity and associated factors in women with fibromyalgia. *BMC musculoskeletal disorders*, 19, 121.
- LASLETT, L. L. & JONES, G. 2014. Capsaicin for osteoarthritis pain. *Capsaicin as a Therapeutic Molecule*. Springer.
- LEGHA, A., RILEY, R. D., ENSOR, J., SNELL, K. I., MORRIS, T. P. & BURKE, D. L. 2018. Individual participant data meta - analysis of continuous outcomes: A comparison of approaches for specifying and estimating one - stage models. *Statistics in Medicine*, 37, 4404-4420.
- LEQUESNE, M., MERY, C., SAMSON, M. & GERARD, P. 1987. Indexes of severity for osteoarthritis of the hip and knee: validation-value in comparison with other assessment tests. *Scandinavian Journal of Rheumatology*, 16, 85-89.
- LEWIS, G. N., RICE, D. A. & MCNAIR, P. J. 2012. Conditioned pain modulation in populations with chronic pain: a systematic review and meta-analysis. *The Journal of Pain*, 13, 936-944.
- LEYLAND, K., GATES, L., SANCHEZ-SANTOS, M., ALHAMBRA, D. P., JUDGE, A., COLLINS, G., CLEVELAND, R., FELSON, D., JORDON, J. & CALLAHAN, L. 2017. Knee osteoarthritis and the risk of premature mortality in the community: an international individual patient level meta-analysis in 9889 subjects. *Osteoarthritis and Cartilage*, 25, S29-S30.

- LI, T., PUHAN, M. A., VEDULA, S. S., SINGH, S. & DICKERSIN, K. 2011. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Medicine*, 9, 1-5.
- LILLIE, E. O., PATAY, B., DIAMANT, J., ISSELL, B., TOPOL, E. J. & SCHORK, N. J. 2011. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized Medicine*, 8, 161-173.
- LIN, F. Y., ZHANG, W. Y., JONES, A. & DOHERTY, M. 2004. Efficacy of topical non-steroidal anti-inflammatory drugs in the treatment of osteoarthritis: meta-analysis of randomised controlled trials. *BMJ*, 329, 324-326B.
- LOW, P. A., OPFER-GEHRKING, T. L., DYCK, P. J., LITCHY, W. J. & O'BRIEN, P. C. 1995. Double-blind, placebo-controlled study of the application of capsaicin cream in chronic distal painful polyneuropathy. *Pain*, 62, 163-8.
- LUMLEY, T. 2002. Network meta - analysis for indirect treatment comparisons. *Statistics in medicine*, 21, 2313-2324.
- LYDERSEN, S. 2015. Statistical review: frequently given comments. *Annals of the rheumatic diseases*, 74, 323-325.
- MAINKA, T., MALEWICZ, N. M., BARON, R., ENAX-KRUMOVA, E. K., TREEDE, R. D. & MAIER, C. 2016. Presence of hyperalgesia predicts analgesic efficacy of topically applied capsaicin 8% in patients with peripheral neuropathic pain. *European Journal of Pain*, 20, 116-129.
- MAKRIS, U. E., KOHLER, M. J. & FRAENKEL, L. 2010. Adverse effects of topical nonsteroidal antiinflammatory drugs in older adults with osteoarthritis: A systematic literature review. *Journal of Rheumatology*, 37, 1236-1243.
- MALFAIT A, S. T. 2013. Towards a mechanism-based approach to pain management in osteoarthritis. *Nature Reviews Rheumatology*, 9, 654-664.
- MALY, M. R. & KRUPA, T. 2007. Personal experience of living with knee osteoarthritis among older adults. *Disability and rehabilitation*, 29, 1423-1433.
- MAPP PI, W. D. 2012. Mechanisms and targets of angiogenesis and nerve growth in osteoarthritis. *Nature Reviews Rheumatology*, 390-398.
- MARCH, L., IRWIG, L., SCHWARZ, J., SIMPSON, J., CHOCK, C. & BROOKS, P. 1994. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ*, 309, 1041-1046.
- MARICAR, N., PARKES, M. J., CALLAGHAN, M. J., HUTCHINSON, C. E., GAIT, A. D., HODGSON, R., FELSON, D. T. & O'NEILL, T. W. 2017.

Structural predictors of response to intra-articular steroid injection in symptomatic knee osteoarthritis. *Arthritis Research & Therapy*, 19, 88.

MARTINI, C. H., YASSEN, A., KREBS-BROWN, A., PASSIER, P., STOKER, M., OLOFSEN, E. & DAHAN, A. 2013. A novel approach to identify responder subgroups and predictors of response to low- and high-dose capsaicin patches in postherpetic neuralgia. *European Journal of Pain*, 17, 1491-1501.

MARTINOLI, C. & BIANCHI, S. 2007. Knee. *Ultrasound of the Musculoskeletal System. Medical Radiology (Diagnostic Imaging)*. Berlin, Heidelberg: Springer.

MASON, L., MOORE, R. A., DERRY, S., EDWARDS, J. E. & MCQUAY, H. J. 2004. Systematic review of topical capsaicin for the treatment of chronic pain. *BMJ*, 328, 991.

MCALINDON, T., BANNURU, R., SULLIVAN, M., ARDEN, N., BERENBAUM, F., BIERMA-ZEINSTRAS, S., HAWKER, G., HENROTIN, Y., HUNTER, D. & KAWAGUCHI, H. 2014. OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis and Cartilage*, 22, 363-388.

MCALINDON, T. E., DRIBAN, J. B., HENROTIN, Y., HUNTER, D. J., JIANG, G. L., SKOU, S. T., WANG, S. & SCHNITZER, T. 2015. OARSI Clinical Trials Recommendations: Design, conduct, and reporting of clinical trials for knee osteoarthritis. *Osteoarthritis and Cartilage*, 23, 747-760.

MCALINDON, T. E., LAVALLEY, M. P., GULIN, J. P. & FELSON, D. T. 2000. Glucosamine and chondroitin for treatment of osteoarthritis: A systematic quality assessment and meta-analysis. *JAMA*, 283, 1469-1475.

MCCARTHY, G. M. & MCCARTY, D. J. 1992. Effect of topical capsaicin in the therapy of painful osteoarthritis of the hands. *Journal of Rheumatology*, 19, 604-7.

MCCLEANE, G. 2000. The analgesic efficacy of topical capsaicin is enhanced by glyceryl trinitrate in painful osteoarthritis: a randomized, double blind, placebo controlled study. *European Journal of Pain*, 4, 355-60.

MCHUGH, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica*, 22, 276-282.

MCWILLIAMS, D. F., LEEB, B. F., MUTHURI, S. G., DOHERTY, M. & ZHANG, W. 2011. Occupational risk factors for osteoarthritis of the knee: a meta-analysis. *Osteoarthritis and Cartilage*, 19, 829-839.

MELZACK, R. 1975. The McGill Pain Questionnaire: Major properties and scoring methods. *PAIN*, 1, 277-299.

- MILLER RE, T. P., OBEIDAT AM, RAGHU P, ISHIHARA S, MILLER RJ, MALFAIT A 2015. The role of peripheral nociceptive neurons in the pathophysiology of osteoarthritis pain. *Curr Osteoporos Rep*, 318-326.
- MILLS, E. J., THORLUND, K. & IOANNIDIS, J. P. A. 2013. Demystifying trial networks and network meta-analysis. *BMJ*, 346.
- MILNER, R. & DOHERTY, C. 2015. TRPV1 and Capsaicin. In: TUBBS, R., RIZK, E., SHOJA, M., LOUKAS, M., BARBARO, N. & SPINNER, R. (eds.) *Nerves and Nerve Injuries: Pain, Treatment, Injury, Disease and Future Directions*. London: Academic Press, Elsevier.
- MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G. & GROUP, P. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*, 6, e1000097.
- MOORE, R. A., TRAMER, M. R., CARROLL, D., WIFFEN, P. J. & MCQUAY, H. J. 1998. Quantitative systematic review of topically applied non-steroidal anti-inflammatory drugs. *BMJ* 316, 333-338.
- MORETON, B. J., TEW, V., DAS NAIR, R., WHEELER, M., WALSH, D. A. & LINCOLN, N. B. 2015. Pain Phenotype in Patients With Knee Osteoarthritis: Classification and Measurement Properties of painDETECT and Self - Report Leeds Assessment of Neuropathic Symptoms and Signs Scale in a Cross - Sectional Study. *Arthritis Care & Research*, 67, 519-528.
- MURRAY, C. J. L., RICHARDS, M. A., NEWTON, J. N., FENTON, K. A., ANDERSON, H. R., ATKINSON, C., BENNETT, D., BERNABÉ, E., BLENCOWE, H., BOURNE, R., BRAITHWAITE, T., BRAYNE, C., BRUCE, N. G., BRUGHA, T. S., BURNEY, P., DHERANI, M., DOLK, H., EDMOND, K., EZZATI, M., FLAXMAN, A. D., FLEMING, T. D., FREEDMAN, G., GUNNELL, D., HAY, R. J., HUTCHINGS, S. J., OHNO, S. L., LOZANO, R., LYONS, R. A., MARCENES, W., NAGHAVI, M., NEWTON, C. R., PEARCE, N., POPE, D., RUSHTON, L., SALOMON, J. A., SHIBUYA, K., VOS, T., WANG, H., WILLIAMS, H. C., WOOLF, A. D., LOPEZ, A. D. & DAVIS, A. 2013. UK health performance: findings of the Global Burden of Disease Study 2010. *The Lancet*, 381, 997-1020.
- NCCCC 2008. National Institute for Health and Clinical Excellence: Guidance. *Osteoarthritis: National Clinical Guideline for Care and Management in Adults*. London: National Collaborating Centre for Chronic Conditions, Royal College of Physicians (UK).
- NEOGI, T. 2017. Structural correlates of pain in osteoarthritis. *Clin Exp Rheumatol*, 35, S75-S78.
- NEOGI, T., FELSON, D., NIU, J., NEVITT, M., LEWIS, C. E., ALIABADI, P., SACK, B., TORNER, J., BRADLEY, L. & ZHANG, Y. 2009. Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. *BMJ*, 339.

- NEVITT, S. J., MARSON, A. G., DAVIE, B., REYNOLDS, S., WILLIAMS, L. & SMITH, C. T. 2017. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ*, 357, j1390.
- NICE. 2014. Osteoarthritis: Care and management in adults [CG177]. Available: <http://www.nice.org.uk/guidance/cg177/evidence/cg177-osteoarthritis-full-guideline3> [Accessed 16/10/2015].
- NICE. 2018. Neuropathic pain in adults: pharmacological management in non-specialist settings [CG173].
- NIETHARD, F. U., GOLD, M. S., SOLOMON, G. S., LIU, J. M., UNKAUF, M., ALBRECHT, H. H. & ELKIK, F. 2005. Efficacy of topical diclofenac diethylamine gel in osteoarthritis of the knee. *Journal of Rheumatology*, 32, 2384-2392.
- NÜESCH, E., TRELLE, S., REICHENBACH, S., RUTJES, A. W., TSCHANNEN, B., ALTMAN, D. G., EGGER, M. & JÜNI, P. 2010. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ*, 341, c3515.
- O'LEARY, H., SMART, K. M., MOLONEY, N. A. & DOODY, C. M. 2017. Nervous System Sensitization as a Predictor of Outcome in the Treatment of Peripheral Musculoskeletal Conditions: A Systematic Review. *Pain Practice*, 17, 249-266.
- O'NEILL, J., BROCK, C., OLESEN, A. E., ANDRESEN, T., NILSSON, M. & DICKENSON, A. H. 2012. Unravelling the Mystery of Capsaicin: A Tool to Understand and Treat Pain. *Pharmacological Reviews*, 64, 939-971.
- O'REILLY, S. & DOHERTY, M. 2003. Signs, symptoms, and laboratory tests. In: BRANDT, K., DOHERTY, M. & LOHMANDER, L. S. (eds.) *Osteoarthritis*. 2nd ed ed. Oxford: Oxford University Press.
- OHMANN, C., BANZI, R., CANHAM, S., BATTAGLIA, S., MATEI, M., ARIYO, C., BECNEL, L., BIERER, B., BOWERS, S. & CLIVIO, L. 2017. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ open*, 7, e018647.
- OLIVO, S. A., MACEDO, L. G., GADOTTI, I. C., FUENTES, J., STANTON, T. & MAGEE, D. J. 2008. Scales to assess the quality of randomized controlled trials: a systematic review. *Physical therapy*, 88, 156-175.
- OTTILLINGER, B., GÖMÖR, B., MICHEL, B. A., PAVELKA, K., BECK, W. & ELSASSER, U. 2001. Efficacy and safety of eltenac gel in the treatment of knee osteoarthritis. *Osteoarthritis and Cartilage*, 9, 273-280.

- PAICE, J. A., FERRANS, C. E., LASHLEY, F. R., SHOTT, S., VIZGIRDA, V. & PITRAK, D. 2000. Topical capsaicin in the management of HIV-associated peripheral neuropathy. *Journal of Pain and Symptom Management*, 19, 45-52.
- PAN, F., TIAN, J., AITKEN, D., CICUTTINI, F., WINZENBERG, T. & JONES, G. 2017. The association of knee structural pathology with pain at the knee is modified by pain at other sites in those with knee osteoarthritis. *Clinical Rheumatology*, 36, 2549-2555.
- PATRONO, C. 2015. Nonsteroidal antiinflammatory drugs. In: HOCHBERG, M., SILMAN, A. J., SMOLEN, J., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed. Philadelphia: Elsevier Mosby.
- PEDRO. 1999. *PEDro scale* [Online]. Centre of Evidence-Based Physiotherapy. Available: <http://www.pedro.org.au/english/downloads/pedro-scale/> [Accessed 08/02/2016].
- PEIKERT, A., HENTRICH, M. & OCHS, G. 1991. TOPICAL 0.025-PERCENT CAPSAICIN IN CHRONIC POSTHERPETIC NEURALGIA - EFFICACY, PREDICTORS OF RESPONSE AND LONG-TERM COURSE. *Journal of Neurology*, 238, 452-456.
- PEREIRA, D., PELETEIRO, B., ARAÚJO, J., BRANCO, J., SANTOS, R. A. & RAMOS, E. 2011. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. *Osteoarthritis and Cartilage*, 19, 1270-1285.
- PERSSON, M. S. M., SARMANOVA, A., DOHERTY, M. & ZHANG, W. 2018. Conventional and biologic disease-modifying anti-rheumatic drugs for osteoarthritis: a meta-analysis of randomized controlled trials. *Rheumatology*, key131-key131.
- PETERSEN, K. K., SIMONSEN, O., ANDERSEN, M. M., OLESEN, A. E. & ARENDT-NIELSEN, L. 2017. The effect of facilitated temporal summation of pain, widespread pressure hyperalgesia and pain intensity in patients with knee osteoarthritis on the responds to Non-Steroidal Anti-Inflammatory Drugs-A preliminary analysis. *Scandinavian journal of pain*, 16, 182-182.
- PHAM, T., VAN DER HEIJDE, D., ALTMAN, R. D., ANDERSON, J. J., BELLAMY, N., HOCHBERG, M., SIMON, L., STRAND, V., WOODWORTH, T. & DOUGADOS, M. 2004. OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis and Cartilage*, 12, 389-399.
- POLANIN, J. R. & TERZIAN, M. 2019. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *Journal of Clinical Epidemiology*, 106, 60-69.

- POOLE, J. L. 2011. Measures of hand function: Arthritis Hand Function Test (AHFT), Australian Canadian Osteoarthritis Hand Index (AUSCAN), Cochin Hand Function Scale, Functional Index for Hand Osteoarthritis (FIHOA), Grip Ability Test (GAT), Jebsen Hand Function Test (JHFT), and Michigan Hand Outcomes Questionnaire (MHQ). *Arthritis Care & Research*, 63, S189-S199.
- PUHAN, M. A., SCHÜNEMANN, H. J., MURAD, M. H., LI, T., BRIGNARDELLO-PETERSEN, R., SINGH, J. A., KESSELS, A. G. & GUYATT, G. H. 2014. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*, 349, g5630.
- PUIG-JUNOY, J. & ZAMORA, A. R. 2015. Socio-economic costs of osteoarthritis: A systematic review of cost-of-illness studies. *Seminars in Arthritis and Rheumatism*, 44, 531-541.
- RAO, P. & KNAUS, E. E. 2008. Evolution of nonsteroidal anti-inflammatory drugs (NSAIDs): cyclooxygenase (COX) inhibition and beyond. *Journal of Pharmacy & Pharmaceutical Sciences*, 11, 81-110.
- RAT, A.-C., POUCHOT, J., COSTE, J., BAUMANN, C., SPITZ, E., RETEL-RUDE, N., BAUMANN, M., LE QUINTREC, J.-S., DUMONT-FISCHER, D. & GUILLEMIN, F. 2006. Development and testing of a specific quality-of-life questionnaire for knee and hip osteoarthritis: OAKHQOL (OsteoArthritis of Knee Hip Quality Of Life). *Joint Bone Spine*, 73, 697-704.
- RICARD, E., SAULE, E., BARNETCHE, T., TREVES, R., VERGNE-SALLE, P. & BERTIN, P. 2013. Efficacité de la capsaïcine dans le traitement des douleurs neuropathiques: méta-analyse des essais randomisés contrôlés. *Douleurs: Evaluation-Diagnostic-Traitement*, 14, 286-291.
- RILEY RD, L. P., ABO-ZAID G 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, 340.
- ROBERTS, E., DELGADO NUNES, V., BUCKNER, S., LATCHER, S., CONSTANTIN, M., MILLER, P., DOHERTY, M., ZHANG, W., BIRRELL, F., PORCHERET, M., DZIEDZIC, K., BERNSTEIN, I., WISE, E. & CONAGHAN, P. G. 2016. Paracetamol: not as safe as we thought? A systematic literature review of observational studies. *Ann Rheum Dis*, 75, 552-9.
- ROBINSON, W. H., LEPUS, C. M., WANG, Q., RAGHU, H., MAO, R., LINDSTROM, T. M. & SOKOLOVE, J. 2016. Low-grade inflammation as a key mediator of the pathogenesis of osteoarthritis. *Nature Reviews Rheumatology*, 12, 580-592.
- ROOS, E. M. & LOHMANDER, L. S. 2003. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health and Quality of Life Outcomes*, 1, 64.

- ROOS, E. M., ROOS, H. P., LOHMANDER, L. S., EKDAHL, C. & BEYNNON, B. D. 1998. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. *J Orthop Sports Phys Ther*, 28, 88-96.
- ROSE, W., MANZ, G. & LEMMEL, E. M. 1991. Behandlung der aktivierten gonarthrose mit topisch appliziertem piroxicam-gel. [Topical application of Piroxicam-gel in the treatment of activated gonarthrosis]. *Münch Med Wschr.*, 133, 562-566.
- ROTH, S. H. & SHAINHOUSE, J. Z. 2004. Efficacy and safety of a topical diclofenac solution (Pennsaid) in the treatment of primary osteoarthritis of the knee - A randomized, double-blind, vehicle-controlled clinical trial. *Archives of Internal Medicine*, 164, 2017-2023.
- ROTHER, M. & CONAGHAN, P. G. 2013. A Randomized, Double-blind, Phase III Trial in Moderate Osteoarthritis Knee Pain Comparing Topical Ketoprofen Gel with Ketoprofen-free Gel. *Journal of Rheumatology*, 40, 1742-1748.
- ROTHER, M., LAVINS, B. J., KNEER, W., LEHNHARDT, K., SEIDEL, E. J. & MAZGAREANU, S. 2007. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral cefecoxib and placebo in osteoarthritis of the knee: Multicentre randomised controlled trial. *Annals of the Rheumatic Diseases*, 66, 1178-1183.
- ROUZER, C. A. & MARNETT, L. J. 2009. Cyclooxygenases: structural and functional insights. *Journal of lipid research*, 50, S29-S34.
- ROVENSKÝ, J., MICEKOVÁ, D., GUBZOVÁ, Z., FIMMERS, R., LENHARD, G., VÖGTLE-JUNKERT, U. & SCHREYGER, F. 2001. Treatment of knee osteoarthritis with a topical non-steroidal antiinflammatory drug. Results of a randomized, double-blind, placebo-controlled study on the efficacy and safety of a 5% ibuprofen cream. *Drugs under experimental and clinical research*, 27, 209-21.
- RUNHAAR, J., ROZENDAAL, R. M., VAN MIDDELKOOP, M., BIJLSMA, H. J. W., DOHERTY, M., DZIEDZIC, K. S., LOHMANDER, L. S., MCALINDON, T., ZHANG, W. & BIERMA ZEINSTRAS, S. 2017. Subgroup analyses of the effectiveness of oral glucosamine for knee and hip osteoarthritis: a systematic review and individual patient data meta-analysis from the OA trial bank. *Ann Rheum Dis*, 76, 1862-1869.
- SALAFFI, F., STANCATI, A., SILVESTRI, C. A., CIAPETTI, A. & GRASSI, W. 2004. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *European Journal of Pain*, 8, 283-291.
- SANDELIN, J., HARILAINEN, A., CRONE, H., HAMBERG, P., FORSSKAHL, B. & TAMELANDER, G. 1997. Local NSAID gel (eltenac) in the treatment of osteoarthritis of the knee - A double blind study

comparing eltenac with oral diclofenac and placebo gel.
Scandinavian Journal of Rheumatology, 26, 287-292.

SARMANOVA, A. 2017. *Synovial changes detected by ultrasound in community-derived people with knee pain*. Doctor of Philosophy, University of Nottingham.

SARMANOVA, A., HALL, M., FERNANDES, G. S., BHATTACHARYA, A., VALDES, A. M., WALSH, D. A., DOHERTY, M. & ZHANG, W. 2017. Association between ultrasound-detected synovitis and knee pain: a population-based case-control study with both cross-sectional and follow-up data. *Arthritis Research & Therapy*, 19, 281.

SCHAIBLE, H.-G. 2015. Scientific Basis of Pain. In: HOCHBERG, M. C., SILMAN, A. J., SMOLEN, J., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatology*. 6th ed ed. Philadelphia: Elsevier Mosby.

SCHEDLOWSKI, M., ENCK, P., RIEF, W. & BINGEL, U. 2015. Neuro-bio-behavioral mechanisms of placebo and nocebo responses: implications for clinical trials and clinical practice. *Pharmacological reviews*, 67, 697-730.

SCHEFFLER, N. M., SHEITEL, P. L. & LIPTON, M. N. 1991. Treatment of painful diabetic neuropathy with capsaicin 0.075%. *Journal of the American Podiatric Medical Association*, 81, 288-93.

SCHNITZER, T., MORTON, C. & COKER, S. 1994. Topical capsaicin therapy for osteoarthritis pain: Achieving a maintenance regimen. *Seminars in Arthritis and Rheumatism*, 23, 34-40.

SCHNITZER, T. J., PELLETIER, J. P., HASELWOOD, D. M., ELLISON, W. T., ERVIN, J. E., GORDON, R. D., LISSE, J. R., ARCHAMBAULT, W. T., SAMPSON, A. R., FEZATTE, H. B., PHILLIPS, S. B. & BERNSTEIN, J. E. 2012. Civamide cream 0.075% in patients with osteoarthritis of the knee: A 12-week randomized controlled clinical trial with a longterm extension. *Journal of Rheumatology*, 39, 610-620.

SCHORK, N. J. 2015. Personalized medicine: Time for one-person trials. *Nature*, 520, 609-611.

SCHUIT, E., LI, A. H. & IOANNIDIS, J. 2018. How often can meta-analyses of individual-level data individualize treatment? A meta-epidemiologic study. *International Journal of Epidemiology*, dyd239.

SCHULZ, K. F., CHALMERS, I., HAYES, R. J. & ALTMAN, D. G. 1995. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 273, 408-412.

SERRANO, A., TORRES, D., VECIANA, M., CARO, C., MONTERO, J. & MAYORAL, V. 2017. Quantitative Thermal Testing Profiles as a Predictor of Treatment Response to Topical Capsaicin in Patients

with Localized Neuropathic Pain. *Pain research and treatment*, 2017.

- SHOARA, R., HASHEMPUR, M. H., ASHRAF, A., SALEHI, A., DEHSHAHRI, S. & HABIBAGAH, Z. 2015. Efficacy and safety of topical *Matricaria chamomilla* L.(chamomile) oil for knee osteoarthritis: a randomized controlled clinical trial. *Complementary Therapies in Clinical Practice*, 21, 181-187.
- SILVERWOOD, V., BLAGOJEVIC-BUCKNALL, M., JINKS, C., JORDAN, J., PROTHEROE, J. & JORDAN, K. 2015. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis and Cartilage*, 23, 507-515.
- SIMON, L. S., GRIERSON, L. M., NASEER, Z., BOOKMAN, A. A. & ZEVSCHAINHOUSE, J. 2009. Efficacy and safety of topical diclofenac containing dimethyl sulfoxide (DMSO) compared with those of topical placebo, DMSO vehicle and oral diclofenac for knee osteoarthritis. *Pain*, 143, 238-45.
- SMITH, T. C., SPIEGELHALTER, D. J. & THOMAS, A. 1995. Bayesian approaches to random - effects meta - analysis: a comparative study. *Statistics in Medicine*, 14, 2685-2699.
- SRIKANTH, V. K., FRYER, J. L., ZHAI, G., WINZENBERG, T. M., HOSMER, D. & JONES, G. 2005. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis and Cartilage*, 13, 769-781.
- STAMM, T., VAN DER GIESEN, F., THORSTENSSON, C., STEEN, E., BIRRELL, F., BAUERNFEIND, B., MARSHALL, N., PRODINGER, B., MACHOLD, K. & SMOLEN, J. 2009. Patient perspective of hand osteoarthritis in relation to concepts covered by instruments measuring functioning: a qualitative European multicentre study. *Annals of the Rheumatic Diseases*, 68, 1453-1460.
- STENGEL, D., BAUWENS, K., SEHOULI, J., RADEMACHER, G., MUTZE, S., EKKERNKAMP, A. & PORZSOLT, F. 2005. Emergency ultrasound - based algorithms for diagnosing blunt abdominal trauma. *Cochrane Database of Systematic Reviews*, 14, CD004446.
- STEWART, L. A., CLARKE, M., ROVERS, M., RILEY, R. D., SIMMONDS, M., STEWART, G. & TIERNEY, J. F. 2015. Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *JAMA*, 313, 1657-1665.
- STEWART LA, T. J. 2002. To IPD or not to IPD? Advantages and Disadvantages of Systematic Reviews Using Individual Patient Data. *Evaluation & the Health Professions*, 25, 76-97.
- SUN, X., BRIEL, M., WALTER, S. D. & GUYATT, G. H. 2010. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*, 340, c117.

- SUTTON, A. J. & ABRAMS, K. R. 2001. Bayesian methods in meta-analysis and evidence synthesis. *Statistical methods in medical research*, 10, 277-303.
- SZALLASI, A. & BLUMBERG, P. M. 1999. Vanilloid (capsaicin) receptors and mechanisms. *Pharmacological Reviews*, 51, 159-212.
- SZEBENYI, B., HOLLANDER, A. P., DIEPPE, P., QUILTY, B., DUDDY, J., CLARKE, S. & KIRWAN, J. R. 2006. Associations between pain, function, and radiographic features in osteoarthritis of the knee. *Arthritis & Rheumatism*, 54, 230-235.
- SZOLCSANYI, J. 2014. Capsaicin and sensory neurones: a historical perspective. *Progress in Drug Research*, 68, 1-37.
- TAICHMAN, D. B., BACKUS, J., BAETHGE, C., BAUCHNER, H., DE LEEUW, P. W., DRAZEN, J. M., FLETCHER, J., FRIZELLE, F. A., GROVES, T., HAILEAMLAK, A., JAMES, A., LAINE, C., PEIPERL, L., PINBORG, A., SAHNI, P. & WU, S. 2016. Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, 374, 384-386.
- TANAMAS, S., HANNA, F. S., CICUTTINI, F. M., WLUKA, A. E., BERRY, P. & URQUHART, D. M. 2009. Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review. *Arthritis Care & Research*, 61, 459-467.
- TANDAN, R., LEWIS, G. A., KRUSINSKI, P. B., BADGER, G. B. & FRIES, T. J. 1992. Topical capsaicin in painful diabetic neuropathy. Controlled study with long-term follow-up. *Diabetes Care*, 15, 8-14.
- TEIXEIRA, M. J., MENEZES, L. M. B., SILVA, V., GALHARDONI, R., SASSON, J., OKADA, M., DUARTE, K. P., YENG, L. T. & DE ANDRADE, D. C. 2015. Liposomal topical capsaicin in post-herpetic neuralgia: a safety pilot study. *Arquivos De Neuro-Psiquiatria*, 73, 237-240.
- THAKUR, M., DICKENSON, A. H. & BARON, R. 2014. Osteoarthritis pain: nociceptive or neuropathic? *Nature Reviews Rheumatology*, 10, 374.
- TIERNEY, J. F., VALE, C., RILEY, R., SMITH, C. T., STEWART, L., CLARKE, M. & ROVERS, M. 2015. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med*, 12, e1001855.
- TOWHEED, T. E. 2006. PENNSAID therapy for osteoarthritis of the knee: a systematic review and metaanalysis of randomized controlled trials. *Journal of Rheumatology*, 33, 567-573.
- TRAMÈR, M. R., REYNOLDS, D. J. M., MOORE, R. A. & MCQUAY, H. J. 1997. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ*, 315, 635-640.

- TRAN, G., SMITH, T. O., GRICE, A., KINGSBURY, S. R., MCCRORY, P. & CONAGHAN, P. G. 2016. Does sports participation (including level of performance and previous injury) increase risk of osteoarthritis? A systematic review and meta-analysis. *British Journal of Sports Medicine*, 50, 1459-1466.
- TREASURY, H. 2013. Spending Review 2010 Executive Summary. In: TREASURY, H. (ed.).
- TRNAVSKY, K., FISCHER, M., VOGTLE-JUNKERT, U. & SCHREYGER, F. 2004. Efficacy and safety of 5% ibuprofen cream treatment in knee osteoarthritis. Results of a randomized, double-blind, placebo-controlled study. *Journal of Rheumatology*, 31, 565-572.
- TRUSHEIM, M. R., BERNDT, E. R. & DOUGLAS, F. L. 2007. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nature Reviews Drug Discovery*, 6, 287-293.
- UNDERWOOD, M., ASHBY, D., CROSS, P., HENNESSY, E., LETLEY, L., MARTIN, J., MT-ISA, S., PARSONS, S., VICKERS, M. & WHYTE, K. 2008. Advice to use topical or oral ibuprofen for chronic knee pain in older people: randomised controlled trial and patient preference study. *BMJ*, 336, 138-142.
- VAN HASELEN, R. A. & FISHER, P. A. 2000. A randomized controlled trial comparing topical piroxicam gel with a homeopathic gel in osteoarthritis of the knee. *Rheumatology*, 39, 714-9.
- VAN MIDDELKOOP, M., ARDEN, N., ATCHIA, I., BIRRELL, F., CHAO, J., REZENDE, M., LAMBERT, R., RAVAUD, P., BIJLSMA, J. & DOHERTY, M. 2016. The OA Trial Bank: meta-analysis of individual patient data from knee and hip osteoarthritis trials show that patients with severe pain exhibit greater benefit from intra-articular glucocorticoids. *Osteoarthritis and Cartilage*, 24, 1143-1152.
- VAN MIDDELKOOP, M., DZIEDZIC, K. S., DOHERTY, M., ZHANG, W., BIJLSMA, J. W., MCALINDON, T. E., LOHMANDER, S. L. & BIERMA-ZEINSTRAS, S. M. 2013. Individual patient data meta-analysis of trials investigating the effectiveness of intra-articular glucocorticoid injections in patients with knee or hip osteoarthritis: an OA Trial Bank protocol for a systematic review. *Systematic reviews*, 2, 1-8.
- VANDERWEELE, T. J. & KNOL, M. J. 2011. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Annals of Internal Medicine*, 154, 680-683.
- VARADI, G., ZHU, Z., BLATTNER, T., HOSLE, M., LOHER, A., POKORNY, R., FREY, D. & CARTER, S. G. 2013. Randomized clinical trial evaluating transdermal ibuprofen for moderate to severe knee osteoarthritis. *Pain Physician*, 16, 749-762.

- VERONESE, N., CEREDA, E., MAGGI, S., LUCHINI, C., SOLMI, M., SMITH, T., DENKINGER, M., HURLEY, M., THOMPSON, T. & MANZATO, E. 2016. Osteoarthritis and mortality: a prospective cohort study and systematic review with meta-analysis. *Seminars in arthritis and rheumatism*, 46, 160-167.
- VISSER, A. W., BOYESEN, P., HAUGEN, I. K., SCHOONES, J. W., VAN DER HEIJDE, D. M., ROSENDAAL, F. R. & KLOPPENBURG, M. 2015. Instruments Measuring Pain, Physical Function, or Patient's Global Assessment in Hand Osteoarthritis: A Systematic Literature Search. *Journal of Rheumatology*, 42, 2118-34.
- VOHRA, S., SHAMSEER, L., SAMPSON, M., BUKUTU, C., SCHMID, C. H., TATE, R., NIKLES, J., ZUCKER, D. R., KRAVITZ, R. & GUYATT, G. 2015. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *Journal of Clinical Epidemiology*, 350, h1738.
- WAKEFIELD, R. J., BALINT, P. V., SZKUDLAREK, M., FILIPPUCCI, E., BACKHAUS, M., D'AGOSTINO, M. A., SANCHEZ, E. N., IAGNOCCO, A., SCHMIDT, W. A., BRUYN, G. A., KANE, D., O'CONNOR, P. J., MANGER, B., JOSHUA, F., KOSKI, J., GRASSI, W., LASSERE, M. N., SWEN, N., KAINBERGER, F., KLAUSER, A., OSTERGAARD, M., BROWN, A. K., MACHOLD, K. P. & CONAGHAN, P. G. 2005. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *Journal of Rheumatology*, 32, 2485-7.
- WALKER, J. S., SHEATHER - REID, R. B., CARMODY, J. J., DAY, R. O. & VIAL, J. H. 1997. Nonsteroidal antiinflammatory drugs in rheumatoid arthritis and osteoarthritis. Support for the concept of "responders" and "nonresponders". *Arthritis & Rheumatism*, 40, 1944-1954.
- WAN, X., WANG, W., LIU, J. & TONG, T. 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, 14, 135.
- WANDEL, S., JÜNI, P., TENDAL, B., NÜESCH, E., VILLIGER, P. M., WELTON, N. J., REICHENBACH, S. & TRELLE, S. 2010. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. *BMJ*, 341, c4675.
- WARNER, S. C. & VALDES, A. M. 2016. The genetics of osteoarthritis: A review. *Journal of Functional Morphology and Kinesiology*, 1, 140-153.
- WEGMAN, A. C. M., VAN DER WINDT, D. A. W. M., DE HAAN, M., DEVILLÉ, W. L. J. M., FO, C. T. C. A. & DE VRIES, T. P. G. M. 2003. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. *Annals of the Rheumatic Diseases*, 62, 1156-1161.

- WEITZ, E., KLEIBOER, A., VAN STRATEN, A., HOLLON, S. D. & CUIJPERS, P. 2017. Individual patient data meta-analysis of combined treatments versus psychotherapy (with or without pill placebo), pharmacotherapy or pill placebo for adult depression: a protocol. *BMJ open*, 7, e013478.
- WHITE, I. 2015. Network meta-analysis. *Stata Journal*, 951-981.
- WHITE, I. R., BARRETT, J. K., JACKSON, D. & HIGGINS, J. P. 2012. Consistency and inconsistency in network meta - analysis: model estimation using multivariate meta - regression. *Research synthesis methods*, 3, 111-125.
- WIDRIG, R., SUTER, A., SALLER, R. & MELZER, J. 2007. Choosing between NSAID and arnica for topical treatment of hand osteoarthritis in a randomised, double-blind study. *Rheumatology International*, 27, 585-591.
- WILKINSON, C. E., CARR, A. J. & DOHERTY, M. 2005. Does increasing the grades of the knee osteoarthritis line drawing atlas alter its clinimetric properties? *Annals of the Rheumatic Diseases*, 64, 1467-1473.
- WOLFE, F., CLAUW, D. J., FITZCHARLES, M.-A., GOLDENBERG, D. L., HÄUSER, W., KATZ, R. L., MEASE, P. J., RUSSELL, A. S., RUSSELL, I. J. & WALITT, B. 2016. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Seminars in Arthritis and Rheumatism*, 46, 319-329.
- WOOLF, C. J. 2011. Central sensitization: implications for the diagnosis and treatment of pain. *Pain*, 152, S2-S15.
- WU, X. M., BENNETT, D. H., RITZ, B., CASSADY, D. L., LEE, K. & HERTZ-PICCIOTTO, I. 2010. Usage pattern of personal care products in California households. *Food Chem Toxicol*, 48, 3109-19.
- WYATT, L. & DOHERTY, M. 2016. Morphological aspects of pathology. In: DOHERTY, M., BIJLSMA, J., ARDEN, N., HUNTER, D. & DALBETH, N. (eds.) *Osteoarthritis and Crystal Arthropathy*. 3rd ed. ed. Oxford: Oxford University Press.
- XING, D., XU, Y., LIU, Q., KE, Y., WANG, B., LI, Z. & LIN, J. 2016. Osteoarthritis and all-cause mortality in worldwide populations: grading the evidence from a meta-analysis. *Scientific reports*, 6, 24393.
- YAMAMOTO, S., KODAMA, T., HASEGAWA, Y. & SAEKI, K. 1979. [Clinical evaluation of indomethacin ointment on osteoarthritis--evaluation by a multicenter double-blind trial (author's transl)]. *Ryumachi. [Rheumatism]*, 19, 328-336.
- YELLAND, M. J., NIKLES, C. J., MCNAIRN, N., DEL MAR, C. B., SCHLUTER, P. J. & BROWN, R. M. 2007. Celecoxib compared with sustained-

release paracetamol for osteoarthritis: a series of n-of-1 trials. *Rheumatology*, 46, 135-140.

- YOO, B., CHOI, S. W., LEE, M. S. & MOON, H. B. 1996. Treatment of Knee Osteoarthritis with Ketoprofen (Ketotop): A Double-blind Placebo-controlled Randomized Trial. *The Journal of the Korean Rheumatism Association*, 3, 70-75.
- ZENG, C., WEI, J., PERSSON, M. S. M., SARMANOVA, A., DOHERTY, M., XIE, D., WANG, Y., LI, X., LI, J., LONG, H., LEI, G. & ZHANG, W. 2018. Relative efficacy and safety of topical non-steroidal anti-inflammatory drugs for osteoarthritis: a systematic review and network meta-analysis of randomised controlled trials and observational studies. *British Journal of Sports Medicine*, 52, 642-650.
- ZHANG, W. & DOHERTY, M. 2018. Efficacy paradox and proportional contextual effect (PCE). *Clinical Immunology*, 186, 82-86.
- ZHANG, W., DOHERTY, M., ARDEN, N., BANNWARTH, B., BIJLSMA, J., GUNTHER, K.-P., HAUSELMANN, H. J., HERRERO-BEAUMONT, G., JORDAN, K., KAKLAMANIS, P., LEEB, B., LEQUESNE, M., LOHMANDER, S., MAZIERES, B., MARTIN-MOLA, E., PAVELKA, K., PENDLETON, A., PUNZI, L., SWOBODA, B., VARATOJO, R., VERBRUGGEN, G., ZIMMERMANN-GORSKA, I. & DOUGADOS, M. 2005. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Annals of the Rheumatic Diseases*, 64, 669-681.
- ZHANG, W., DOHERTY, M., LEEB, B., ALEKSEEVA, L., ARDEN, N., BIJLSMA, J., DINÇER, F., DZIEDZIC, K., HAUSELMANN, H. & HERRERO-BEAUMONT, G. 2007. EULAR evidence based recommendations for the management of hand osteoarthritis: report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Annals of the Rheumatic Diseases*, 66, 377-388.
- ZHANG, W., DOHERTY, M., LEEB, B. F., ALEKSEEVA, L., ARDEN, N. K., BIJLSMA, J. W., DINCER, F., DZIEDZIC, K., HAUSELMANN, H. J., KAKLAMANIS, P., KLOPPENBURG, M., LOHMANDER, L. S., MAHEU, E., MARTIN-MOLA, E., PAVELKA, K., PUNZI, L., REITER, S., SMOLEN, J., VERBRUGGEN, G., WATT, I. & ZIMMERMANN-GORSKA, I. 2009. EULAR evidence-based recommendations for the diagnosis of hand osteoarthritis: report of a task force of ESCISIT. *Annals of the Rheumatic Diseases*, 68, 8-17.
- ZHANG, W., DOHERTY, M., PEAT, G., BIERMA-ZEINSTRAS, M. A., ARDEN, N. K., BRESNIHAN, B., HERRERO-BEAUMONT, G., KIRSCHNER, S., LEEB, B. F., LOHMANDER, L. S., MAZIÈRES, B., PAVELKA, K., PUNZI, L., SO, A. K., TUNCER, T., WATT, I. & BIJLSMA, J. W. 2010a. EULAR evidence-based recommendations for the

diagnosis of knee osteoarthritis. *Annals of the Rheumatic Diseases*, 69, 483-489.

- ZHANG, W. & LI WAN PO, A. 1994. The effectiveness of topically applied capsaicin. *European Journal of Clinical Pharmacology*, 46, 517-522.
- ZHANG, W., MOSKOWITZ, R. W., NUKI, G., ABRAMSON, S., ALTMAN, R. D., ARDEN, N., BIERMA-ZEINSTRAS, S., BRANDT, K. D., CROFT, P., DOHERTY, M., DOUGADOS, M., HOCHBERG, M., HUNTER, D. J., KWOH, K., LOHMANDER, L. S. & TUGWELL, P. 2008a. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis and Cartilage*, 16, 137-162.
- ZHANG, W., NUKI, G., MOSKOWITZ, R., ABRAMSON, S., ALTMAN, R., ARDEN, N., BIERMA-ZEINSTRAS, S., BRANDT, K., CROFT, P. & DOHERTY, M. 2010b. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis and Cartilage*, 18, 476-499.
- ZHANG, W., ROBERTSON, J., JONES, A., DIEPPE, P. & DOHERTY, M. 2008b. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. *Annals of the rheumatic diseases*, 67, 1716-1723.
- ZIGMOND, A. S. & SNAITH, R. P. 1983. The hospital anxiety and depression scale. *Acta psychiatrica scandinavica*, 67, 361-370.
- ZOU, K., WONG, J., ABDULLAH, N., CHEN, X., SMITH, T., DOHERTY, M. & ZHANG, W. 2016. Examination of overall treatment effect and the proportion attributable to contextual effect in osteoarthritis: meta-analysis of randomised controlled trials. *Annals of the Rheumatic Diseases*, 75, 1964-1970.
- ZUCKER, D. R., RUTHAZER, R. & SCHMID, C. H. 2010. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *Journal of Clinical Epidemiology*, 63, 1312-1323.

APPENDICES

Appendix 1 – Systematic literature search strategies

OA and NSAID search strategy: Medline (via Ovid)

1. Randomized Controlled Trials as Topic/
 2. (randomi?ed controlled trial or double-blind or blind\$ or mask\$ or clinical trial or trial).af.
 3. Administration, Topical/
 4. (stick-on or cutaneous or dermal or transcutaneous or percutaneous or skin or massage or embrocation or gel or ointment or aerosol or cream or lotion or mousse or foam or liniment or spray or rub or balm or salve or emulsion or oil or patch or plaster).af.
 5. exp Osteoarthritis/dt, th [Drug Therapy, Therapy]
 6. (osteoarthr* or OA or arthritis or Degenerative arthritis or degenerative joint disease or arthrosis or osteoarthrosis).af.
 7. Anti-Inflammatory Agents, Non-Steroidal/ad, ae, tu [Administration & Dosage, Adverse Effects, Therapeutic Use]
 8. (nsaid or nonsteroidal antiinflammatory or non-steroidal anti-inflammatory or bufexamac or bufexine or calmaderm or ekzemase or dicoflenac or solaraze or pennsaid or voltarol or emulgen or voltarene or optha or voltaren or etofenamate or afrolate or algesalona or bayro or deiron or etofen or flexium or flogoprofen or rheuma-gel or rheumon or traumalix or traumon or zenavan or felbinac or dolinac or flexfree or napageln or target or traxam or fentiazac or domureuma or fentiazaco or norvedan or riscalon or fepradinol or dalgen or flexidol or cocresol or rangozona or reuflodol or pinazone or zepelin or flufenamic or dignodolin or rheuma or lindofluid or sastridex or lunoxaprofen or priaxim or flubiprofen or fenomel or ocufen or ocuflur or "trans act lat" or tulip or ibuprofen or cuprofen or "deep relief" or fenbid or ibu-cream or ibugel or ibuleve or ibumousse or ibuspray or "nurofen gel" or proflex or motrin or advil or radian or ralgesic or ibutop or indomethacin or indocin or indospray or isonixin or nixyn or ketoprofen or tiloket or oruvail or powergel or solpaflex or ketorolac or acular or trometamol or meclofenamic or naproxen or naprosyn or niflumic or actol or flunir or niflactol topico or niflugel or nifluril or oxyphenbutazone or californit or diflamil or otone or tanderil or piketoprofen or calmatel or triparsean or piroxicam or feldene or pranoprofen or oftalar or pranox or suxibuzone or danilon or flamilon or ufenamate or fenazol).af.
 9. 1 or 2
 10. 3 or 4
 11. 5 or 6
 12. 7 or 8
 13. 9 and 10 and 11 and 12
 14. Limit 13 to randomized controlled trial
-

OA and capsaicin search strategy: Medline (via Ovid)

1. Randomized Controlled Trials as Topic/
 2. (randomi?ed controlled trial or double-blind or blind\$ or mask\$ or clinical trial or trial).af.
 3. Administration, Topical/
 4. (stick-on or cutaneous or dermal or transcutaneous or percutaneous or skin or massage or embrocation or gel or ointment or aerosol or cream or lotion or mousse or foam or liniment or spray or rub or balm or salve or emulsion or oil or patch or plaster).af.
 5. exp Osteoarthritis/dt, th [Drug Therapy, Therapy]
 6. (osteoarthr* or OA or arthritis or Degenerative arthritis or degenerative joint disease or arthrosis or osteoarthrosis).af.
 7. Capsaicin/ad, ae, tu [Administration & Dosage, Adverse Effects, Therapeutic Use]
 8. (8-Methyl-N-Vanillyl-6-Nonenamide or 8 methyl n vanillyl 6 nonenamide or axsain or capsaicin or capsaicine or capsicum farmaya or capsidol or capsin or capzasin or gelcen or katum or ngx 4010 or ngx-4010 or ngx4010 or zacin or zostrix or capsici or capsig or capsina or capsiplast or capzasin-p or dolorac or no pain-hp or priltam or r-gel or chilli or chili).af
 9. 1 or 2
 10. 3 or 4
 11. 5 or 6
 12. 7 or 8
 13. 9 and 10 and 11 and 12
 14. Limit 13 to randomized controlled trial
-

NP in Medline (via Ovid) – this replaced points 5 and 6 from the OA strategy

1. Neuralgia/
 2. Diabetic Neuropathies/
 3. Neuralgia, Postherpetic/
 4. Trigeminal Neuralgia/
 5. Facial Neuralgia/
 6. HIV Infections/ or Pain/ or Acquired Immunodeficiency Syndrome/
 7. Pain, Postoperative/
 8. Phantom Limb/
 9. Polyneuropathies/
 10. Nerve Compression Syndromes/
 11. Radiculopathy/
 12. Sciatica/
 13. Spinal Cord Injuries/
 14. Stroke/
 15. Multiple Sclerosis/
 16. Complex Regional Pain Syndromes/ or Causalgia/
 17. Polyarteritis Nodosa/
 18. (((((((((((((neuropath* adj3 pain).ab. or neurogen*.mp.) adj3 pain.ab.) or diabet*.mp.) adj3 neuropath*.ab.) or postherp*.mp.) adj3 neuropath*.ab.) or trigemin*.mp.) adj3 neuralg*.ab.) or central*.mp.) adj3 pain*.ab.) or (fac* adj3 (pain* or neuralg*))).ab. or (HIV adj3 (pain* or neuralg* or neuropath*))).ab. or (pain* adj3 (post-treatment* or post treatment* or posttreatment* or surg* or post-op* or postop* or post op*))).ab. or ((phantom adj3 limb) or ((stump or (postamputation or post-amputation or post amputation)) adj3 pain*))).ab. or (pain* adj3 (polyneuropath* or alcohol* or amyloid* or plasmocyt* or hypothyroid* or heredit* or fabry* or neuroborrel* or vitamin B or toxic*))).ab. or (((entrap* or compress* or periph*) adj3 (neuropath* or nerv*)) or cruralg* or brachalg* or radiculopath* or radicular pain or carpal tunnel or interosseous or pronator teres or struther* or cubital tunnel or guyon* canal or radial).ab. or spinal cord.mp.) adj3 injur*.ab.) or stroke.mp.) adj3 pain*.ab.) or ((complex adj3 region* adj3 pain*) or crps or sympathetic dystroph* or sudec* atroph* or (hand* adj3 shoulder adj3 syndrom*))).ab. or ((posttraum* or post traum* or post-traum*) adj3 (neuralg* or neuropath*))).ab. or isch?em*.mp.) adj3 neuropath*.ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
 19. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
-

Appendix 2 – Data extraction form and instructions

Data extraction form

General

Trial ID: Enter text **Author:** Enter text **Year:** Enter text **Source:** Enter text

Country: Enter text **Source of funding:** ☐Public ☐Commercial ☐Unclear

Participants

Setting: ☐Hospital ☐Community ☐Unclear

Joints affected: ☐Hand ☐Knee ☐Hip ☐Other: Please state

Diagnosis made using:

☐ Radiographic changes	☐ American College of Rheumatology criteria
☐ Pain	☐ Clinically
☐ Physical function	☐ Patient Global Assessment
☐ Stiffness	☐ Other: Please specify

Patient characteristics:

	Treatment	Control	Other (if applicable)	Total
Number of patients	Enter text	Enter text	Enter text	Enter text
Number of withdrawals	Enter text	Enter text	Enter text	Enter text
Withdrawals from lack of effect	Enter text	Enter text	Enter text	Enter text
Age (mean $\pm$ SD)	Enter text	Enter text	Enter text	Enter text
Female n (%)	Enter text	Enter text	Enter text	Enter text
Mean BMI	Enter text	Enter text	Enter text	Enter text

Study

Design: ☐Parallel ☐Crossover

Blinding (as stated in paper): ☐Un-blinded ☐Single-blinded ☐Double-blinded ☒Unclear

	Treatment	Control	Other (if applicable)
Drug	Enter text.	Enter text.	Enter text.
Formulation	Enter text.	Enter text.	Enter text.
Dose/concentration	Enter text.	Enter text.	Enter text.
Frequency	Enter text.	Enter text.	Enter text.

Duration: Please select all that apply

	1	2	3	4	5	6	7	8	9	10	11	12	Other
Duration of intervention (weeks)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Please specify.
Timing of follow-up (weeks)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Please specify.

Quality					
<i>Please see instructions criteria</i>					
	Yes	No	Unclear	Comments	
1. Was the randomisation procedure adequate?	☐	☐	☐		
2. Was the treatment allocation adequately concealed?	☐	☐	☐		
3. Were participants blinded to the intervention?	☐	☐	☐		
4. Were physicians blinded to the intervention?	☐	☐	☐		
5. Were outcome assessors blinded to the intervention?	☐	☐	☐		
6. Incomplete outcome data: Is the attrition rate <15%?	☐	☐	☐		
7. Are all pre-specified outcomes of interest reported in the pre-specified way?	☐	☐	☐		
8. Was intention-to-treat analysis used?	☐	☐	☐		
9. Were the treatment and control group similar at baseline?	☐	☐	☐		

Outcome												
Primary and secondary outcomes. Outcomes at all durations of follow-up given in paper. Data is not to be extracted from graphs. Please copy/paste tables as appropriate to cover all groups, durations, and outcomes. If possible, enter intention-to-treat (ITT) and per protocol (pp) population values.												

Group: Enter text. Duration: Choose an item. ITT or PP: Choose an item.												
Outcome	Definition	Baseline			Endpoint			Change				
		N	Mean	SD	N	Mean	SD	N	Mean	SD	95%	SE
Primary	Text	n	mean	SD	N	mean	SD	N	Mean	SD	CI	SE
Secondary	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
1												
2	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
3	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
4	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
5	Text	N	mean	SD	N	mean	SD	N	mean	SD	CI	SE

Group: Enter text. Duration: Choose an item. ITT or PP: Choose an item.												
Outcome	Definition	Baseline			Endpoint			Change				
		N	Mean	SD	N	Mean	SD	N	Mean	SD	95%	SE
Primary	Text	n	mean	SD	N	mean	SD	N	Mean	SD	CI	SE
Secondary	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
1												
2	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
3	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
4	Text	N	Mean	SD	N	mean	SD	N	Mean	SD	CI	SE
5	Text	N	mean	SD	N	mean	SD	N	mean	SD	CI	SE

Instructions for data extraction

General

Trial ID: entered by MP **Author:** entered by MP **Year:** entered by MP **Source:** entered by MP

Country: location of trial **Source of funding:** **Public** (e.g. University, state, charity funding), **Commercial** (e.g. pharmaceutical funding), **Unclear** (source of funding not clearly stated)

Participants

Setting: **Hospital** (inpatient population), **Community** (e.g. outpatients, clinic, general practice, "centre"), **Unclear** (source of participants not clearly stated)

Joints affected: ☐Hand ☐Knee ☐Hip ☐Other: Please state

Diagnosis made using: ☐Radiographic changes ☐American College of Rheumatology criteria

Criteria used to determine eligibility

☐Pain

☐Clinically

☐Physical function

☐Patient Global Assessment

☐Stiffness

☐Other: Please specify

Patient characteristics:

	Description
Number of patients	Total number of patients initially randomised into group
Number of withdrawals	Total number of withdrawals from group
Withdrawals from lack of effect	Number of withdrawals due to lack of effect or adverse event + lack of effect
Age (mean $\pm$ SD)	Mean age $\pm$ standard deviation of group
Female n (%)	Number of females in group (percent of group that is female). Note that if the number of males is given, please convert this to females (total n of participants – n male participants = n female participants)
Mean BMI	Mean body mass index of group. If number of participants in each BMI range is given instead of mean, please write the ranges and numbers of participants (or percentage)

Study

Design: **Parallel** (participants remain in one group throughout the whole study,) **Crossover** (participants switch groups during study. E.g. group 1 does treatment A then B, whilst group 2 does treatment B then A)

Blinding (as stated in paper): ☐Un-blinded ☐Single-blinded ☐Double-blinded ☐Unclear

Description

Drug	Name of active component of treatment, "placebo", or "vehicle"
Formulation	e.g. gel, solution, cream, tablet
Dose/concentration	Dose of active component
Frequency	Number of times/day that treatment is taken

Duration: Please select all that apply

Description

Duration of intervention (weeks)	Number of weeks treatment is prescribed for participant
Timing of follow-up (weeks)	The timing, in weeks after baseline, when outcome measures are assessed

Quality

Note that if the method description is incomplete, then papers conducted by the same study group on the same drug can be used to make the judgment (e.g. the drug and control may be described as identical in the first paper, but subsequent papers simply state double-blinded. If subsequent papers use the same drug and control, these are most likely identical, as previously described).

1. Was the randomisation procedure adequate?	YES – A random component in the sequence generation process is used, such as: • Using to a random number table • Using a computer random number generator • Tossing a coin • Shuffling cards or envelopes • Throwing dice • Drawing lots • Minimisation • Randomisation generated by outside statistician/central group • Random permuted block scheme	NO – a non-random component is used in the allocation process, such as: • Sequence generated using odd or even date of birth • Sequence based on a rule based on date of admission or hospital number • Allocation decision is made by clinician or participant • Allocation is based on clinical or laboratory test findings

Unclear – Study uses statements such as “we randomly allocated” or “using a randomised design”, but insufficient detail is given of the process to allow a decision to be made regarding the adequacy of the method. If the randomisation process is inadequately described this is deemed unclear.

2. Was the treatment allocation adequately concealed?	YES – participants and physicians could not foresee allocation assignment, because an adequate method was used to conceal allocation. Adequate allocation concealment methods include: • Central allocation • Identical drug containers are sequentially numbered • Opaque, sealed, sequentially numbered assignment envelopes are used • Remotely labelled study kit with no indication of treatment within • Statement that allocation was concealed from investigators	NO – participants or physicians could potentially foresee allocation. Inadequate allocation concealment procedures include: • The use of an open random allocation schedule • Assignment envelopes are not appropriate, such that envelopes may be unsealed, not sequentially numbered, or see-through • Rotation or alternation • Methods based on date of birth, case record number, or other patient identifiers

Unclear – No mention of precautions taken to conceal treatment allocation.

3. Were participants blinded to the intervention?	<i>YES – patients were adequately blinded and it is unlikely that the blinding could be broken. A statement describing that the treatments were identical or “matched” is expected to allow a decision.</i> <i>Unclear – Study reported as single- or double-blinded but no description of the process of blinding, there not allowing a decision to be made regarding the adequacy.</i>	<i>NO – No blinding, incomplete blinding, or blinding attempted but likely to have been broken</i>
4. Were physicians blinded to the intervention?	<i>YES – physicians were adequately blinded and it is unlikely that the blinding could be broken. If the treatments are identical a statement that the physicians were blinded allows a yes</i>	<i>NO – No blinding, incomplete blinding, or blinding attempted but likely to have been broken</i>
	<i>Unclear – Study reported as single- or double-blinded but no description of the process of blinding, there not allowing a decision to be made regarding the adequacy.</i>	
5. Were outcome assessors blinded to the intervention?	<i>YES – treatment choice is not evident when measuring outcome.</i> <i>• If outcome is patient-reported, then the answer for “Were participants blinded to the intervention?” will be the same for this question.</i> <i>• For physician-assessed outcomes, this is dependent on whether the treatment allocation can be identified from examination of the patient or their tests</i>	<i>NO – treatment choice is likely to be evident when measuring outcome</i> <i>• If outcome is patient-reported, then the answer for “Were participants blinded to the intervention?” will be the same for this question.</i> <i>• For physician-assessed outcomes, this is dependent on whether the treatment allocation can be identified from examination of the patient or their tests</i>
6. Incomplete outcome data: Is the attrition rate <15%?	<i>YES – drop-out rate is less than 15%</i>	<i>NO – drop-out rate is greater than 15%</i>
	<i>UNCLEAR – flow of patients not given. Unable to ascertain number of withdrawals</i>	
7. Are all pre-specified outcomes of interest reported in the pre-specified way?	<i>YES – all the study’s pre-specified outcomes of interest are reported in the pre-specified way</i>	<i>NO – Outcome measures were not pre-specified, such as:</i> <i>• Not all the pre-specified outcomes have been reported</i> <i>• Outcomes are not reported using the measurements or methods pre-specified</i> <i>• Outcomes reported had not been pre-specified, such as post-hoc analyses</i>

- Outcomes of interest have been incompletely reported
- The report does not report outcomes that would be expected from such a study

8. Was intention-to-treat analysis used?	<i>YES – all randomised participants are analysed according to the group they were allocated to, regardless of non-compliance, protocol violations, drop-outs etc.</i>	<i>NO – the study states that ITT was used, but patients have been excluded from analysis, post-randomisation, for variety of reasons. Alternatively, per-protocol analysis is used to analyse only participants that adhered fully to their allocated treatment, excluding any drop-outs or moves between treatment groups.</i>
9. Were the treatment and control group similar at baseline?	<i>UNCLEAR – does not mention if ITT or PP used and n for each outcome not presented</i> <i>YES – treatment and control group were similar regarding patient demographics and baseline parameters.</i>	<i>NO – Treatment and control groups were dissimilar with regards to patient demographics and baseline parameters. Alternatively, characteristics are not presented in either text or table</i>

UNCLEAR- if characteristics are mentioned in text, but data is not presented

Outcome

Primary and secondary outcomes. Please extract all outcomes at all durations of follow-up presented in the study. Data is not to be extracted from graphs, unless the exact outcome value is given for each point. Please copy/paste tables as appropriate to cover all groups, durations, and outcomes. If possible, enter intention-to-treat (ITT) and per protocol (pp) population values. Please use the labels ITT or PP as described by the author (i.e. if an author states they used ITT, please label the outcome as ITT, regardless of your answer to question 8 in the Risk of Bias assessment)

N = number of participants analysis is based on

SD = standard deviation

95% = 95% confidence interval

SE = standard error

Group: Treatment group **Duration:** timing of outcome measure **ITT or PP:** analysis is ITT or PP

Outcome	Definition	Baseline	Endpoint	Change
		N Mean SD	N Mean SD	N Mean SD 95% SE
Primary	Outcome measure, as described by study author	Baseline outcome measures	Outcome measure at "duration" stated above	Difference in outcome measure between baseline and endpoint value

Appendix 3 – GRADE assessment: Topical NSAIDs in OA

GRADE criteria	Rating	Footnotes	Quality of evidence
Study design	High	RCTs	⊕⊕⊕○ Moderate
Risk of Bias	No	Overall, low risks of bias evident for randomisation and blinding. Low/unclear risk of bias treatment allocation. Higher risk for ITT, however, many used a modified ITT which was unlikely to greatly bias the results. Higher risk studies were generally smaller and contributed less to overall ES. Treatment effect was unchanged in sensitivity analyses using high quality trials	
Inconsistency	Serious (-1)	There is some inconsistency. The ES varies from -0.17 to +1.06. Not all 95%CI overlap, especially comparing older to newer studies. I^2 indicates substantial heterogeneity	
Indirectness	No	PICO match MA aims. Time points included ranged from 1-12 weeks, however, all are still felt to reflect “short to medium term”	
Imprecision	No	Overall sample size is 6191, this is likely sufficient All of 95%CI is held within the same range of interpretation (statistically significant, but does not meet MCID)	
Publication bias	No	Small study effects evident on funnel plot, however, the conclusions are unchanged when using the trim and fill method. Trials not included in the MA were identified and included in the IPD MA. This did not substantially change the ES, therefore the subset of trials included are likely representative	

Appendix 4 – GRADE assessment: Capsaicin in OA

GRADE criteria	Rating	Footnotes	Quality of evidence
Study design	High	RCTs	⊕○○○ Very low
Risk of Bias	Serious (-1)	No trials blinded participants. As pain is subjective, this risk of bias may potentially alter results. High risks of bias were identified for other domains, such as a failure to conduct ITT analyses in all trials. In addition, randomisation and treatment allocation procedures were unclear across the body of evidence. It was not possible to conduct sensitivity analyses for high quality trials as none met the criteria.	
Inconsistency	Serious (-1)	ES vary from -0.18 to +0.78 in the trials. However, all confidence intervals do overlap. I ² indicates substantial heterogeneity, which was reduced (I ² =0) when excluding 0.0125% capsaicin. However, subgroup analyses are only plausible – comparisons are between studies, the hypothesis did not precede analysis, and it is not consistent across outcomes at other durations	
Indirectness	No	PICO match MA aims. One study used physician's assessment of pain, which may not be the most informative. However it is likely correlated with patient's pain assessment	
Imprecision	Serious (-1)	Sample size is 415, which is borderline using the "rule of thumb" Estimate is imprecise – at its lowest bound it indicates that capsaicin is worse than placebo. At its highest bound it indicates that capsaicin meets the MCID	
Publication bias	Serious (-1)	There is some asymmetry in the funnel plot. All evidence is from small, pharmaceutical-sponsored studies, therefore increasing the likelihood of publication bias	

Appendix 5 – GRADE assessment: NMA in OA

GRADE criteria	Rating	Footnotes	Quality of evidence
DIRECT EVIDENCE		None available	
INDIRECT EVIDENCE			⊕○○○ Very low
Study design	High	RCTs	
Risk of Bias	Serious (-1)	From direct estimate of capsaicin	
Inconsistency	Serious (-1)	From direct estimate of capsaicin	
Indirectness	Serious (-1)	From direct estimate of capsaicin	
Publication bias	Serious (-1)	From direct estimate of capsaicin	
Intransitivity	Serious (-1)	Capsaicin is likely not successfully blinded, whilst NSAIDs probably are. The placebo groups are therefore different and may modify the treatment effects	
NETWORK ESTIMATE			⊕○○○ Very low
Imprecision	Serious (-1)	Wide confidence interval with differing interpretations at each boundary. At either boundary, the treatment effect indicates small-to-moderate effects in favour of capsaicin or topical NSAID, respectively	
Incoherence	-	Not assessable	

Appendix 6 – GRADE assessment: Overall effect of topical NSAIDs in OA

GRADE criteria	Rating	Footnotes	Quality of evidence
Study design	Low	Observational	⊕○○○ Very low
Risk of Bias	No	Overall, low risks of bias for blinding. Higher risk for ITT, however, many used a modified ITT which was unlikely to greatly bias the results. Treatment effect was unchanged in sensitivity analyses using high quality trials	
Inconsistency	Serious (-1)	Inconsistency present. Estimates range from 0.62 to 2.12. Many confidence intervals do not overlap. $I^2 = 92.3\%$.	
Indirectness	No	PICO match MA aims. Time points included ranged from 1 - 12 weeks, however, all are still felt to reflect “short to medium term”	
Imprecision	No	Overall sample size is 3183, this is likely sufficient All of 95%CI is held within the same range of interpretation (meets MCID of 0.5 SD)	
Publication bias	No	Range of small and large studies available. Slight funnel plot asymmetry (figure not shown). Overall ES is largely unchanged when including previously unpublished placebo-controlled trials in IPD MA (overall ES 30mm (95%CI 25 to 35); data not shown.	

Appendix 7 – GRADE assessment: Overall effect of capsaicin in OA

GRADE criteria	Rating	Footnotes	Quality of evidence
Study design	Low	Observational	⊕○○○ Very low
Risk of Bias	Serious (-1)	No trials successfully blinded participants. High risks of bias were identified for other domains, such as a failure to conduct ITT analyses in all trials. No trials met criteria for “high quality” used for sensitivity analyses.	
Inconsistency	Serious (-1)	ES ranged for 0.18 to 1.62. Non-overlapping confidence intervals. $I^2=89\%$	
Indirectness	No	PICO match MA aims.	
Imprecision	No	Sample size is 206, below the “rule of thumb” threshold of 400. However, the interpretation of the estimate is unchanged at either bounds of the confidence interval, using the MCID threshold of 0.5 SD	
Publication bias	Serious (-1)	Only small, industry-funded trials are available. Evidence of funnel plot asymmetry (data not shown)	

Appendix 8 – Contact letter to authors



Dear [author],

We are currently conducting an individual patient data meta-analysis to compare “Relative efficacy of topical non-steroidal anti-inflammatory drugs and topical capsaicin between osteoarthritis and neuropathic pain”. We found that your study titled *[title]* fulfils our inclusion criteria and would appreciate your collaboration for this project that is funded by Arthritis Research UK. We feel that inclusion of data from your excellent study is essential to this project, therefore would like to invite you to be a collaborator and co-author for this project.

The project

Background

Current National Institute for Health and Care Excellence (NICE) guidelines recommend topical non-steroidal anti-inflammatory drugs (NSAIDs) for hand and knee osteoarthritis (OA)¹ and topical capsaicin for OA and chronic neuropathic conditions¹⁻². However, the relative efficacy of these treatments between these conditions, and whether there are patient characteristics that predict treatment outcome, remains unknown.

Aims

- To identify the efficacy of topical NSAIDs and capsaicin in OA, post-herpetic neuralgia (PHN), and painful diabetic neuropathy (PDN).
- To investigate whether treatment outcomes are influenced by patient-specific characteristics.

Brief method

- Conventional, network, and individual patient data (IPD) meta-analysis of randomised controlled trials.
- IPD data will be collated and subgroup analysis will be conducted.

Using IPD of several randomised controlled trials permits sufficient power to undertake subgroup analysis to identify treatment responders and predictors of response. Collaboration from the original investigators is essential in ensuring that the project gathers the information required for analysis.

Funding and host

This project is funded by the Arthritis Research UK Pain Centre and is hosted by the University of Nottingham. The project will be supervised by Professor Weiya Zhang, Professor David Walsh, and Professor Michael Doherty.

Collaborators

- All original trial PIs/corresponding authors
- OA Trial Bank
 - An international collaboration to develop an individual patient database of all randomised controlled trials in osteoarthritis (OA). The collaboration has undertaken an IPD meta-analysis to investigate the use of intra-articular steroid injections in OA and is now undertaking IPD meta-analyses for glucosamine, chondroitin and exercise therapy in OA. This current project also is part of the OA Trial Bank collaboration. More information is available at: www.oatrialbank.com
 - It is formalised with a chair, steering committee, and executive coordinator. The steering committee consists of international clinical researchers in the field of OA and all members are involved in the development of the protocol. Please see Figure 1 for further information.

Other projects

The data will also be used for the identification of placebo response and predictors of response in order to better understand the treatment effect in osteoarthritis.

Your participation

Your participation is essential to this project as it relies fully on the contribution of the data from the original trials. If you agree to participate:

- 1) You will be a collaborator of the project, involved in the database development, being a co-author or being included in the acknowledgements, whichever you prefer.
 - a) Data collected for your RCT will be transferred to the OA Trial Bank. These data should include background information collected about study participants (including demographic factors and disease factors), intervention, comparator, outcomes of the trial. Full details will be provided once the agreement is reached.
- 2) Both parties will sign a licence agreement. This will:
 - a) Cover input data, obligations, term and termination, authorship, and publication
 - b) Specify exactly what can and what cannot be published with your data

What next?

Please let us know if you are interested in participating in our project and on what terms you would wish to do so. Please let us also know if you have any unpublished trial data that may be eligible for inclusion to this project.

If you have any enquiries, please do not hesitate to contact us.

We look forward to hearing from you.

Kind regards,

Monica Persson, BMBS BMedSci

PhD Student

University of Nottingham
Academic Rheumatology
Clinical Sciences Building
City Hospital
Nottingham
NG5 1PB

Tel: +44 (0)115 823 1759

1. NICE. Osteoarthritis: Care and management in adults. *NICE Clinical Guideline CG 177* [Online], 2014.

<http://www.nice.org.uk/guidance/cg177/evidence/cg177-osteoarthritis-full-guideline3> (accessed 16/10/2015)

2. NICE. Neuropathic pain – pharmacological management: The pharmacological management of neuropathic pain in adults in non-specialist settings. *NICE Clinical Guideline CG 173* [Online], 2014.

www.nice.org.uk/guidance/CG173 (accessed 16/10/2015)

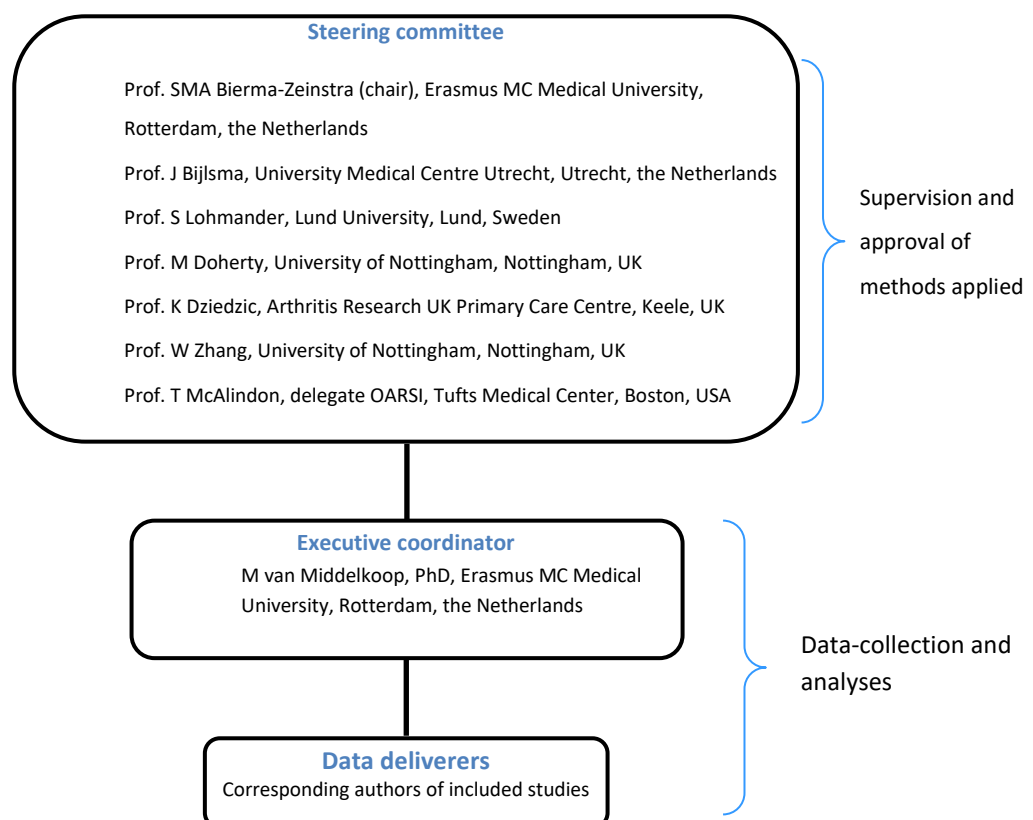


Figure 1. Organisation of OA Trial Bank

Appendix 9 – Contact email to institutions

Dear ,

I am a PhD student at the University of Nottingham; supervised by Professors **Weiya Zhang**, **David Walsh**, and **Michael Doherty**.

I am emailing regarding the study conducted at your university: **[title]**. Published in *[DETAILS]*. I have not been able to contact the author, and wanted to know if your university has kept the data and would be interested in collaborating on our project.

We will be conducting an **individual patient data (IPD) meta-analysis** of topical NSAIDs and topical capsaicin in osteoarthritis and neuropathic pain. This project is funded by Arthritis Research UK in collaboration with the OA Trial Bank. We hope that our project will lead to a greater understanding of the pain mechanisms in osteoarthritis and that it will help guide clinical decision-making by identifying patient subgroups that benefit most from the previously mentioned treatments. An IPD meta-analysis is dependent on receiving raw study data from all eligible studies, and as such, your collaboration would be essential to the project. We will invite all collaborators to be authors or be acknowledged on the project, as per your preference.

Please find further details on the collaboration attached to this email.

Please let me know if you have any information on the study and if so, whether you would be interested in collaborating on this project.

Best wishes,

Monica Persson, *BMBS BMedSci*

PhD Student

University of Nottingham
Academic Rheumatology
Clinical Sciences Building
City Hospital
Nottingham
NG5 1PB
Tel: +44 (0)115 823 1759

Appendix 10 – Contact letter to pharmaceutical companies



We are currently conducting an individual patient data meta-analysis to compare “Relative efficacy of topical non-steroidal anti-inflammatory drugs and topical capsaicin between osteoarthritis and neuropathic pain”. We are interested in randomised controlled trials investigating topical NSAIDs or topical capsaicin in osteoarthritis, painful diabetic neuropathy, or post-herpetic neuralgia. We are writing to enquire whether you have any unpublished data which may be eligible for inclusion to this project that is funded by Arthritis Research UK. Inclusion of unpublished data is essential to this project and, if any such data is available, we would like to invite you to be a collaborator and co-author for this project.

The project

Background

Current National Institute for Health and Care Excellence (NICE) guidelines recommend topical non-steroidal anti-inflammatory drugs (NSAIDs) for hand and knee osteoarthritis (OA) (NICE, 2014) and topical capsaicin for OA and chronic neuropathic conditions (NICE, 2014, NICE, 2018). However, the relative efficacy of these treatments between these conditions, and whether there are patient characteristics that predict treatment outcome, remains unknown.

Aims

- To identify the efficacy of topical NSAIDs and capsaicin in OA, post-herpetic neuralgia (PHN), and painful diabetic neuropathy (PDN).
- To investigate whether treatment outcomes are influenced by patient-specific characteristics.

Brief method

- Conventional, network, and individual patient data (IPD) meta-analysis of randomised controlled trials.
- IPD data will be collated and subgroup analysis will be conducted.

Using IPD of several randomised controlled trials permits sufficient power to undertake subgroup analysis to identify treatment responders and predictors of response. Collaboration from the original investigators is essential in ensuring that the project gathers the information required for analysis.

Funding and host

This project is funded by the Arthritis Research UK Pain Centre and is hosted by the University of Nottingham. The project will be supervised by Professor Weiya Zhang, Professor David Walsh, and Professor Michael Doherty.

Collaborators:

- All original trial PIs/corresponding authors
- OA Trial Bank
 - An international collaboration to develop an individual patient database of all randomised controlled trials in osteoarthritis (OA). The collaboration has undertaken an IPD meta-analysis to investigate the use of intra-articular steroid injections in OA (More information can be found at: www.oatrialbank.com) and is now undertaking IPD meta-analyses for glucosamine, chondroitin and exercise therapy in OA. This current project also is part of the OA Trial Bank collaboration.
 - It is formalised with a chair, steering committee, and executive coordinator. The steering committee consists of international clinical researchers in the field of OA and all members are involved in the development of the protocol. Please see Figure 1 for further information.

Your participation

Your participation is essential to this project as it relies fully on the contribution of the data from the original trials. If you agree to participate:

1. You will be a collaborator of the project, involved in the database development, being a co-author or being included in the acknowledgements, whichever you prefer.
 - a. Data collected for your RCT will be transferred to the University of Nottingham. These data should include background information collected about study participants (including demographic factors and disease factors), intervention, comparator, outcomes of the trial. Full details will be provided once the agreement is reached.
2. Both parties will sign a licence agreement. This will:
 - a. Cover input data, obligations, term and termination, authorship, and publication
 - b. Specify exactly what can and what cannot be published with your data

What next?

Please let us know if you are interested in participating in our project and on what terms you would wish to do so. If you could please attach the trial protocol for any studies you think may be eligible for inclusion, we will further assess these.

If you have any enquiries, please do not hesitate to contact us.

We look forward to hearing from you.

Kind regards,

Monica Persson, *BMBS BMedSci*

PhD Student

University of Nottingham
Academic Rheumatology
Clinical Sciences Building
City Hospital
Nottingham
NG5 1PB
Tel: +44 (0)115 823 1759

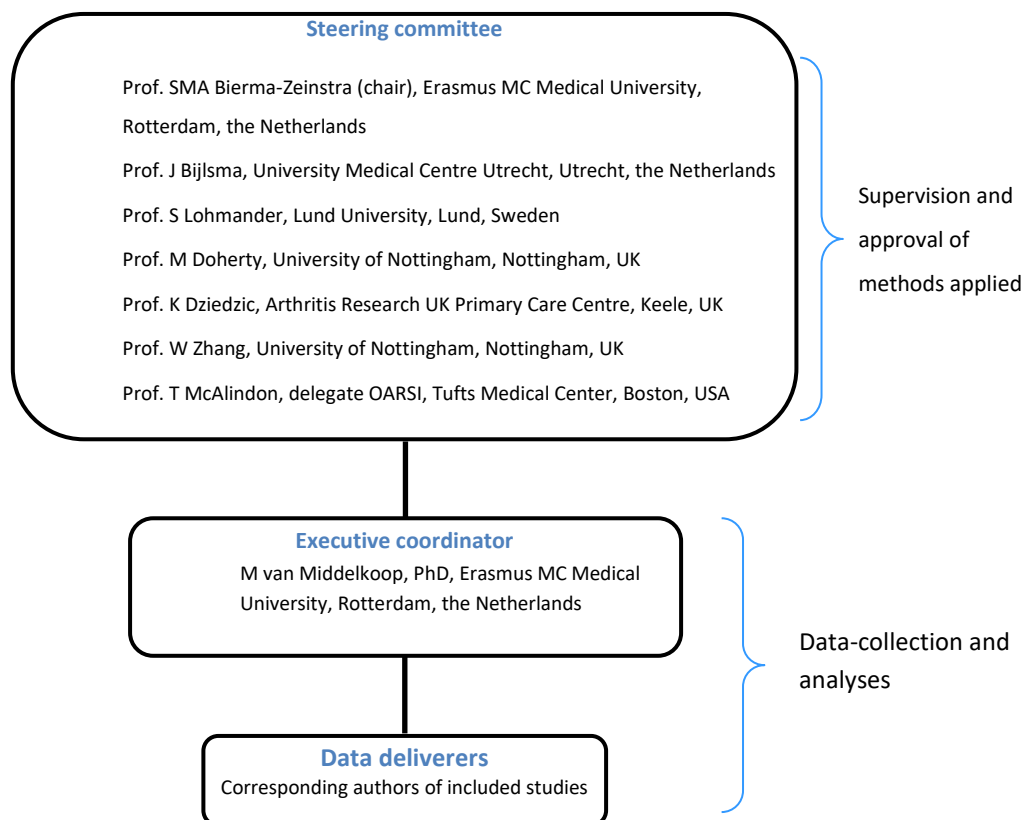


Figure 1. Organisation of OA Trial Bank

Appendix 11 – OA Trial Bank DTA

DATA DELIVERY AGREEMENT

Erasmus MC Medical University, Rotterdam, the Netherlands, an institution organised in accordance with the public law of the Netherlands, more specifically acting for and on behalf of its Department of General Practice, in particular for the OA Trial Bank, having its administrative offices at 's Gravendijkwal 230 in (3015 CE) Rotterdam, the Netherlands, duly represented in this matter by **Monica Persson**

and

[_____]

Jointly referred to as the “Parties”

Whereas:

- Osteoarthritis (OA) is a very heterogeneous disease;
- Given the wide range of available treatments in OA and their small to moderate effectiveness, there is a need for research on clinical predictors of response to different treatments;
- Combining individual patient data from several randomised clinical trials on specific types of treatment would facilitate researchers in defining the clinical predictors of response;
- Erasmus MC intends to create a central databank including data from several randomised clinical trials evaluating one or more interventions in patients with OA of the knee or hip (the OA Trial Bank) in order to enable researchers to perform meta-analyses on individual patient data to define sub-groups that are specifically responsive to certain treatments;
- The OA Trial Bank will be supervised by the OA Trial Bank Steering Committee;
- The Data Deliverer is an experienced researcher in the field of OA and is willing to provide Erasmus MC with data resulting from one or more randomised clinical trials for the OA Trial Bank.

And therefore the Parties agree as follows:

1. Definitions

1.1 In this Agreement, the following words and phrases have the following meanings:

Input Data: pseudonymised or anonymised digital data resulting from one or more clinical trials provided by the Data Deliverer, described in **Annex 1**;

IPR: patents, trade marks, trade names, service marks, domain names, copyrights, moral rights, rights in and to databases (including rights to prevent the extraction or reutilisation of information from a database), design rights, topography rights and all rights or forms of protection

of a similar nature, or having equivalent or the similar effect to any of them which may subsist anywhere in the world, whether or not any of them are registered, and including applications for registration of any of them;

OA: osteoarthritis;

OA Trial Bank: databank including data from randomised controlled trials evaluating one or more interventions in patients with OA of the knee or hip;

OA Trial Bank Steering Committee: the team of international OA researchers, renown in the field, representing the main disciplines in clinical OA research, supervising the access to and use of data in the OA Trial Bank, setting rules for data sharing, development of study questions and of data pooling and analysis;

Output Data: the data and other material generated through analysis and/or manipulation of the Input Data provided by the Data Deliverer as well as equivalent data provided by other participating centres;

Permitted Users: the people who are authorised by the OA Trial Bank Steering Committee to access and use (parts of) the OA Trial Bank, including, but not confined to, the research coordinator of the OA Trial Bank, members of the OA Trial Bank Steering Committee for monitoring and/or auditing purposes as well as Data Deliverers that are actively involved in data analyses.

2. Delivery of Input Data for the OA Trial Bank

2.1 The Data Deliverer grants Erasmus MC a royalty-free, non-exclusive licence in respect of Input Data to import and include such Input Data in the OA Trial Bank so as to allow the use and processing of such Input Data for further statistical and research analysis by Permitted Users in accordance with the rules set by the OA Trial Bank Steering Committee and in accordance with **Annex 1**.

2.2 The Data Deliverer will be contacted and its permission will be requested for additional analyses not described in **Annex 1**.

2.3 The Data Deliverer acknowledges that Input Data will be integrated with data provided by other participating centres in order to enable Permitted Users to conduct further research in accordance with **Annex 1** and to create Output Data.

3. Input Data

3.1 The Data Deliverer warrants and undertakes that the Input Data are the result of one or more randomised clinical trials conducted in accordance with Good Clinical Practice and Academic Standards, in accordance with the laws and procedures as to the direction and conduct of medical studies involving patients applicable in the country where the clinical trial is performed, in particular in accordance with the rules and regulations regarding patient's informed consent. These

Standards, Procedures and Laws are considered compliant with the requirements of the Declaration of Helsinki (amended by the 59th WMA General Assembly, Seoul, October 2008) on the subject of Clinical Trials.

- 3.2 The Data Deliverer warrants and undertakes that it has complied with and shall continue to comply with all relevant legislation, regulations, codes of practices, guidance and other requirements of any relevant government or governmental agency as may apply to the Data Deliverer's possession and disclosure of the Input Data to Erasmus MC.
- 3.3 The Data Deliverer warrants that it has the authority to enter into this Agreement and that the processing of the Input Data by Erasmus MC in the manner envisaged by this Agreement does not and shall not breach any provision of any applicable legislation or agreement or understanding with other parties or individuals.
- 3.4 The Data Deliverer shall indemnify and hold harmless Erasmus MC against any and all loss, claim or expense suffered by Erasmus MC as a result of the Data Deliverer's breach of this clause.

4. Obligations of Erasmus MC/ OA Trial Bank

- 4.1 Erasmus MC will process any data comprised within the Input Data only for such purposes as are reasonably necessary in accordance with this Agreement. Erasmus MC will ensure that it has in place and will maintain appropriate technical and organisational measures against unauthorised or unlawful processing of the Input Data and against accidental loss or destruction of or damage to the Input Data.

5. IPR

- 5.1 All IPR on the Input Data is and will remain with the Data Deliverer. All IPR on the Output Data is and will remain the exclusive property of Erasmus MC. Neither Party will make any claim against the IPR of the other Party.

6. Term and termination

- 6.1 This Agreement shall become effective on the date this Agreement is signed by the last Party and shall remain in force until termination in accordance with this clause.
- 6.2 Either Party may terminate this Agreement subject to six months notice in writing to the other Party.
- 6.3 Either Party may terminate this Agreement with immediate effect by giving notice in writing to the other Party upon the occurrence of any of the following events:

a.) If the defaulting Party commits a material breach of its obligations under this Agreement and (if the breach is capable of remedy) fails to remedy the breach within 30 days of being specifically required to do so in writing by the non-defaulting Party;

- b.) If any distress, execution, sequestration or other similar process is levied or enforced upon or against property of the defaulting Party which is not discharged within thirty days, or an encumbrancer takes possession of, or an administrator, an administrative receiver, a receiver, a trustee or a liquidator is appointed over the whole or any substantial part of the defaulting Party's undertaking, property or assets, or an order is made or a resolution is passed for the winding-up or analogous proceedings in any jurisdiction of the defaulting Party.
- 6.4** In the event that Erasmus MC identifies serious and/or persistent non-compliance with the obligations outlined in clause 3.1 of this Agreement, Erasmus MC will have the right to terminate the Agreement with immediate effect and to remove the Input Data from the OA Trial Bank.
- 6.5** In the event of termination, Input Data that have already been analysed and used to generate Output Data will be kept available by the Data Deliverer for monitoring and/or auditing purposes for a period of 15 years after termination.
- 7. Authorship and publication**
- 7.1** Erasmus MC intends to publish the results of the analyses of the Input Data in reputable scientific and medical journals and at scientific conferences.
- 7.2** Authorship and acknowledgements follow the criteria established by the International Committee of Medical Journal Editors (ICMJE). According to these guidelines, authorship credit is based only on (i) substantial contribution to concept and design, or acquisition of data, or analysis and interpretation of data; and (ii) drafting or revising the manuscript for essential intellectual content; and (iii) approval of the final version to be published. All three aforementioned criteria must be fulfilled. Consistent with these and major journal guidelines, those individuals who meet all authorship criteria should be named as authors and those who do not should be acknowledged elsewhere, if appropriate.
- 7.3** In addition to clause 7.2, the Parties agree on the fact that the research coordinator will be first author in case the data analysis is performed by the research coordinator. Where a Data Deliverer or one of its employees is willing to perform the data analyses under the supervision of the research coordinator and is permitted to do so by the OA Trial Bank Steering Committee, the Data Deliverer or one of its employees will be first author while the research coordinator will be mentioned as co-author. At least two members of the OA Trial Bank Steering Committee will be named as co-author in each publication, in accordance with clause 7.2.
- 7.4** Where a Data Deliverer or one of its employees is willing to perform the data analyses and is permitted to do so by the OA Trial Bank Steering Committee, the Data Deliverer and the research coordinator will agree upon a timeframe for the analyses of the Data, writing a

draft publication and submitting this publication to a renowned medical journal.

- 7.5 Where several Data Deliverers are willing to perform the data analysis and the parties involved disagree about the right person to do so, the parties involved shall make all reasonable efforts to settle disputes arising from or in connection with this issue in an amicable way. Any disputes that remain unresolved shall be decided upon by the OA Trial Bank Steering Committee, taking into account the amount of Input Data delivered by each Data Deliverer and the experience of the Data Deliverers in comparable data analyses.

8. Notices

- 8.1 Any notices which are required to be given or which shall be given under this Agreement shall be in writing delivered by facsimile or by regular mail (airmail if not domestic) addressed to the Parties as follows:

Erasmus MC:	Monica Persson	Data Deliverer:
	Room A26, Academic Rheumatology	
	Clinical Sciences Building	
	City Hospital	
	Nottingham	
	NG5 1PB	
	UK	

9. Assignment

- 9.1 This Agreement shall not be assignable by either Party without the prior written consent of the other Party.

10. Independent contractor

- 10.1 For the purposes of this Agreement and all services to be provided hereunder, each Party shall be, and shall be deemed to be, an independent contractor and not an agent or employee of the other Party. Neither Party shall have the authority to make any statements, representations or commitments of any kind or to take any action which shall be binding on the other Party, except as may be explicitly authorised by the other Party in writing.

11. Governing law

- 11.1 The validity and interpretation of this Agreement and the legal relationship of the Parties to it shall, in all respects, be governed by the laws of the Netherlands. Any and all disputes between the Parties that cannot be settled amicably shall be subject to the exclusive jurisdiction of the court having competence in any such matter at Rotterdam, the Netherlands, except for disputes described in clause 7.5.

12. Entire agreement

- 12.1 Unless otherwise specified, this Agreement (including the annexes thereto) embodies the entire understanding between Erasmus MC and

the Data Deliverer, and any prior or contemporaneous representations, either oral or written, are hereby superseded. No amendments or changes to this Agreement shall be effective unless made in writing and signed by authorised representatives of the Parties.

AGREED AND SIGNED BY BOTH PARTIES:

Erasmus MC:

Data deliverer:

Name: _____

Name: _____

Date: _____

Date: _____

Signature:

Signature:

Annex 1: Description of data provided by the Data Deliverer and research protocol for analysis of Input Data

ANNEX 1 (DATA DELIVERY AGREEMENT)

Methods/Design

We will conduct an individual patient data (IPD) meta-analysis of randomised controlled trials (RCTs) to investigate the efficacy and predictors of topical NSAIDs or topical capsaicin in patients with OA. This is an alternative method that increases the power of subgroup analysis compared to a single RCT. In addition, we will conduct a network meta-analysis to examine the relative effect between topical NSAIDs and topical capsaicin.

Study selection

Type of studies

Published and unpublished RCTs that evaluate the efficacy of either topical NSAIDs or topical capsaicin in participants with OA will be included. There will be no language or geographical restrictions applied to the studies.

Participants

Men and women with physician diagnosed OA, as defined by the criteria endorsed by the ACR (Altman et al., 1990, Altman et al., 1986). Participants with chronic joint pain, not due to rheumatoid arthritis or other forms of arthritis, will also be included. Studies investigating only OA patients and studies with a subgroup of OA patients will be included, as long as IPD data can be collected separately for the OA participants. Participants with back pain and other arthritic pain, such as rheumatoid arthritis pain, will be excluded.

Types of interventions

Only topical NSAIDs or topical capsaicin will be studied. These can include any formulation, dosage, or drug within the group. Salicylates, although pharmacologically related to NSAIDs, will be excluded as their principal action when applied topically is different from that of topical NSAIDs (Derry et al., 2014). Interventions involving ultrasound or magnetophoresis of the topical treatment will also be excluded. Controls will include placebo, active comparators, continuing usual treatment, no treatment, and head to head comparison between topical NSAIDs and topical capsaicin. Both open label and blinded studies will be included.

Types of baseline assessments

As a minimum, studies will need to have assessed the level of pain and important patient characteristics, including age and gender, at baseline.

Types of outcomes

For inclusion, trials must have an adequate assessment of pain. As recommended by the OMERACT-OARSI Initiative (Pham et al., 2004), measures of functional impairment and patient's global assessment will also be included in analysis when available.

Duration of follow up will be a minimum of one week, and the primary outcome will be measured at four weeks or closest to four weeks.

Identification of eligible studies

A systematic database search of Medline (OVID interface, 1946 onwards), Embase (OVID interface, 1974 onwards), AMED (OVID interface, 1985 onwards), CINAHL (EBSCOhost interface, all years), Web of Science (Core Collection, 1900 onwards), and the Cochrane library (Wiley interface, current issue) will be conducted to identify all relevant RCTs of topical NSAIDs or topical capsaicin in OA up until 16/11/2015. The search strategies are based on those from the Cochrane reviews of topical NSAIDs (Derry S, 2012) and topical capsaicin (Derry and Moore, 2012, Derry S, 2013), with the inclusion of other terms found through searching the internet. The search strategies used can be found in Appendix 1. Further studies will also be identified by contacting pharmaceutical suppliers of topical NSAIDs and topical capsaicin, identified via the British National Formulary, the electronic Medicines Compendium, and Clinicaltrials.gov, to ask for any unpublished trial data. Reference lists of included trials and reviews in the area will also be hand-searched for any trials not found from searching the databases. Review authors and collaborating authors will be asked if they are aware of any eligible studies not yet included.

References will be exported to Endnote, where duplicates will be removed and eligibility for inclusion can be assessed. One of the investigators (MP) will assess if studies meet the inclusion criteria by screening the titles, the abstracts, and finally the full texts of articles found through the systematic search. The list of included and excluded studies will then be independently reviewed by another investigator (YF), who will assess all full texts. Any disagreements regarding the inclusion of studies will be discussed between the two review authors. If no consensus can be reached, a third investigator will be consulted.

Data collection

The primary or corresponding author for all eligible trials will be contacted and asked to collaborate on the project. If they cannot be contacted, the author listed last will be emailed. If none of the authors can be reached, the institution where the research was carried out will be contacted. Authors or institutions interested in collaborating will be asked for access to the IPD from the original study. The data will be accepted in any format and will be transferred into Microsoft Access for analysis. The data will be contributed to the OA Trial Bank by the authors, as described by van Middelkoop *et al.* (van Middelkoop *et al.*, 2013).

Data extraction

Data extracted from the published reports will include the distribution of participant demographics and disease characteristics, study information and design, details of the intervention, and outcome measures. These will be entered into a data extraction form by MP, and will be independently verified by another investigator.

Data contributed by the original study authors will include patient characteristics such as age, gender, and body mass index (BMI). If available, other characteristics will be collected including duration of complaints, radiographic change, level of inflammation (e.g., clinical or ultrasound effusion, CRP), type of pain (e.g., neuropathic like pain through PainDetect), central sensitization (e.g., pain elsewhere through a questionnaire or pain threshold through quantitative sensory testing), and psychological assessments (e.g., depression, anxiety, sleep disturbance, catastrophising). Study-level characteristics, such as setting (community or hospital), allocation concealment, blinding, sample size, risk of bias (low vs. high), intervention(s) used, control(s) used, doses and formulations, and duration of follow-up will also be required. Finally, authors will be asked to contribute the outcome measures of pain, function, and patient's global assessment at baseline and at all subsequent assessments.

All participants that have been randomised will be included in the pooled database and will be included in intention-to-treat (ITT) analyses. Secondary per-protocol analyses will also be carried out.

Quality assessment

The quality of included studies will be assessed using a modified version of the Risk of Bias tool recommended by the Cochrane Collaboration (Higgins JPT, 2011) (Table 1 and Appendix 2). It is composed of 9 questions that assess each of the domains set out in the Risk of Bias tool. In contrast with the narrative nature of the Cochrane tool, the modified version uses questions that are scored as yes, no, or unclear. Each question evaluates the risk of bias in an objective manner, so as to minimise the inter-rater disagreement that is associated with the more subjective domains in the Risk of Bias tool (Hartling et al., 2009). As the reviewer carrying out the quality assessment will vary by the language of the study, these modifications were aimed at reducing inter-rater variability in the quality scoring.

The first five questions of the modified tool assess selection bias, performance bias, and detection bias. Each of these questions will be scored using the criteria outlined by Cochrane (Higgins JPT, 2011). The sixth question is a measure of attrition bias, and similarly to the PEDro and van Tulder quality scales (Olivo et al., 2008), a cut-off for acceptable drop-out rates was used to determine the risk. After discussion between review authors, a consensus of 15% cut-off for attrition bias was agreed upon (PEDro, 1999). The seventh question addresses selective reporting by comparing the outcomes pre-specified in the methods with those presented in the results. Within the domain "Other bias", two questions cover the use of intention-to-treat analysis and whether the treatment and control group were similar at baseline. Although not pre-specified in Cochrane's risk of bias (Higgins JPT, 2011), these two sources of bias are assessed in a variety of other RCT quality scales (Olivo et al., 2008).

Similar to the protocol of van Middelkoop *et al.* (van Middelkoop et al., 2013), each study will be categorised as "low risk of bias" if it meets at least 5 of the criteria items in the list. Studies will then be categorised as "low risk of bias", "high risk of bias", or "unclear". With the exception of non-English

publications, the quality assessment of each study will be undertaken independently by two authors. If the data extracted is from a study in a language not spoken by either or both authors, the risk of bias assessment will be done only by the author or a colleague that speaks the language. The risk of bias score will be used to assess study-level predictors of response.

Table 1. Modified Risk of Bias assessment

Source of bias	Yes ^a	No	Unclear	Comments
1. Was the randomisation procedure adequate?	☐	☐	☐	
2. Was the treatment allocation adequately concealed?	☐	☐	☐	
3. Were participants adequately blinded to the intervention?	☐	☐	☐	
4. Were physicians adequately blinded to the intervention?	☐	☐	☐	
5. Were outcome assessors adequately blinded to the intervention?	☐	☐	☐	
6. Incomplete outcome data: Is the attrition rate <15%?	☐	☐	☐	
7. Are all pre-specified outcomes of interest reported in the pre-specified way?	☐	☐	☐	
8. Was intention-to-treat analysis used?	☐	☐	☐	
9. Were the treatment and control group similar at baseline?	☐	☐	☐	

Modified from Cochrane's Risk of Bias tool (Higgins JPT, 2011). ^aSee Appendix 2 for criteria for a judgment of "yes"

In addition, the quality of the IPD data will be assessed. The quality indicators include number of trials eligible per treatment, number of trials with the IPD data obtained, and percentage of the data obtained per trial.

In order to determine how representative the IPD analysis is, a comparison will be made of the patient characteristics between all eligible trials and those where IPD was obtained. Characteristics of studies that are not included, due to missing raw data, unwillingness to collaborate, or inability to reach study authors, will be presented and compared to those of the included studies. Furthermore, sensitivity analyses will be conducted to compare the summary effects of the included studies with the summary effects extracted from the published reports of the studies that were not included. Finally, assessments of funnel plot asymmetry will be conducted to assess for publication bias.

Data analysis

Once data have been received, the databases will be cleaned and merged. Means and standard deviations will be used to describe normally-distributed continuous data, medians and interquartile range will be used to describe continuous data that are not normally distributed, and frequencies and

percentages will be used for categorical data. Data will be described using 95% confidence intervals. P-values <0.05 will be considered significant. Any missing data will be presumed to be missing at random, therefore, a multiple imputation method will be used within each trial before pooling the data. I^2 will be calculated as a measure of the heterogeneity of the included trials. Data analysis will be conducted using the statistical program Stata SE 14 (StataCorp, College Station, Texas)

The primary outcome will be pain at 4 weeks treatment duration, or closest to 4 weeks. Secondary outcomes will include pain at other durations of follow-up, area under the curve (AUC) for pain scores at the different time points, function, and global assessment measures.

Overall analysis

A conventional meta-analysis will be undertaken using study-level variables such as mean pain score, mean age, mean BMI, and sample size. This analysis will help to determine the treatment effect and its variation between studies. A network meta-analysis will be conducted to determine the relative efficacy between topical NSAIDs and capsaicin. Finally, an IPD meta-analysis using a mixed effects model will be conducted, using both study-level and IPD-level variables, to determine the treatment effect and its potential predictors.

Conventional meta-analysis

An aggregate data meta-analysis, using a random-effects model, will be performed to estimate the treatment effect of topical NSAIDs and topical capsaicin over placebo. This will be used for the sensitivity analyses comparing the treatment effect in the studies included in the IPD analysis with all eligible studies. The relative efficacy between the two drugs will be examined directly by comparing topical NSAIDs and topical capsaicin within any available head-to-head comparison trials and indirectly by conducting a network meta-analysis using a common comparator.

Subgroup analysis: When sufficient data are available, subgroup analysis will be performed for the primary outcome - pain at four weeks according to: pain elsewhere (yes/no), type of pain (dull/neuropathic), pain severity (low/high intensity), disease severity (marked/modest disease radiographically), level of inflammation (low/high), and duration of pain.

Network meta-analysis

A number of trials have been undertaken for topical NSAIDs and topical capsaicin in OA. However, there is no head-to-head comparison between these two commonly used topical analgesics. Due to the different mechanisms of actions, we assume that topical capsaicin may be better than NSAIDs in neuropathic-like pain, such as post-herpetic neuralgia, diabetic neuropathy, and end-stage OA with more neuropathic damage. This begs a comparison between these two agents in different pain models. A network meta-analysis has therefore been proposed (Lumley, 2002).

We will use topical placebo as a common comparator (or “node”) to network topical NSAIDs and topical capsaicin in each condition to calculate the relative difference between the two active treatments. To maximise the information, we will also include trials comparing topical NSAIDs or capsaicin (if any exist) with oral NSAIDs or other analgesics. In this scenario, oral NSAIDs or other analgesics will be the “node” to link topical NSAIDs and topical capsaicin. If needed, more than one “node” will be used to link the two treatments, and a more complicated network may be developed.

Once the network is developed, the relative efficacy between topical NSAIDs and topical capsaicin will be calculated. Statistical pooling will be undertaken taking into consideration the variations between studies and comparisons. In order to increase the precision of the estimate, the Bayesian statistical approach will be applied (Smith et al., 1995).

IPD meta-analysis

Two stage modelling: A regression model will be developed for each trial and significant interactions/predictors will be pooled between trials. Depending on the definition of the treatment effect (dependent variable), the model will be built with or without interaction terms in order to identify the predictors. For the specific treatment effect (i.e., the difference between treatment and placebo), an interaction term between treatment and potential predictor will be used to define the predictors of treatment response. The partial regression coefficient of this interaction will be selected and pooled with the same interaction terms obtained from the other trials. For the overall treatment or placebo response (i.e., improvement from baseline), the prediction model is developed within each arm, therefore there is no need to use an interaction term to identify a predictor. In this case, the partial regression coefficient of the predictor will be used and pooled between trials.

One stage modelling: Treatment responders will be determined using a one-step random-effects IPD meta-analysis approach, i.e., taking into account both study level and individual patient level covariates in the regression model. Individual patient level covariates will be centred to the mean of the covariate in each trial. The overall treatment effect (i.e., change from baseline in participants in the treatment group) and specific treatment effect (i.e. difference between treatment and placebo) will be calculated (Zou et al., 2016). Two multi-level regression models will be developed to consider patient level and study level (i.e., cluster) effects, one to examine predictors of the specific treatment effect and the other to examine the predictors of the overall treatment effect. The first model will include participants from both intervention groups where the specific treatment effect will be the dependent variable and treatment (active or placebo) and patient characteristics will be independent variables. Interaction between treatment and predictors will be used to identify the effect-modifiers. The second model will include participants only in the treatment group where the overall treatment effect will be the dependent variable and patient characteristics will be independent variables. This model is based on the assumption that any treatment effect includes both specific and non-specific contextual effect (i.e., placebo effect) (Zou et al., 2016), and in clinical practice we only give treatment not placebo. The aim of this model is to identify the treatment responders and factors

related to the response. The threshold for response will be established based on the distribution of the overall treatment effect (e.g., comparing the distributions between baseline pain score versus endpoint pain score to find a cut-off point), and a binomial model will be developed as appropriate.

Potential predictors

All baseline variables described in “Data extraction” will be considered potential predictors of treatment response. As a minimum, we will utilise age, gender, BMI, baseline pain, pain elsewhere, evidence of inflammation (clinical or imaging), and radiographic findings as patient level variables. For study level variables, we will include joint treated, sample size, setting (community/hospital), and allocation concealment as a minimum.

Appendix 12 – UoN drafted DTA

Data Transfer Agreement



This agreement is made this __ day of ____ between:

- 1 _____, of [address] (referred to as the "Provider"); and
- 2 **The University of Nottingham** of University Park, Nottingham, NG7 2RD, (the "Recipient").

Whereas:

The Provider has collected and owns certain data sets and collections of interpretable data as more particularly set out in the Appendix (the "Data") derived from information provided by participants in various studies (the "Study Participants").

The Recipient wishes to use the Data in the research described in the Appendix 1 (the "Research"). The Provider is willing to supply the Recipient with the Data to conduct the Research under the terms and conditions of this Agreement.

It is agreed as follows:

1. The Provider shall make the Data available to the Recipient for a term of 15 years (the "Term"), subject to earlier termination of this Agreement in accordance with its terms.
2. The Data remains the property of the Provider. There is no transfer or licence or implied transfer or licence of rights in the Data from the Provider to the Recipient including (without limitation) of any intellectual property rights therein. This Agreement does not restrict the rights of the Provider in the Data, including (without limitation) its ability to distribute any of the Data to other institutions or to publish any document relating to or including the same.
3. The Recipient shall maintain the confidentiality of the Data. Where the Data have been anonymised, the Recipient shall not attempt to identify any Study Participants. The Recipient shall retain the Data in a secure location on its premises and will not permit the Data or any part of it to come into the possession or control of any other organisation or any individual other than those employees who are involved in the Research. The Recipient shall ensure that such employees are bound to it by contractual commitments which maintain the confidentiality of the Data.

4. The Recipient shall not transfer the Data in whole or in part to third parties without prior written permission from the Provider, which the Provider may grant in its absolute discretion.
5. The Recipient will use the Data only to carry out the Research described in the Appendix 1 to this Agreement, and only for Research that has appropriate ethical approval. The Recipient shall not, without the prior written permission of the Provider, use the Data or any parts thereof for any commercial purpose or for any purpose that may render the Data subject to rights in favour of third parties.
6. The Provider may terminate this Agreement (i) at any time, on giving no less than three months' written notice; or (ii) if the Recipient is in breach of any of the conditions of this Agreement, by immediate notice. Unless terminated earlier, this Agreement shall expire at the end of the Term.
7. The Provider may modify or withdraw all or part of the Data at any time if such action is required by law or by the withdrawal of consent of a Study Participant. Immediately following termination or such modification or withdrawal, the Recipient shall, at the Provider's option, return or destroy the relevant part of the Data then in its possession or under its control and will confirm in writing to the Provider such action as has been taken.
8. The Results shall be the property of the Recipient save that the Recipient grants to the Provider a royalty-free, irrevocable, perpetual non-exclusive right to use such Results for internal non-commercial research and teaching.
9. The Recipient will acknowledge the Provider in all publications relating to the Results arising from the Research.
10. The Data is supplied without cost but, if agreed between the parties in writing, the Recipient shall reimburse the Provider for the costs of preparing and sending the Data to the Recipient.
11. The Recipient will neither assign, transfer, mortgage, charge nor part with any of its interests, rights, duties or obligations under this Agreement.

Signed on behalf of

.....*sign here*

by

.....*print name*

Authorised signatory

Signed on behalf of
The University of
Nottingham

.....*sign here*

by

Appendix 1:

The Research Project Title:	The Relative efficacy of topical NSAIDs and topical capsaicin in osteoarthritis and neuropathic pain
Investigator:	Monica Persson
Description of Data:	Data contributed by the original study authors will include patient characteristics such as age, gender, BMI. If available, authors will be asked to provide disease-specific data including pain elsewhere, character of pain, pain severity, disease severity, level of inflammation, duration of pain. Study-level characteristics, such as trial number, types of interventions, and doses will also be required. Finally, authors will be asked to contribute the outcome measures of pain, function, and patient's global assessment at baseline and follow-up.
Description of the Research:	The Investigator will be conducting an individual patient data meta-analysis of randomised controlled trials investigating topical NSAIDs and topical capsaicin in osteoarthritis and neuropathic pain (painful diabetic neuropathy and post-herpetic neuralgia). It will aim to identify predictors of treatment response.

Appendix 13 – List of variables from IPD MA trials

Relative efficacy of topical non-steroidal anti-inflammatory drugs and topical capsaicin in osteoarthritis: protocol for an individual patient data meta-analysis



The University of Nottingham

Title of publication	Randomized clinical trial evaluating transdermal ibuprofen for moderate to severe knee osteoarthritis		
Corresponding author	Gyula Varadi		
Year of publication	2013		
Journal	Pain Physician		

BASELINE DATA	AVAILABLE	TYPE OF DATA (numeric/continuous/date/string)	REMARKS
Patient ID	☒		
Date of randomisation/inclusion	☐		
Age (or date of birth)	☐		
Gender	☐		
Weight	☐		
Height	☐		
BMI	☐		
Joint affected by OA	☐		
Duration of complaints	☐		
Type of intervention	☐		
Signs of inflammation:	☐		<i>Please specify which are available</i> ☐ Physical examination ☐ Imaging (sonography, MRI): Please specify ☐ Additional testing (ESR, CRP, inflammatory markers): Please specify
OA characteristics:	☐		☐ Other: Please specify <i>Please specify which are available</i> ☐ ACR criteria ☐ Severity ☐ Radiographic changes ☐ Kellgren and Lawrance score ☐ Other: Please specify
Nature of pain (dull/neuropathic)	☐		<i>Please specify which are available</i> ☐ PainDETECT questionnaire ☐ DN4 questionnaire ☐ Other: Please specify
Central sensitization:	☐		<i>Please specify which are available</i> ☐ Pain elsewhere (questionnaire): Please specify ☐ Quantitative sensory testing ☐ Other: Please specify
Psychological assessments: Anxiety, depression, sleep disturbance, catastrophising etc.	☐		<i>Please specify which are available</i> ☐ Questionnaire: Please specify ☐ Other: Click here to enter text.
Other variables	☐		<i>Please specify which other variables are available:</i> Click here to enter text.

OUTCOME MEASURES	MEASURED AT		TIMINGS OF FOLLOW-UP MEASUREMENT (weeks)	TYPE OF MEASUREMENT (VAS, WOMAC, KOOS)	TYPE OF DATA (numeric/continuous/ categorical)	REMARKS
	Baseline	Follow-up				
Pain severity (VAS, NRS, WOMAC, etc.)	☐	☐				
Physical functioning (WOMAC, KOOS, KSS, etc.)	☐	☐				
Patient global assessment	☐	☐				
Quality of life (SF-36, SIP, NHP, OAKHQoL, etc.)	☐	☐				
Other outcomes	☐	☐				<i>Please specify</i>

FORMAT IN WHICH DATA ARE SAVED: (SPSS, SAS, STATA, EXCEL,...)	

Appendix 14 – DTA for collaboration with VSM



The University of Nottingham
University Park,
Nottingham NG7 2RD,
United Kingdom

Clinical Research

Date: November 8, 2017
Initials:

Contact: Bernadette Krom
Phone: +31(0)72-566 11 46
Fax: +31(0)72-562 38 83
E-Mail: Bernadette.Krom@vsm.nl

183-16 (100)

Dear Ms. Persson,

In this letter we would like to record our arrangements regarding the analysis of raw data of the clinical trial „A randomized controlled trial comparing topical piroxicam gel with a homeopathic gel in osteoarthritis of the knee" (study number P9124), we conducted from April 1993 until August 1995.

You, The University of Nottingham, are planning to conduct an individual patient data (IPD) meta-analysis of topical NSAIDs and topical capsaicin in osteoarthritis and neuropathic pain at the University of Nottingham. Your analysis is intended to include, amongst others, the P9124 data. We will provide you the corresponding raw data of the P9124 trial free of cost and to an extent which is at our discretion.

You will use the provided raw data only for the above mentioned analysis. Any other use beyond the extent of this analysis shall require prior written consent by VSM.

The results of the analysis ("Results") shall be the property of The University of Nottingham save that The University of Nottingham grants to the VSM a royalty-free, irrevocable, perpetual non-exclusive right to use such Results for internal, non-commercial research and teaching. For the avoidance of doubt, VSM, DHU and its affiliates are permitted to use the Results for commercial purposes such as, without limitation to, advertising, promotion and product development, after publication of the Results

In accordance with normal academic practice, The University of Nottingham will acknowledge VSM in all publications relating to the Results arising from the analyses utilising VSM's data. You will send the publication manuscript composed on the basis of the raw data to VSM at least 30 days prior to any submission to (i) a publishing house, (ii) to a sponsor of your project (above all the Arthritis Research UK and the OA Trial Bank) or (iii) to an internet-based publication platform or (iv) in case of any other publication.

Upon receipt of the manuscript, we, as well as our then study coordinator Dr. Robert van Haselen (13 Rue de la Mesange, 67000 Strasbourg, France) and Deutsche Homöopathie Union, Ottostraße 24, 76227 Karlsruhe, Germany, (DHU) will assess its content. You agree not to publicize the Results before this assessment has been completed, which is when we submit our consent in writing. Such consent shall not be unreasonably withheld.

The University of Nottingham shall maintain the confidentiality of the data and any information or instructions received with it and which is marked as confidential from VSM (hereinafter "data"). Where the data have been anonymised, The University of Nottingham shall not attempt to identify any study participants. The University of Nottingham shall retain the data in a secure location on its premises and will not permit the data or any part of it to come into the possession or control of any other organisation or any individual other than those employees who are involved in the analyses. The University of Nottingham shall ensure that such employees are bound to it by contractual commitments which maintain the confidentiality of the data. The University of Nottingham shall not transfer the data in whole or in part to third parties without prior written permission from VSM, which VSM may grant in its absolute discretion.

Any modifications of and amendments to this agreement as well as ancillary agreements shall only be effective, if agreed in writing. The same applies to a modification or cancellation of this written form clause.



This agreement shall be governed by the laws of the Netherlands. The place of jurisdiction shall be Alkmaar, Netherlands. In the event of any dispute, claim, question, or disagreement arising from or relating to this agreement or the breach thereof, the parties shall use their best efforts to settle amicably the dispute, claim, question, or disagreement. To this effect, they shall consult and negotiate with each other in good faith and, recognizing their mutual interests, attempt to reach a just and equitable solution satisfactory to both parties.

Please return one signed copy of this letter to Mrs Josephine van Vreeswijk. Hereby, the agreement will come into existence.

Yours faithfully

VSM Netherlands
Alkmaar

Mrs Josephine van Vreeswijk
General Manager

Date: 09-11-2017

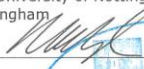
Bernadette Krom

Mrs Bernadette Krom
Scientific Affairs Manager

Date: 09-11-2017

Agreed and accepted:

The University of Nottingham,
Nottingham



Ryan Keyworth
Research and Innovation
Director, Research and Innovation
University of Nottingham

Date: 4 Dec 2017

VSM
Postbus 9321, 1800 GH Alkmaar
Berenkoog 35, 1822 BH Alkmaar

www.vsm.nl
info@vsm.nl
Telefoon +31(0)72 566 11 22

ABN AMRO Nederland
BIC ABNANL2A
IBAN NL16ABNA0643752897

KvK Alkmaar 37046253
BTW nr VSKI NL002843845B01

Page 2/2

Schwabe Group
From Nature. For Health.

Appendix 15 – Pilot n-of-1 trial series UoN ethics approval



E-mail: FMHS-ResearchEthics@nottingham.ac.uk

10th April 2017

Dr Monica Persson
PhD Student
c/o Professor Weiya Zhang, Professor of Epidemiology
Academic Rheumatology,
Rheumatology, Orthopaedics and Dermatology
School of Medicine
Clinical Sciences Building, CHN Campus
Nottingham University Hospitals
Hucknall Road, Nottingham
NG5 1PB

**Faculty of Medicine and
Health Sciences**

Research Ethics Committee
C/o Faculty PVC Office
School of Medicine Education Centre
8 Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham
NG7 2UH

Dear Dr Persson

Ethics Reference No B10022017 – please always quote	
Study Title: Identifying treatment responders to a topical non-steroidal anti-inflammatory drug (NSAID) or topical capsaicin in painful knee osteoarthritis: a pilot series of n-of-1 trials.	
Short Title: Ibuprofen gel or capsaicin cream for my painful knee osteoarthritis?	
Chief Investigator/Supervisor: Professor Weiya Zhang, Professor of Epidemiology, Academic Rheumatology, Rheumatology, Orthopaedics and Dermatology.	
Lead Investigators/student: Dr Monica Persson, PhD Student, Academic Rheumatology	
Other Key Investigators: Professor Michael Doherty, Professor of Rheumatology, Professor David Walsh, Professor of Rheumatology, Dr Joanne Stocks, Translational Research Facilitator, Dr Gwen Fernandes Research Fellow, Dr Aliya Sarmanova PhD student.	
Type of Study: PhD project	
Proposed Start Date: 03/2017	Proposed End Date: 30/09/2018 18 mths
No of Subjects: 22	Age: 18+years
School: Medicine	

Thank you for submitting the above application which was considered by the Committee at its meeting on 10th February 2017 and the following documents were received:

Short Title: Ibuprofen gel or capsaicin cream for my painful knee osteoarthritis?

- FMHS Research Ethics Application form version 1.0: 03/02/2017
- MHRA confirmation that review not required dated 15/11/2016
- Evidence from UoN Sponsors Office that review not required 19/01/2017
- Protocol and supporting documents version 1.0: 03/02/2017

These have been reviewed and are satisfactory and the study is approved.

Approval is given on the understanding that the conditions set out below are followed:

1. You must follow the protocol agreed and inform the Committee of any changes using a notification of amendment form (please request a form).
2. You must notify the Chair of any serious or unexpected event.

3. This study is approved for the period of active recruitment requested. The Committee also provides a further 5 year approval for any necessary work to be performed on the study which may arise in the process of publication and peer review.
4. An End of Project Progress Report is completed and returned when the study has finished (Please request a form).

Yours sincerely

pp *Ravi Mahajan*

Professor Ravi Mahajan
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix 16 – Association between baseline characteristics and overall treatment effect of topical therapy in pilot n-of-1 trial series

Characteristic	Primary analysis			Sensitivity analyses	
	β	95%CI	p value	NRS 4-8 p value	MCID p value
Basic demographics					
Age; years	-0.06	-0.12 to 0.01	0.111	0.179	0.989
Sex; women vs men	1.27	0.12 to 2.42	0.031*	0.048*	0.225
BMI; kg/m ²	0.11	0.00 to 0.22	0.043	0.233	0.372
Questionnaire – comorbidities					
Anxiety; HADS	-0.09	-0.26 to 0.07	0.264	0.233	0.076
Depression; HADS	-0.02	-0.23 to 0.18	0.820	0.619	0.514
Fibromyalgia; fibromyalgia severity score	-0.08	-0.24 to 0.08	0.328	0.191	0.704
Questionnaire – OA features					
Baseline pain; NRS severity	0.80	0.61 to 0.99	<0.001*	<0.001*	0.089
Neuropathic-like pain; Rasch-transformed PDQ	0.54	-0.44 to 1.52	0.277	0.805	0.691
Physical function; KOOS	0.01	-0.02 to 0.05	0.465	0.220	0.035*
Illness perception; Brief Illness Perception score	-0.02	-0.10 to 0.07	0.728	0.278	0.963
Treatment perception; Brief illness perception, treatment sub score	-0.26	-0.53 to 0.00	0.053	0.102	0.889
Central Mechanisms Traits	0.00	-0.18 to 0.18	0.997	0.802	0.211
Examination findings (index knee)					
Quadriceps strength; kg	-0.03	-0.16 to 0.10	0.637	0.967	0.686
Radiographic severity					
Total NLDA score	-0.05	-0.14 to 0.03	0.202	0.221	0.258
NLDA osteophyte score	-0.07	-0.17 to 0.04	0.214	0.216	0.249
NLDA JSN score	-0.26	-0.57 to 0.06	0.107	0.069	0.225
Tibiofemoral KL grade	0.17	-0.44 to 0.78	0.594	0.740	0.654
Patellofemoral KL grade	-0.52	-1.12 to 0.08	0.089	0.113	0.264
Ultrasound features					
Synovial thickness; mm	-0.01	-0.23 to 0.22	0.946	0.702	0.931
Effusion; mm	-0.10	-0.27 to 0.08	0.276	0.127	0.510
Power Doppler positive	-1.40	-3.14 to 0.36	0.118	0.135	0.134
QST features					
Localised PPT (MJL); kPA	-0.00	-0.00 to 0.00	0.365	0.582	0.710
Distal PPT (proximal tibia); kPA	0.00	-0.00 to 0.00	0.991	0.759	0.670
Remote PPT (sternum); kPA	0.00	-0.00 to 0.01	0.503	0.353	0.253
Temporal summation; 0-100 NRS	-0.00	-0.03 to 0.02	0.774	0.552	0.116

Data presented as β coefficient and associated 95%CI from univariate analysis for potential predictors of the overall effect of topical treatments in OA (either topical NSAID or capsaicin). Sensitivity analyses conducted were a) using only treatment periods where baseline pain was 4-8 (inclusive); or b) dichotomising the NRS change scores to represent whether or not the MCID was met. p values for the interaction effect are shown for the sensitivity analyses. * Significant p values (p<0.05) are indicated

Appendix 17 – Treatment-by-baseline characteristic interaction effects in pilot n-of-1 trial series

17a. Primary analysis

Characteristic	β	95%CI	p value
<u>Basic demographics</u>			
Age			
Younger (57 years)	-0.55	-1.27 to 0.17	0.310
Older (76 years)	-0.00	-0.75 to 0.75	
Sex			
Men	-0.10	-0.84 to 0.63	0.487
Women	-0.47	-1.19 to 0.26	
BMI			
Less overweight (25.1 kg/m ²)	0.13	-0.61 to 0.87	0.126
More overweight (36.3 kg/m ²)	-0.70	-1.43 to 0.04	
<u>Questionnaire – comorbidities</u>			
Anxiety; HADS			
Less features of anxiety (2.8 points)	-0.36	-1.14 to 0.42	0.795
More features of anxiety (10.7 points)	-0.22	-0.93 to 0.49	
Depression; HADS			
Less features of depression (1.8 points)	-0.29	-1.04 to 0.48	0.977
More features of depression (8.2 points)	-0.28	-1.03 to 0.47	
Fibromyalgia; fibromyalgia severity score			
Less features of fibromyalgia (5.0 points)	0.07	-0.67 to 0.81	0.199
More features of fibromyalgia (13.0 points)	-0.64	-1.38 to 0.12	
<u>Questionnaire – OA features</u>			
Baseline pain; NRS severity			
Less pain (3.0 points)	-0.02	-0.78 to 0.74	0.167
More pain (7.7 points)	-0.56	-1.31 to 0.20	
Neuropathic-like pain; Rasch-transformed PDQ			
More nociceptive pain (-1.3 points)	-0.08	-0.78 to 0.63	0.405
More neuropathic pain (-0.0 points)	-0.51	-1.24 to 0.23	
Physical function; KOOS			
Worse function (41.3 points)	-0.54	-1.28 to 0.19	0.344
Better function (75.4 points)	-0.04	-0.76 to 0.68	
Illness perception; Brief Illness Perception score			
Less threatening view (29.8 points)	0.00	-0.74 to 0.74	0.297
More threatening view (45.2 points)	-0.55	-1.28 to 0.17	
Treatment perception; Brief Illness perception sub score			
Treatment perceived more helpful (1.5 point)	-0.50	-1.22 to 0.23	0.800
Treatment perceived less helpful (6.1 points)	-0.07	-0.82 to 0.69	

Continued

Characteristic	β	95%CI	p value
Central Mechanisms Traits			
Less central pain (7.0 points)	-0.20	-0.95 to 0.55	0.777
More central pain (14.4 points)	-0.35	-1.06 to 0.36	
Examination findings (index knee)			
Quadriceps strength; kg			
Weaker (12.5 kg)	-.39	-1.13 to 0.36	0.714
Stronger (22.9 kg)	-0.19	-0.91 to 0.52	
Radiographic severity			
Total NLDA score			
Less severe (6.2 points)	-0.38	-1.18 to 0.43	0.790
More severe (22.1 points)	-0.23	-0.96 to 0.51	
NLDA osteophyte score			
Less severe (3.3 points)	-0.40	-1.20 to 0.40	0.719
More severe (16.1 points)	-0.20	-0.95 to 0.54	
NLDA JSN score			
Less narrowing (2.6 points)	-0.18	-0.94 to 0.58	0.696
More narrowing (6.5 points)	-0.38	-1.06 to 0.30	
Tibiofemoral KL grade			
0	-0.78	-2.25 to 0.70	0.206
1	-0.90	-2.18 to 0.38	
2	0.59	-0.40 to 1.58	
3	-0.29	-1.06 to 0.48	
4	-1.92	-4.47 to 0.63	
Patellofemoral KL grade			
0	-1.61	-2.93 to -0.29	0.138
1	0.25	-0.57 to 1.08	
2	-0.24	-1.72 to 1.24	
3	-0.38	-1.38 to 0.62	
Synovial thickness; mm			
Less inflammation (≤ 2.6 mm)	-0.71	-1.43 to 0.01	0.101
More inflammation (≥ 8.4 mm)	0.14	-0.58 to 0.85	
Effusion; mm			
Less inflammation (≤ 5.2 mm)	-0.50	-1.24 to 0.23	0.413
More inflammation (≥ 12.5 mm)	-0.07	-0.79 to 0.65	
Power Doppler			
Positive (present)	-0.43	-0.98 to 0.12	0.175
Negative (absent)	0.58	-0.77 to 1.93	
Localised PPT (MJL); kPA			
Lower, more pain sensitivity (170.8 kPA)	0.05	-0.71 to 0.82	0.242
Higher, less pain sensitivity (659.3 kPA)	-0.56	-1.24 to 0.13	
Distal PPT (proximal tibia); kPA			
Lower, more pain sensitivity (181.8 kPA)	0.06	-0.68 to 0.81	0.210
Higher, less pain sensitivity (612.9 kPA)	-0.58	-1.26 to 0.11	
Continued			

Characteristic	β	95%CI	p value
Remote PPT (sternum); kPA			
Lower, more pain sensitivity (129.9 kPA)	0.14	-0.61 to 0.88	0.131
Higher, less pain sensitivity (437.5 kPA)	-0.63	-1.31 to 0.05	
Temporal summation; 0-100 NRS			
Lower, less pain sensitivity (13.6 points)	-0.73	-1.46 to -0.00	0.094
Higher, more pain sensitivity (62.1 points)	0.18	-0.57 to 0.92	

Data presented as treatment coefficient (β) and associated 95%CI per predictor subgroup category. Continuous data were subgrouped as a) one SD below and one SD above the mean predictor value. Treatment coefficients <0 favour capsaicin, >0 favour ibuprofen, and =0 show no difference between the treatments.

p values for the interaction coefficient are provided, with significant p values (p<0.05) indicated (*).

BMI, body mass index; HADS, Hospital Anxiety and Depression Scale; JSN, joint space narrowing; KL, Kellgren-Lawrence; KOOS, Knee Injury and Osteoarthritis Outcome Score; MJL, medial joint line; NLDA, Nottingham Line Drawing Atlas; NRS, numeric rating scale; PDQ, painDETECT questionnaire; PPT, pressure pain threshold

17b. Sensitivity analysis – baseline pain 4-8

Characteristic	β	95%CI	p value
<u>Basic demographics</u>			
Age			
Younger (57 years)	-0.74	-1.58 to 0.10	0.606
Older (76 years)	-0.41	-1.37 to 0.55	
Sex			
Men	-0.52	-1.42 to 0.39	0.760
Women	-0.72	-1.66 to 0.22	
BMI			
Less overweight (25.1 kg/m ²)	-0.29	-1.24 to 0.66	0.378
More overweight (36.3 kg/m ²)	-0.87	-1.74 to 0.00	
<u>Questionnaire – comorbidities</u>			
Anxiety; HADS			
Less features of anxiety (2.8 points)	-0.90	-1.92 to 0.13	0.435
More features of anxiety (10.7 points)	-0.37	-1.20 to 0.46	
Depression; HADS			
Less features of depression (1.8 points)	-0.54	-1.50 to 0.41	0.900
More features of depression (8.2 points)	-0.63	-1.53 to 0.28	
Fibromyalgia; fibromyalgia severity score			
Less features of fibromyalgia (5.0 points)	-0.22	-1.19 to 0.75	0.280
More features of fibromyalgia (13.0 points)	-0.92	-1.78 to -0.07	
<u>Questionnaire – OA features</u>			
Baseline pain; NRS severity			
Less pain (3.0 points)	-0.94	-1.89 to 0.02	0.347
More pain (7.7 points)	-0.28	-1.21 to 0.64	
Neuropathic-like pain; Rasch-transformed PDQ			
More nociceptive pain (-1.3 points)	-0.68	-1.67 to 0.31	0.774
More neuropathic pain (-0.0 points)	-0.49	-1.34 to 0.37	
Physical function; KOOS			
Worse function (41.3 points)	-0.68	-1.55 to 0.19	0.689
Better function (75.4 points)	-0.42	-1.36 to 0.51	
Illness perception; Brief Illness Perception score			
Less threatening view (29.8 points)	-0.53	-1.66 to 0.60	0.861
More threatening view (45.2 points)	-0.66	-1.48 to 0.17	
Treatment perception; Brief Illness perception sub score			
Treatment perceived more helpful (1.5 point)	-0.80	-1.70 to 0.10	0.533
Treatment perceived less helpful (6.1 points)	-0.38	-1.34 to 0.59	
Continued			

Characteristic	β	95%CI	p value
Central Mechanisms Traits			
Less central pain (7.0 points)	-0.64	-1.60 to 0.33	0.211
More central pain (14.4 points)	-0.52	-1.34 to 0.29	
<u>Examination findings (index knee)</u>			
Quadriceps strength; kg			
Weaker (12.5 kg)	-0.61	-1.55 to 0.33	0.959
Stronger (22.9 kg)	-0.58	-1.50 to 0.34	
Radiographic severity			
Total NLDA score			
Less severe (6.2 points)	-0.47	-1.37 to 0.44	0.837
More severe (22.1 points)	-0.59	-1.47 to 0.29	
NLDA osteophyte score			
Less severe (3.3 points)	-0.49	-1.39 to 0.42	0.151
More severe (16.1 points)	-0.59	-1.48 to 0.31	
NLDA JSN score			
Less narrowing (2.6 points)	-0.41	-1.36 to 0.54	0.686
More narrowing (6.5 points)	-0.67	-1.51 to 0.18	
Tibiofemoral KL grade			
0	-1.78	-3.69 to 0.13	0.102
1	-1.11	-2.49 to 0.27	
2	-0.63	-0.51 to 1.76	
3	-0.74	-1.78 to 0.30	
4	-1.99	-4.64 to 0.67	
Patellofemoral KL grade			
0	-1.50	-2.85 to -0.16	0.083
1	0.45	-0.55 to 1.44	
2	-1.05	-2.71 to 0.62	
3	-0.94	-2.16 to 0.27	
Synovial thickness; mm			
Less inflammation (2.6 mm)	-0.86	-1.70 to -0.03	0.327
More inflammation (8.4 mm)	-0.22	-1.21 to 0.78	
Effusion; mm			
Less inflammation (5.2 mm)	-0.89	-1.83 to 0.06	0.446
More inflammation (12.5 mm)	-0.41	-1.25 to 0.44	
Power Doppler			
Positive (present)	-0.65	-1.36 to 0.05	0.774
Negative (absent)	-0.37	-2.15 to 1.40	
Localised PPT (MJL); kPA			
More pain sensitivity (170.8 kPA)	-0.11	-1.01 to 0.78	0.132
Less pain sensitivity (659.3 kPA)	-1.01	-1.85 to -0.18	
Distal PPT (proximal tibia); kPA			
More pain sensitivity (181.8 kPA)	0.06	-0.85 to 0.96	0.049*
Less pain sensitivity (612.9 kPA)	-1.14	-1.98 to -0.31	
Continued			

Characteristic	β	95%CI	p value
Remote PPT (sternum); kPA			
More pain sensitivity (129.9 kPA)	-0.05	-0.93 to 0.82	0.079
Less pain sensitivity (437.5 kPA)	-1.05	-1.86 to -0.24	
Temporal summation; 0-100 NRS			
Less pain sensitivity (13.6 points)	-1.41	-2.34 to -0.48	0.020*
More pain sensitivity (62.1 points)	0.15	-0.75 to 1.05	

Sensitivity analysis including only periods where baseline pain were 4-8. Data presented as treatment coefficient (β) and associated 95%CI per predictor subgroup category. Continuous data were subgrouped as a) one SD below and one SD above the mean predictor value. Treatment coefficients <0 favour capsaicin, >0 favour ibuprofen, and =0 show no difference between the treatments. p values for the interaction coefficient are provided, with significant p values (p<0.05) indicated (*).

BMI, body mass index; HADS, Hospital Anxiety and Depression Scale; JSN, joint space narrowing; KL, Kellgren-Lawrence; KOOS, Knee Injury and Osteoarthritis Outcome Score; MJL, medial joint line; NLDA, Nottingham Line Drawing Atlas; NRS, numeric rating scale; PDQ, painDETECT questionnaire; PPT, pressure pain threshold

17c. Sensitivity analysis – MCID outcome

Characteristic	OR	95%CI	p value
<u>Basic demographics</u>			
Age			
Younger (57 years)	0.80	0.19 to 3.46	0.669
Older (76 years)	1.28	0.31 to 5.22	
Sex			
Men	0.95	0.25 to 3.63	0.901
Women	1.07	0.28 to 4.10	
BMI			
Less overweight (25.1 kg/m²)	1.09	0.29 to 4.10	0.879
More overweight (36.3 kg/m²)	0.92	0.19 to 4.43	
<u>Questionnaire – comorbidities</u>			
Anxiety; HADS			
Less features of anxiety (2.8 points)	1.80	0.39 to 8.43	0.350
More features of anxiety (10.7 points)	0.65	0.17 to 2.49	
Depression; HADS			
Less features of depression (1.8 points)	1.88	0.49 to 7.22	0.209
More features of depression (8.2 points)	0.52	0.13 to 2.19	
Fibromyalgia; fibromyalgia severity score			
Less features of fibromyalgia (5.0 points)	0.43	0.11 to 1.67	0.083
More features of fibromyalgia (13.0 points)	2.77	0.60 to 12.86	
<u>Questionnaire – OA features</u>			
Baseline pain; NRS severity			
Less pain (3.0 points)	0.61	0.14 to 2.68	0.390
More pain (7.7 points)	1.67	0.33 to 8.34	
Neuropathic-like pain; Rasch-transformed PDQ			
More nociceptive pain (-1.3 points)	0.55	0.15 to 2.10	0.196
More neuropathic pain (-0.0 points)	2.15	0.48 to 9.63	
Physical function; KOOS			
Worse function (41.3 points)	2.42	0.60 to 9.74	0.089
Better function (75.4 points)	0.39	0.09 to 1.72	
Illness perception; Brief Illness Perception score			
Less threatening view (29.8 points)	0.47	0.12 to 1.84	0.114
More threatening view (45.2 points)	2.35	0.56 to 9.78	
Treatment perception; Brief Illness perception sub score			
Treatment perceived more helpful (1.5 point)	0.77	0.21 to 2.90	0.553
Treatment perceived less helpful (6.1 points)	1.34	0.36 to 5.00	
Continued			

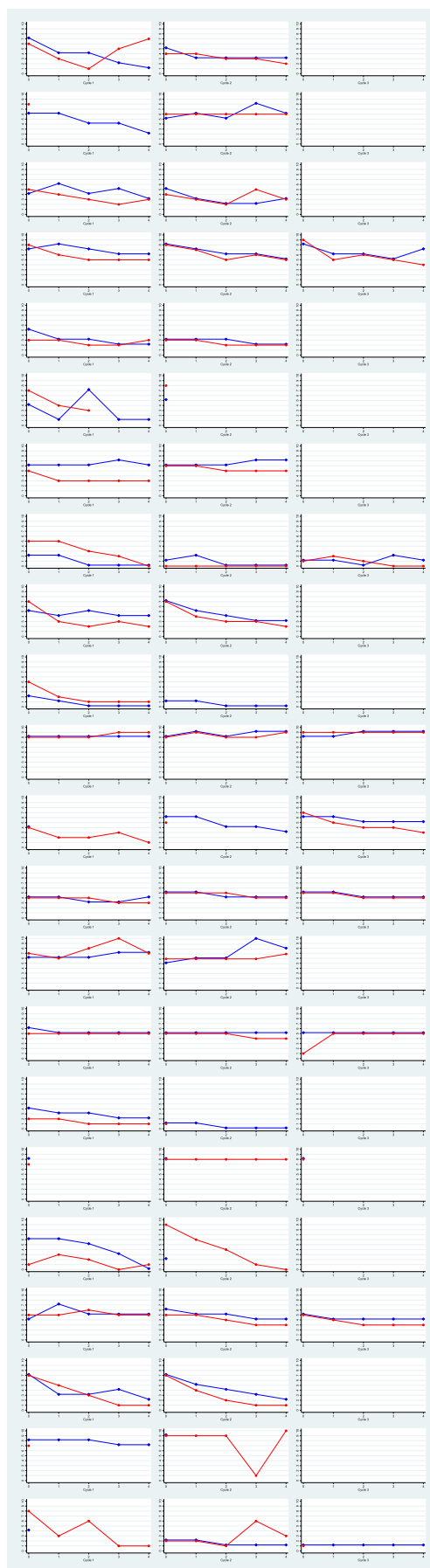
Characteristic	OR	95%CI	p value
Central Mechanisms Traits			
Less central pain (7.0 points)	1.12	0.28 to 4.47	0.845
More central pain (14.4 points)	0.92	0.23 to 3.64	
<u>Examination findings (index knee)</u>			
Quadriceps strength; kg			
Weaker (12.5 kg)	1.26	0.32 to 5.00	0.686
Stronger (22.9 kg)	0.85	0.23 to 3.15	
Radiographic severity			
Total NLDA score			
Less severe (6.2 points)	1.69	0.38 to 7.53	0.506
More severe (22.1 points)	0.81	0.19 to 3.40	
NLDA osteophyte score			
Less severe (3.3 points)	1.50	0.34 to 6.53	0.633
More severe (16.1 points)	0.89	0.21 to 3.72	
NLDA JSN score			
Less narrowing (2.6 points)	0.88	0.21 to 3.74	0.792
More narrowing (6.5 points)	1.15	0.31 to 4.25	
Tibiofemoral KL grade			
0	1.05	0.15 to 7.27	0.973
1	1.04	0.29 to 3.68	
2	1.02	0.39 to 2.64	
3	1.01	0.28 to 3.62	
4	0.99	0.14 to 6.97	
Patellofemoral KL grade			
0	2.84	0.45 to 17.96	0.253
1	1.62	0.51 to 5.18	
2	0.83	0.34 to 2.05	
3	0.53	0.10 to 2.79	
Synovial thickness; mm			
Less inflammation (2.6 mm)	3.08	0.72 to 13.14	0.042*
More inflammation (8.4 mm)	0.32	0.07 to 1.45	
Effusion; mm			
Less inflammation (5.2 mm)	1.16	0.29 to 4.56	0.794
More inflammation (12.5 mm)	0.88	0.22 to 3.62	
Power Doppler			
Positive (present)	1.04	0.38 to 2.90	0.829
Negative (absent)	0.77	0.06 to 9.72	
Localised PPT (MJL); kPA			
More pain sensitivity (170.8 kPA)	0.99	0.26 to 3.84	0.948
Less pain sensitivity (659.3 kPA)	1.06	0.27 to 4.21	
Distal PPT (proximal tibia); kPA			
More pain sensitivity (181.8 kPA)	1.11	0.30 to 4.12	0.859
Less pain sensitivity (612.9 kPA)	0.94	0.24 to 3.60	
Continued			

Characteristic	OR	95%CI	p value
Remote PPT (sternum); kPA			
More pain sensitivity (129.9 kPA)	1.39	0.37 to 5.23	0.517
Less pain sensitivity (437.5 kPA)	0.69	0.15 to 3.17	
Temporal summation; 0-100 NRS			
Less pain sensitivity (13.6 points)	1.59	0.38 to 6.73	0.423
More pain sensitivity (62.1 points)	0.70	0.18 to 2.66	

Sensitivity analysis using dichotomised pain outcome as MCID met (NRS change score ≥ 1) or MCID not met (NRS change score < 1) per period. Data presented as treatment coefficient (OR) and associated 95%CI per predictor subgroup category. Continuous data were subgrouped as a) one SD below and one SD above the mean predictor value. OR > 1 favours capsaicin and < 1 favours ibuprofen. p values for the interaction coefficient are provided, with significant p values ($p < 0.05$) indicated (*).

BMI, body mass index; HADS, Hospital Anxiety and Depression Scale; JSN, joint space narrowing; KL, Kellgren-Lawrence; KOOS, Knee Injury and Osteoarthritis Outcome Score; MJL, medial joint line; NLDA, Nottingham Line Drawing Atlas; NRS, numeric rating scale; PDQ, painDETECT questionnaire; PPT, pressure pain threshold

Appendix 18 – Weekly NRS pain scores in pilot n-of-1 trial series



Treatment responses for topical NSAID (blue) and capsaicin (red) using weekly 0-10 NRS pain intensity scores. Each participant is represented on one row. Cycles 1, 2, and 3 are presented in the first, second, and third column

Appendix 19 – GRADE assessment: Capsaicin in NP

GRADE criteria	Rating	Footnotes	Quality of evidence
Study design	High	RCTs	⊕○○○ Very low
Risk of Bias	Serious (-1)	Majority of trials likely did not successfully blind capsaicin or adequately describe allocation concealment. No trials report analysing all data. Incomplete outcome reporting is a serious concern. Sensitivity analyses using only high quality trials changed the conclusions	
Inconsistency	Serious (-1)	ES vary from -2.32 to +0.87 in the trials. Confidence intervals do not overlap. High levels of heterogeneity evident ($I^2=91\%$)	
Indirectness	No	PICO match MA aims. Outcomes were assessed using different approaches (not all continuous), but are likely to be measuring the same underlying effect	
Imprecision	Serious (-1)	Estimate is imprecise – at its lowest bound it indicates that placebo is better than capsaicin (meeting MCID). At its highest bound it indicates that capsaicin is better, and clinically significant	
Publication bias	Serious (-1)	There is some asymmetry in the funnel plot. All evidence is from small studies (predominantly industry-sponsored), therefore increasing the likelihood of publication bias	

Appendix 20 – GRADE assessment: Overall effect of capsaicin in NP

GRADE criteria	Rating	Footnotes	Quality of evidence
Study design	Low	Observational	⊕○○○ Very low
Risk of Bias	Serious (-1)	Incomplete outcome reporting is a serious concern – overall ES was only calculated for one trial as the majority did not report measures to calculate it. Participants were likely not adequately blinded	
Inconsistency	-	Not assessable	
Indirectness	Serious (-1)	Analysis is based only on liposomal 0.025% capsaicin, applied two to three times per day. This is not available in the UK and may therefore not be applicable	
Imprecision	Serious (-2)	Only 14 participants (in cross over trial) were analysed. Greatly underpowered. Interpretation of ES varies between the bounds of the confidence interval (from marginal effect to meeting MCID)	
Publication bias	Serious (-1)	Only small trial available for analysis. Additional trials have been identified but not included in analyses. The analysed trial is likely not representative of the body of evidence available	