BMJ Open

Defining Acute Flares in Knee Osteoarthritis: A Systematic Review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019804
Article Type:	Research
Date Submitted by the Author:	27-Sep-2017
Complete List of Authors:	Parry, Emma; Keele University, Research Institute for Primary Care and Health Sciences Thomas, M; Keele University, Research Institute for Primary Care and Health Sciences Peat, George; Keele University, Research Institute for Primary Care & Health Sciences
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	General practice / Family practice
Keywords:	Osteoarthritis, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Flare

DEFINING ACUTE FLARES IN KNEE OSTEOARTHRITIS: A SYSTEMATIC REVIEW

Emma L. Parry[1] , e.parry@keele.ac.uk

Martin J. Thomas[1], m.thomas@keele.ac.uk

George Peat[1*], g.m.peat@keele.ac.uk

[1] *Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, UK*

*Correspondence: Professor George Peat, Professor of Clinial Epidemiology, Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG. Tel: +44 (0) 1782 732929. Fax:+44 (0) 1782 734719. Email: g.m.peat @keele.ac.uk

Word count: 3722

Keywords: knee osteoarthritis, flare, systematic review

ABSTRACT

Objective: To identify and critically synthesise definitions of acute flares in knee osteoarthritis (OA) reported in the medical literature.

Design: Systematic review and narrative synthesis. We searched MEDLINE, EMBASE, Web of science and 6 other electronic databases (inception to July 2017) for original articles and conference abstracts reporting a definition of acute flare (or synonym) in humans with knee OA. There were no restrictions by language or study design (apart from iatrogenic induced flare-ups e.g. injection-induced). Data extraction comprised: definition, pain scale used, flare duration or withdrawal period, associated symptoms, definition rationale, terminology (e.g. exacerbation or flare), baseline OA severity, age, gender, sample size and study design.

Results: Sixty-nine articles were included (46 flare-design trials, 17 observational studies, 6 other designs; sample sizes: 15-6085). Domains used to define flares included: worsening of signs and symptoms (61 studies, 27 different measurement tools), specifically increased pain intensity; minimum pain threshold at baseline (44 studies); minimum duration (7 studies, range 8-48 hours); speed of onset (2 studies, defined as 'sudden' or 'quick'); requirement for increased medication (2 studies). No definitions included activity interference.

Conclusions: The concept of OA flare appears in the medical literature but most often in the context of flare design trials. Key domains, used to define acute events in other chronic conditions, appear relevant to OA flare and could provide the basis for

consensus on a single, agreed definition of 'naturally occurring' OA flares for research and clinical application.

PROSPERO registration: CRD42014010169

Strengths and limitations of this study

Strengths

- Identified key domains that are used to define acute events by undertaking a comprehensive synthesis of definitions used in the medical literature.
- Broad search strategy covering a wide range of databases including bibliography checks and conference abstracts.
- Prospectively registered with Prospero

Limitations

- Did not search grey literature
- Did not include potential synonyms as search terms ('attack', 'episode', 'fluctuations')

INTRODUCTION

Recurrent acute events or episodes feature in the natural history of many chronic health conditions. The extent to which they characterise the condition varies, as do the presumed pathophysiological mechanisms, and scientific and lay terms used to describe them (e.g. an acute exacerbation of chronic obstructive pulmonary disease (COPD) or asthma, an attack of gout or a rheumatoid arthritis flare). With recognition of their importance has come concerted effort to define these phenomena. Definitions for exacerbations or flares currently exist for COPD[1, 2] , asthma[3], systemic lupus erythematosus (SLE)[4], and ankylosing spondylitis (AS)[5] and there are working groups currently trying to define these for rheumatoid arthritis[6-8], gout[9], and atopic dermatitis/eczema[10]. Despite the different language used, these definitions share some common, core domains: the onset or worsening of symptoms and signs above normal day-to-day variability; speed of onset; duration of sustained worsening; and change in medication/healthcare usage.

Osteoarthritis (OA) appears to comprise multiple disease trajectories[11-15] and symptom variability over time and the presence of intermittent pain is well-recognised[16]. Although OA does not typically have the same very obvious acute events as conditions like gout, flares in OA joints are encountered in practice, these

phenomena appear in patient literature[17], have been discussed in expert reviews[18], and are mentioned in 'flare design' trials in OA[19]**.** These studies invoke acute episodes of pain or flare-ups by asking patients to withdraw their usual medication.

In 2009 Marty et al proposed scoring criteria for knee OA flares based on nocturnal awakening, knee effusion, morning stiffness and limping[20] but it is unclear whether this has contributed to a common understanding, shared terminology and criteria. A common definition of OA flare could be important for a number of reasons; (i) to facilitate communication between researchers, (ii) to allow more direct comparisons between studies on frequencies, determinants and course of events, (iii) to facilitate new insights into novel pathophysiological mechanisms and treatments through valid and homogenous case definitions, and (iv) to help clinicians with prompt diagnosis and management.

The aim of this systematic review was to explore the extent to which a concept of OA flare is reported in the medical literature and the prospects for a common, shared definition of these for research and clinical application.

94

95

96

97

98

METHODS

This systematic review was registered with PROSPERO registration number CRD42014010169. The review protocol has not been published.

Literature sources and study selection

We searched electronic databases from inception to July 2017; ASSIA, EMBASE, Web of Science, Health Management Information Consortium (HMIC), SPORTDiscus, Medline, CINAHL, PsycINFO, AMED, Ageline, Cochrane Database of Systematic Reviews and Cochrane Controlled Clinical Trials (CENTRAL). The search was developed using previously piloted terms for knee OA and a literature search for common terms used to describe acute events. Searches used combined and/or truncated key terms including: ("KNEE OSTEOARTHRITIS" OR (knee N3 pain) OR (knee N3 arthrosis) OR (knee N3 joint) OR (knee N3 osteoarthritis)) AND (exacerbation OR flare OR (pain AND (diary OR diaries)) OR (pain N3 variab*) OR (pain N3 pattern$) OR (daily N3 pain)). A database search strategy is included in the online supplement . Reference lists of all included full text articles retrieved for detailed examination were manually searched.

Studies were included in the final full text peer review if they contained a description or definition, with or without classification criteria based on measurement, of an acute exacerbation or flare-up of knee OA in human adults (18 years or over) in the general population, primary care or hospital settings. There were no restrictions on study dates or design. All non-English language articles were translated to identify a flare definition. Theses, dissertations, book chapters and guidelines, and animal studies were excluded. Conference abstracts were included if they contained a definition for an OA flare-up. Studies were excluded if the flare was induced by an iatrogenic source, for example, injection-induced flares[21]. As these may have been caused by a different pathophysiological process. Abstracts were included in this study as the main outcome of interest was the definition of flare used and it was decided that including abstracts would ensure a more comprehensive review. For each abstract a search was conducted to identify a corresponding full text paper. Where one was found only the full paper was included in the review.

Data collection

The search and article retrieval was conducted by the first reviewer (ELP). Articles were downloaded into RefWorks© bibliography and database manager (RefWorks Copyright 2009). Duplicates were removed and all titles were screened by ELP

against inclusion criteria. All titles were screened by the first reviewer to meet inclusion criteria. The first 20 titles were checked by two reviewers (ELP and MJT) to check consistency. For qualitative studies, all identified potentially eligible full text articles were obtained as title and abstract searches did not always provide the full information about the article content.

All abstracts were screened by two reviewers (ELP and MJT). Potentially eligible full texts were then screened by the same two reviewers to identify articles to be included in data extraction. Where there was disagreement a third reviewer (GP) acted as adjudicator and following discussion agreement was reached by consensus. Where articles could not be retrieved or if the flare definition used was not included in the text, contact with authors was made.

The final included articles were screened to ensure results from the same studies were not counted as separate studies as this is known to introduce bias as the dataset would more strongly affect the results of the review[22]. For articles containing pooled studies, the original studies were sought and included in the main analysis, where available. If the original articles were not referenced or not available the pooled studies were kept and a note made of this in the analysis. No full text articles were required to be translated.

Data extraction

Information was extracted by the first reviewer (ELP), and recorded in a purpose-built Excel spreadsheet. For all of the information extracted every tenth article was independently checked by a second reviewer (MJT).

The following data pertaining to flares were extracted from full text articles: definition used for change in pain, pain scale used, duration of flare or withdrawal period, associated symptoms, rationale behind definition used, terminology used (e.g. exacerbation or flare), baseline OA severity, age range, gender, geographical location, number of participants and study design. Missing data was described in the data extraction tables.

Quality assessment of included studies

Our aim was to identify and contrast definitions of flare-ups used in the literature. We were not concerned with the methodological rigour of the studies deriving, evaluating or applying those definitions. However, for studies presenting definitions we sought supporting statements that gave the rationale for the definition.

Data analysis

A narrative synthesis was undertaken, guided by Popay et al's[23] four stage process to develop a conceptual framework[24]. This approach was chosen as it allowed the words and text in the definitions to be synthesised to summarise findings[23]. The initial data extracted was grouped into drug withdrawal studies ('flare design') and other studies, and frequencies of components included in definitions was tabulated, these included; terminology used, onset/worsening of symptoms; signs/symptoms above day-to-day variability/minimum threshold; speed of onset of symptoms; duration of worsening and change in medication/healthcare usage.

This initial tabulation helped identify similarities and differences and allowed themes to emerge. This was done with an inductive type approach, where possible i.e. without an *a priori* assumption, but also deductively acknowledging that the reviewers were clinicians i.e. they had some background knowledge of the topic of interest. This allowed further examination of the differences of definitions used in drug withdrawal and non-drug withdrawal study designs, and examination of key components of definitions used.

RESULTS

Study selection

The literature search yielded 2194 articles of which 786 were duplicates (Figure 1). After title screening 336 abstracts were reviewed, 223 were not relevant for the study purpose. 113 articles were examined in full which resulted in a further 60 being excluded. The main reason for exclusion was no definition of flare-up reported in text (n=56). At this stage a further 16 articles were identified from the reference lists of the retrieved full text articles resulting in 69 included studies for synthesis.

Study characteristics

Characteristics of the included studies are described in Table 1[20, 25-92]**.** Studies ranged in size from 15-6085[20, 49] and location. Knee OA was defined by clinical and/or radiological criteria.

215 **Table 1: Characteristics of all included studies**

DRUG WITHDRAWAL DESIGN STUDIES					
First author, year of publication	**Setting, geographic location**	**Participants**	**Joint**	**Severity**	**Study design**
Altman, 2015[25]	Multi-centre, recruitment not specified, USA	403 males & females, ≥40y	Knee and hip	KL grade 2-3	RCT, flare design
Baer, 2005[27]	17 medical centres recruiting from community and physician private practice; Canada	216 males & females, 40-85y	Knee	Radiographic evidence of OA (severity not defined)	RCT, flare design
Baraf, 2011[28]	Primary care, internal medicine, orthopaedic, rheumatology; USA	602 males & females, ≥25y	Knee	Radiographically mild to moderate (KL grade 1-3)	RCT, flare design
Battisti, 2004[29]	Clinical centres, out patients; USA	3980 males & females, ≥40y (age unavailable for Geba 2003 and Weaver 2003)	Knee	ACR functional class rating of I,II or III	RCT, pooled 4 trials, flare design
Bingham, 2007[30] Bingham 2011[76]	2x74 outpatient clinics; USA	1207 males & females, ≥40y	Knee and hip	ARA Functional capacity classification I-III	RCT, flare design

Birbara, 2006[31]	Investigative sites; USA	808 males & females, ≥40y	Knee	ARA functional class, I, II, or III	RCT, flare design
Bocanegra, 1998[32]	Clinic; USA	572 males & females, 28-88y (mean 61-62)	Knee and hip	ARA Functional capacity classification I-III	RCT, flare design
Boswell, 2008[33]	50 centres (Europe & Australia) + 187 centres (Europe & USA)	1908 males & females, ≥40y	Knee	KL scale 2 or 3 and ARA class rating of I,II or III	Pooled RCTs (2; one flare design, one non-flare), flare design
Brandt, 2006[34] (pilot studies)	Community; USA	30 males & females, mean age 62y	Knee	KL ≥2	Cohort design, flare design
Case, 2003[35]	Hospital-rheumatology centre; Chicago, USA	82 males & females, 40-75y	Knee	KL ≥1, and clinical criteria (pre-enrolment ambulatory pain; moderate pain by a 5-point Likert scale or increased pain.	RCT, flare design
Day, 2000[74]	49 investigative sites in 26 countries	809 males & females, mean age range 62-65y	Knee and hip	ARA functional class I-III, symptomatic for at least 6 months	RCT, flare design
Ehrich, 1999[36]	Clinical centres; USA	219 males & females, >40y	Knee	ARA functional class, I, II, or III	RCT, flare design
Essex, 2012[37]	Clinical centre; African-American, USA	322 males & females, ≥45y	Knee	ARA Functional capacity classification I-III	RCT, flare design
Essex 2013[77]	Hispanic population, 31 US centers	≥45y	Knee	ACR criteria, Functional capacity classification I-III	RCT, flare design
Gibofsky, 2014[38]	Not specified, USA	305 males & females, 41-90 y	Knee and hip	KL 2-3	RCT, flare design
Gineyts, 2004[39]	Subset of larger study; France	201 males & females, mean age 61-62y	Knee and hip	ARA I-III	RCT, flare design

Goldberg, 1988[40]	Investigative sites; USA	214 males & females, 40-85y (mean 64)	Knee and hip	Radiographic evidence of knee OA-not further defined	RCT, flare design
Gottesdiener, 2002[41]	Investigative sites; USA	617 males & females, ≥40y	Knee	ARA functional class I,II,III	RCT, flare design
Hochberg, 2011[42]	Centres; USA	1234 males & females, ≥50y	Knee	ACR functional class I-III	Pooled RCTs (2), flare design
Katz, 2010[43]	Clinical sites; USA	113 males & females, 28-83y (median 57))	Knee and hip	OA of hip and knee as diagnosed using ACR criteria-no definition of severity	RCT, flare design
Kivitz, 2001[44]	Investigative sites; USA	491 males & females, 28-91y (mean 58-61)	Knee	Confirmation of OA on weight bearing radiograph- no definition of severity	RCT, flare design
Kivitz, 2004[75]	Outpatient sites; USA	1042 males & females, ≥40y	Knee	ACR rating of I, II, III.	RCT, flare design
Leung, 2002[46]	Clinic; USA	677 males & females, ≥40y	Knee and hip	ARA functional class, I, II, or III	RCT, flare design
Luyten, 2007[46]	Centres; Belgium	181 males & females, ≥40y	Knee and hip	ACR Functional capacity classification I-III	RCT, flare design
Manicourt, 2005[48]	Outpatient clinic; Belgium	90 males & females, 50-81y (mean 63-67)	Knee and hip	Clinical and radiographic evidence of OA-severity not defined.	RCT, flare design
Mazzuca, 2002[49]	Not specified, USA	15 males & females, ≥45y	Knee	KL 2-3	Observational, flare design
McIlwain, 1989[50]	Investigative sites; USA	139 males & females, mean 65y	Knee	Radiological evidence of moderate or severe osteoarthritis- not further defined	RCT, flare design
Mendelsohn, 1991[51]	Investigative sites; USA	139 males & females, 21-88y (mean age 63.3y)	Knee	Radiological evidence of moderate or severe osteoarthritis- not further defined	RCT, flare design

Moskowitz, 2006[52]	Investigative sites; USA	530 males & females, ≥45y	Knee	ACR Functional capacity classification I-III	RCT, flare design
Pareek, 2009[53]	Multi-centre study, India	199 males & females, 40-70y	Knee	Lequesne criteria-score of 5 and above	RCT, flare design
Pareek, 2010[54]	Hospital; India	220 males & females, 40-70y	Knee	Clinical and radiological evidence of OA- severity not defined.	RCT, flare design
Roth, 2004[89]	Physicians private practice or community; USA	326 males & females, 40-85y	Knee	Radiological evidence of OA- severity not defined.	RCT, flare design
Rother, 2007[92]	Outpatient units; Germany	397 males & females, ≥40y	Knee	KL 2-3	RCT, flare design
Schnitzer, 2005[56]	Investigative sites; International (7 countries)	583 males & females, 18-75y	Knee and hip	Diagnosis based on ACR criteria- severity not defined.	RCT, flare design
Scott-Lennox, 2001[57]	Investigative sites; USA	182 males & females, mean 61y	Knee	Not defined	RCT, flare design
Silverfield, 2002[58]	Centres; USA	308 males & females, 35-75y	Knee and hip	Clinical evidence of OA- severity not defined	RCT, flare design
Simon, 2009[90]	Outpatient centres; Canada, USA	775 males & females, 40-85y	Knee	Clinical and radiological evidence of OA- severity not defined	RCT, flare design

Strand, 2011[59]	Investigative sites; Multinational-not specified including USA	875 males & females, 18-80y	Knee and hip	OA according to ACR criteria and requiring NSAID treatment to control symptoms in the month preceding screening	RCT, flare design
Weaver, 1995[91]	Investigative sites; USA	328 males & females, >50y	Knee	ACR clinical criteria-diagnostic	RCT, flare design
Wiesenhutter, 2005[60]	Medical Centres; USA	528 males & females, 40-89y	Knee and hip	ARA functional class, I, II, or III	RCT, flare design
Williams, 2001[61]	Clinical sites; USA	718 males & females, mean 61-62y	Knee	ACR clinical and radiographic criteria I-III	RCT, flare design
Wittenberg, 2006[62]	Centres (not specified) ; Germany	364 males & females, 50y	Knee	Moderate to severe symptomatic OA of the knee according to ACR criteria.	RCT, flare design
Yeasted, 2014[63] (Pooled, abstract)	USA	219 (merged observational), 137 (merged trial)>40y	Not specified	ACR criteria-diagnostic	2 longitudinal observational studies, placebo arms of 2 clinical trials
Yocum, 2000[78]	USA, 62 study centres	774 males & females, ≥40y	Knee or hip	Diagnosis confirmed by XR and clinical symptoms (not further specified)	RCT, flare design
Young, 2014[64] (abstract)	Multicenter,	305 males & females, >40y	Knee or hip	KL 2-3	RCT, flare design
Zhao, 1999[65]	Centre (not specified); USA, Canada	1004 males & females, ≥18y	Knee	ACR Functional capacity classification I-III	RCT, flare design
NON-DRUG WITHDRAWAL DESIGN STUDIES					

Atukorala, 2016[79] (abstract) Atukorala, 2016[26] (abstract)	Not specified, USA + Au + Sri Lanka	213 males & females, mean age 62y 345 males & females, mean age 62y	Knee	Not specified	3-month, web based longitudinal follow up study
Bartholdy, 2016[80]	OA out-patient clinic, Denmark	131 males & females, ≥40y	Knee	Radiographic evidence of OA (severity no defined) and BMI between 20-35 kg/m^2	RCT
Bassiouni 2015[81] (abstract)	Not specified, Egypt	60 participants not further specified	Knee	Not specified	Observational
Cibere, 2004[87] Cibere, 2005[88]	Community, Canada	137 males & females, mean age 65y (43-88) for placebo and 64y (40-83) for glucosamine group	Knee	KL ≥2 on anteroposterior radiograph	RCT
Conrozier 2012[67]	Hospital-rheumatology unit, France	44 males & females, mean age 67.6y	Knee	Radiographic evidence of knee OA-not further defined	Observational
D'Agostino 2005[68]	Hospital-European multicentre	600 males & females, ≥18y	Knee	KL grade 1-4	Observational
Erfani, 2014[45] abstract) Erfani, 2014[82] (abstract) Ferreira[83] 2016	Au	268 males & females, mean age 62y 345 males & females, ≥40y	Knee	ACR criteria- meet at least one, KL ≥2	Web based cross over

Hunter 2014[84] (abstract) Makovey 2015[85] (Protocol)					
Jawad, 2005[69]	GPs in France	3000 (for GP study) males & females	Knee	Not defined	n/a, review of surveys. Definition relates to survey of 3000 French GPs
Marty 2009[20]	Community and hospital, France	6085+641males & females, mean age 66.4y (10.9) for flare group, 66.2y (10.2) no flare group	Knee	OA diagnosis based on ACR criteria- severity not defined	Observational
Murphy, 2015[70]	Community based, pain clinics; USA	45 males & females, 37-83y	Knee	ACR criteria- severity not defined	Qualitative
Parry, 2017[86]	Community, UK	719 males & females, ≥50y	Knee	Self-reported knee pain in previous 12 months	Observational
Ricci 2005[55]	Community, USA	329 males & females, 40-65y	Knee and hip	Clinical evidence of OA- severity not defined	Nested case control
Wise 2010[71]	Primary care, hospital, USA	303 males & females, ≥50y	Knee and hip	Signal joint pain in a hip or knee on at least 15 out of the 30 days prior to enrolment- not further defined	Observational

Zhang 2009[72]	Primary care, hospital, USA	303 males & females, ≥50y	Knee and hip	Signal joint pain in a hip or knee on at least 15 out of the 30 days prior to enrolment-not further defined	Observational
Zhang 2011[73] (abstract)	Not specified	52 males & females, median age 63, (50-72y)	Knee	KL>2	Case-crossover
Zobel, 2016[93]	Hospital databases, Australia	297 males & females, >40y	Knee	ACR criteria, KL ≥2, or patellofemoral OA on radiograph	Web based case-cross over

Acronyms:
KL- Kellgran and Lawrence
RCT- Randomised Controlled Trial
USA- United States of America
Au- Australia
ACR- Arthritis Center Research
ARA- American Rheumatism Association
GP- General Practitioner

Twenty-one included mixed knee and hip OA groups[25, 30, 32, 38-40, 43, 46-48, 55, 56, 58-60, 64, 72, 74, 76, 78]. In total, 46 publications used a drug withdrawal RCT design[25, 27-33, 35-44, 46-54, 56-65, 74-78, 89-92], four of which were pooled studies[29, 33, 42, 63] and one used a cohort drug withdrawal design[34] (Table 1). The remaining 22 publications included seventeen observational studies[20, 26, 45, 55, 66-68, 71-73, 79, 81-86], three RCTs[80, 87, 88], one survey[69] and one qualitative interview study[70]. Nine of the included studies were abstracts[26, 45, 63, 64, 73, 79, 81, 82, 84]. Two abstracts were removed as the corresponding full text article was available[70, 93]. Studies using pooled data or the same dataset were included if they used different definitions of OA flare[29, 45, 53, 54, 63, 66, 71, 72, 75].

Rationale given for flare definitions

Six of the included studies gave rationale for the definition used[20, 55, 57, 70, 86, 87]. Cibere[87] outlined face validity checks. It was specified that the flare definition had been determined by study rheumatologists to be a clinically important change in the WOMAC score. The definition used by Murphy et al[70] was informed by two studies[29, 54] which used a drug withdrawal design and from the research team's own experience. Ricci et al[55] used a combination of data-driven and clinical judgement approaches to establish an agreed cut point. Parry et al based their

definition on OA flare design studies and flare definitions used in other chronic disease such as back pain and COPD.

Marty et al[20] and Scott-Lennox et al[57] were the only studies that undertook empirical investigation of flare definitions.

The study by Marty et al[20] was the only study specifically designed to validate a set of predetermined flare criteria, which they did using logistic regression analysis to assign a weight to each of the items identified. A flare up score was determined using a general practitioner database and this was then validated using a rheumatologist database. Pain was not included in the final model.

Scott-Lennox et al[57] sought to test whether four measures for flare intensity could be combined to form a reliable and valid index using data from an RCT using a confirmatory factor analysis.

Flare definitions in drug withdrawal studies

Terminology used

The majority of publications using a drug withdrawal design used the term “flare” in their description[25-31, 33, 34, 37-44, 46-50, 52, 54, 56-65, 75-78, 89-92] (n=42; Table 2).

Table 2: Definition, terminology and measurement instruments used in all included studies

DRUG WITHDRAWAL STUDY DESIGN							
First author	**Terminology used**	**Onset/worsening of symptoms/signs (symptom/sign: measurement instrument; operational definition)**	**Minimum threshold (symptom/sign: measurement instrument; operational definition)**	**Speed of onset**	**Duration**	**Change in medication/healthcare use**	**Reference / rationale**
Altman, 2015[25]	"Flare"	**Pain**: WOMAC Pain subscale (0-100); increase ≥15mm	**Pain:** WOMAC Pain subscale; ≥40mm	Not specified	Not specified	Not specified	None
Baer, 2005[27]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale (0-20); increase ≥2 points and ≥25%	**Pain:** WOMAC Pain score (0-20); ≥6 and ≥1 item rated 'moderate, severe, or extreme'	Interval between screening and baseline re-measurement unclear	Not specified	Not specified	None
Baraf, 2011[28]	"Flare"	**Pain on movement:** VAS (0-100mm); increase ≥5mm	Not specified	1 week washout	Not specified	Not specified	None
Battisti, 2004[29]	"Flare"	**Global assessment (investigator):** single item, 5-point LK; Worsening ≥1 point	**Pain:** VAS (0-100mm); ≥40mm	Not specified	Not specified	Not specified	None

Bingham, 2007[30] Bingham 2011[76]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); increase ≥15mm (2) **Global assessment of disease status (investigator):** single item, 5-point LK; Worsening ≥1 point	(1) **Pain walking on flat surface:** ≥40mm on WOMAC VAS3.0 Q1 (0-100) (2) **Global assessment (investigator)**: single item, 5-point LK; fair, poor, very poor (*acetaminophen users only*) (3) **Global assessment of disease status (patient):** VAS 0-100mm; ≥40mm (*acetaminophen users only*)	Not specified	Not specified	Not specified	None
Birbara, 2006[31]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global assessment (investigator):** single item, 5-point LK; Worsening ≥1 point	(1) **Pain walking on flat surface:** WOMAC VAS3.0 Q1 (0-100); ≥40mm (2) **Global assessment (investigator):** single item, 5-point LK; Fair, poor or very poor (*paracetamol arm only*)	4-15 day washout	Not specified	Not specified	None
Bocanegra, 1998[32]	"Worsening of symptoms"	Two out of the following three: (1) **Global assessment (physician):** single item, 5-point LK; Increase ≥1 grade (2) **Global assessment (patient):** Patients global assessment (current symptoms and limitation of activity) 5-point LK; Increase ≥1 grade (3) **Composite index:** Lequesne OA Severity Index (0-24); Increase ≥2 points	(1) **Global assessment (physician):** single item, 5-point LK; 'poor/very poor' (2) **Global assessment (patient):** Patients global assessment (current symptoms and limitation of activity) 5-point LK; 'poor/very poor' (3) **Composite index:** Lequesne OA Severity Index (0-24); ≥7	3-14d washout	Not specified	Not specified	None

Boswell, 2008[33]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global assessment (patient):** Patient Global Assessment of Arthritis Condition (PGAC) (unspecified); Worsening ≥1 point	Not specified	Not specified	Not specified	Not specified	None
Brandt, 2006[34] (pilot studies)	"Flare"	Not specified	**Pain:** WOMAC LK Pain subscale (5-25); ≥15 points	5 half-lives of NSAID washout	Not specified	Not specified	None
Case, 2003[35]	Not used	(1) **Pain walking on flat surface:** VAS (0-100mm); Increase ≥10mm (2) **Ambulatory pain**; 5-point LK; worsening ≥1 point	Not specified	14d washout	Not specified	Not specified	None
Day, 2000[74]	Not used	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global Assessment (investigator):** single item, 5-point LK; worsening ≥1 point (3) **Global assessment (patient):** VAS (0-100mm); increase ≥15mm (*acetaminophen users only*)	(1) **Pain walking on a flat surface**: WOMAC VAS Q1 (0-100mm); ≥40mm; (2) **Global Assessment (investigator):** single item, 5-point LK; 'Fair, poor, or very poor'; (3) **Global assessment (patient):** VAS (0-100mm); ≥40mm	Longer than 5 plasma half-lives washout	Not specified	Not specified	None
Ehrich, 1999[36]	Not used	**Pain:** VAS (0-100mm); increase ≥15mm	**Pain:** VAS (0-100mm); ≥40mm	Longer than 5 plasma half-lives washout of NSAID	Not specified	Not specified	None

Essex, 2012[37]	"Flare"	(1) **Global Assessment (Physician):** 5-point LK; increase ≥1 grade (2) **Global Assessment (patient):** 5-point LK; increase ≥1 grade	(1) **Global Assessment (Physician):** 5-point LK; 'Fair, poor or very poor' (2) **Global Assessment (patient):** 5-point LK; 'Fair, poor or very poor' (3) **Pain**: VAS (0-100mm); 40-90mm	48 hour withdrawal	Not specified	Not specified	None
Essex 2013[77]	"Flare"	Not specified	(1) **Global Assessment of arthritis (Physician):** Minimum rating of 3 (2) **Global Assessment of arthritis (patient):** Minimum rating of 3 (3) **Pain**: VAS (0-100mm); 40-90mm	48 hour withdrawal	Not specified	Not specified	None
Gibofksy, 2014[38]	"Flare"	**Pain:** WOMAC Pain VAS; increase ≥15mm	**Pain:** WOMAC Pain VAS; ≥40mm	Not specified	Not specified	Not specified	None
Gineyts, 2004[39]	"Flare"	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2)**Global Assessment (investigator):** 5-point scale: worsening ≥1 point	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); ≥40mm	5 half-lives of NSAID washout	Not specified	Not specified	None
Goldberg, 1988[40]	"Flare"	(1) **Pain:** Investigator assessed pain grade (None/mild/mod/severe): (i) at rest, (ii) on passive motion, (iii) on palpation, (iv) weight bearing; increase ≥1 grade in two items OR increase ≥2 grade in one item	Not specified	2-14 day washout until flare	Not specified	Not specified	None
Gottesdiener, 2002[41]	"Flare"	(1) **Pain on walking**: VAS (0-100mm); increase ≥15mm (2)**Global Assessment (Investigator):** 5-point LK; Increase ≥1 point	(1) **Pain on walking**: VAS (0-100mm); ≥40mm	3-15 day washout	Not specified	Not specified	None

Hochberg, 2011[42]	"Flare"	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm)**;** Increase ≥15mm (2) **Global Assessment (patient):** 5-point LK; worsening ≥1 point	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm)**;** ≥40mm	Not specified	Not specified	Not specified	None
Katz, 2010[43]	"Flare"	Not specified	**Pain**: Pain score (0-10); ≥5	Not specified-washout until flare occurred	Not specified	Not specified	None
Kivitz, 2001[44]	"Flare"	**Pain:** Patients Assessment of Pain Score (0-10) (unspecified); increase ≥2 points	**Pain:** Patients Assessment of Pain Score (0-10) (unspecified); ≥5	5 drug half-lives or 48 hours	Not specified	Not specified	None
Kivitz, 2004[75]	"Flare"	(1) **Pain on walking:** VAS (0-100mm); worsening ≥15mm (2) **Global Assessment (investigator):** 5-point LK; worsening ≥1 point	Not specified	NSAID dependent half-life washout	Not specified	Not specified	None
Leung, 2002[46]	"Flare"	(1) **Pain on walking on a flat surface:** WOMAC VAS Q1 (0-100mm); Increase ≥15mm (2) **Global Assessment (investigator):** 5-point LK; worsening ≥1 point	(1)**Pain on walking on a flat surface:** WOMAC VAS Q1 (0-100mm); ≥40mm (2) **Global Assessment (patient):** **(**0-100mm); ≥40mm (*acetaminophen users only*) (3) **Global Assessment (investigator):** 5-point LK; 'Fair, poor, or very poor' (*acetaminophen users only*)	Determined by drug half-life washout	Not specified	Not specified	None

Luyten, 2007[47]	"Flare"	(1) **Global Assessment (Patient):** 5-point LK; Increase ≥1 grade (2) **Global Assessment (physician):** 5-point LK; increase ≥1 grade (3) **Composite definition:** Lequesne Osteoarthritis Severity Index (0-24); increase ≥2 points	(1) **Global Assessment (Patient):** 5-point LK; 'Fair, poor or very poor' (*Not on treatment – 'Poor or very poor'*) (2) **Global Assessment (physician):** 5-point LK; 'Fair, poor or very poor' (*Not on treatment – 'Poor or very poor'*) (3) **Composite definition:** Lequesne Osteoarthritis Severity Index (0-24); ≥7 (4) **Pain**: VAS (0-100mm); ≥40mm	2-14 day washout	Not specified	Not specified	None
Manicourt, 2005[48]	"Flare"	**Pain when walking on a flat surface:** VAS (0-100mm) ; ≥10mm	Not specified	7-10 days washout	Not specified	Not specified	None
Mazzuca, 2002[49]	"Flare"	**Pain on standing**: WOMAC LK Pain Q5 'severe or extreme' after the washout AND decreased after resumption of usual analgesic drugs and/or NSAIDs	Not specified	Drug washout 5 half lives	Not specified	Not specified	None
McIlwain, 1989[50]	"Flare"	No measurement instrument: Increase in pain on motion, swelling, tenderness, redness and/or heat (unspecified if patient/physician/investigator reported)	Not specified	2-14 day washout	Not specified	Not specified	None
Mendelsohn, 1991[51]	"Worsening of arthritis condition"	(1) **Pain:** Pain scale (0-3) (0=none, 3=severe); worsening score (2) **Global (physician):** (0-100); worsening score	Not specified	Up to 14 days washout	Not specified	Not specified	None

Moskowitz, 2006[52]	"Flare"	(1) **Global assessment (patient)**: 5-point LK; increase ≥1 grade (2) **Global Assessment (physician)**: 5-point LK; ≥ 1 grade increase (3) **Composite index:** Lequesne OA Severity Index (0-24); increase ≥2 points	(1) **Global assessment (patient):** 5-point LK; '(Fair), poor, or very poor' (2) **Global Assessment (physician):** 5-point LK; '(Fair), poor or very poor' (3) **Composite index**: Lequesne OA Severity Index (0-24); Minimum ≥7 (4) **Pain walking on a flat surface:** VAS (0-100mm); ≥40mm	NSAID washout of 5 half-lives or at least 2 days	Not specified	Not specified	None
Pareek, 2009[53]	"Flare-up"	(1) **Pain:** 11-point NRS; increase ≥ 2 points during previous 2-5 days (2) Signs and symptoms suggestive of inflammation, morning stiffness and nocturnal pain interfering with sleep	**Pain**: Pain intensity of at least 4 on a 11-point NRS during physical activity for past 24 hours	Placebo washout for 24-48 hours	2-5 days	Not specified	None
Pareek, 2010[54]	"Flare"	Flare symptoms noted but not part of definition: morning stiffness, erythema, nocturnal pain, and swelling/inflammation	(1) **Pain with physical activity**: VAS 0-10; ≥6 (2) **Composite index:** WOMAC Total LK; ≥25. (3) **Composite index**: Lequesne OA Severity Index (0-24); ≥5	Not specified	2-5 days	Not specified	None
Roth, 2004[89]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale (0-20); increase ≥2 points and ≥25%	**Pain:** WOMAC LK3.1 Pain subscale (0-20); Score ≥'moderate' on at least 1 of the 5 items, (ii) Pain score ≥6	Washout period of at least 3 days per week past month	Not specified	Not specified	None

Rother, 2007[92]	"Flare"	(1) **Pain on walking:** VAS (0-100mm); Increase ≥15mm (2) **Global Assessment (patient):** 5-point LK; increase ≥1 grade	(1) **Pain on walking:** VAS (0-100mm); ≥40mm (2) **Global Assessment (patient):** 5-point LK; 3-5	Not specified	Not specified	Not specified	None
Schnitzer, 2005[56]	"Flare"	No tool: increase in pain	**Pain:** VAS (0-100mm); ≥40mm	Not specified	24 hours	Not specified	None
Scott-Lennox, 2001[57]	"Flare"	(1) **Pain:** VAS (0-100mm)**;** ≥20mm (2) **Pain (physician):** 4-point LK; worsening ≥1 point (3) **Global Assessment (patient):** 4-point LK; worsening ≥1 point (4) **Global Assessment (physician)**:4 point LK; worsening ≥1 point	(1) **Pain:** VAS (0-100mm)**;** ≥40mm at baseline) (2) **Pain (physician):** 4-point LK; ≥2 (3) **Global Assessment (patient):** 4-point LK; ≥2 (4) **Global Assessment (physician)**: 4 point LK; worsening ≥2	14 day washout	Not specified	Not specified	Confirmatory Factor Analysis
Simon, 2009[90]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale; increase ≥2 and ≥25%	**Pain:** WOMAC LK3.1 Pain subscale; ≥'moderate' on ≥1 item	14 day washout	Not specified	Not specified	None
Silverfield, 2002[58]	"Flare"	**Pain:** No measurement tool; significant increase	Not specified	Not specified	Not specified	Pain requiring supplemental analgesic medication and/or an increase in NSAID dose	None
Strand, 2011[59]	"Flare"	**Global Assessment (patient):** 5-point LK; Increase ≥1	(1) **Global Assessment (patient):** 5-point LK; 'Fair, poor or very poor' (2) **Pain**: (0-10 NRS); ≥4 but <9 (3) **Global Assessment (physician)**: 5-point LK; 'Fair, poor or very poor'	14 day washout	Not specified	Not specified	None

Weaver, 1995[91]	"Flare"	(1) **Global Assessment (Physician)**: 5-point Likert; increase ≥1 grade (2) **Global Assessment (patient)**: 5-point LK; increase ≥1 grade (3) **Pain:** Worsening pain on motion and weight bearing	(1) **Global Assessment (Physician)**: 5-point Likert; ≥2 (2) **Global Assessment (patient)**: 5-point LK; ≥2	2-14 day washout	Not specified	Not specified	None
Wiesenhutter, 2005[60]	"Flare"	(1) **Pain on walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); increase ≥15mm (2) **Global Assessment (Investigator):** 5-point LK; worsening ≥1 unit	(1) **Pain on walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); ≥40mm	Not specified	Not specified	Not specified	None
Williams, 2001[61]	"Flare"	(1)**Global Assessment (patient):** 5-point LK; Increase ≥1 point (2) **Global Assessment (physician)**: 5-point LK; increase ≥1 point(3) **Composite Index**: Lequesne OA Severity Index (0-24); Increase ≥2 points	(1) **Global Assessment (patient):** 5-point LK; '(Fair), poor or very poor' (2) **Global Assessment (physician)**: 5-point LK; '(Fair), poor or very poor' (3) **Composite Index**: Lequesne OA Severity Index (0-24); ≥7 (4) **Pain:** VAS (0-100mm); ≥40mm	2-14 days	Not specified	Not specified	None
Wittenberg, 2006[62]	"Flare"	**Pain:** VAS (0-100mm); Increase ≥10mm	**Pain:** VAS (0-100mm); ≥40mm	2-7 day washout	Not specified	Not specified	None
Yeasted, 2014[63] (Pooled, abstract)	"Flare"	**Pain:** 0-10 NRS; Increase ≥2 points over the mean pain score from the previous 3 days	**Pain:** Average daily 0-10 NRS; 4-9	Not specified	Not specified	Not specified	None
Yocum 2000[78]	"Flare"	**Disease activity** (1) **Global (Investigator)**: Reduction of ≥ 1 grade (2) **Global Assessment (Patient)**: 100-mm VAS; Increase of ≥10mm	Not specified	≥3 days washout	Not specified	Not specified	None

		(3) **Pain: Overall assessment (patient):** 100-mm VAS; ≥35mm					
Young, 2014[64]	"Flare"	(3) **Pain:** WOMAC pain subscale; increase >15mm	**Pain:** WOMAC Pain subscale >40mm	Not specified	Not specified	Not specified	None
Zhao, 1999[65]	"Flare"	No measurement tool: Worsening of signs and symptoms after discontinuation of NSAIDs of analgesics	Not specified	2-7 day washout	Not specified	Not specified	None
NON-DRUG WITHDRAWAL STUDY DESIGN							
Atukorala, 2016[79] (abstract) Atukorala, 2016[26] (abstract)	"Flare"	**Pain:** (10-point NRS); increase >2 points from the mildest knee OA pain intensity reported at day 0	Not specified	Not specified	Not specified	Not specified	None
Bartholdy, 2016[80]	"Flare"	Not specified	**Pain:** (10-point NRS): Pain >5	Not specified	Not specified	Not specified	None
Bassiouni 2015[81] (abstract)	"Flare"	Not specified	**Global Assessment (physician):** KOFUS ≥7	Not specified	Not specified	Not specified	None
Cibere, 2004[87] Cibere, 2005[88]	"Flare"	(1) Patients perception of worsening of symptoms (2) **Pain walking on flat surface**: WOMAC VAS3.0 Q1 (0-100mm); increase ≥20mm (3) **Global Assessment (physician):** 5-point LK; worsening ≥1 grade	Not specified	Not specified	Not specified	Not specified	Definition determined by study rheumatologists to be a clinically important change in WOMAC-Ehrich2000/

							Bellamy 1998
Conrozier 2012[67]	"Flare"	Fulfilled 4 following criteria: (1) **Pain:** No measurement tool; 'sudden aggravation of knee pain' (2) causing nocturnal awakenings, (3) clinical evidence of effusion.	Not specified	Sudden aggravation of knee pain, whose beginning was identifiable	Not specified	Not specified	None
D'Agostino 2005[68]	"Flare"	Not specified	**Pain intensity during physical activity:** VAS-(0-100mm); ≥40mm	Not specified	48 hours	Not specified	None
Erfani, 2014[45] abstract) Erfani, 2014[82] (abstract) Ferreira[83] 2016 Hunter 2014[84] (abstract) Makovey 2015[85] (Protocol)	Exacerbation	**Pain:** VAS (0-100mm); Increase ≥20mm from mildest pain score reported at baseline	Not specified	Not specified	Not specified	Not specified	None

Jawad, 2005[69]	Exacerbation	**Pain symptoms:** Increased morning stiffness, night pain and synovial fluid effusion	Not specified	Not specified	Not specified	Not specified	None
Marty 2009[20]	"Flare"	No measurement tool: Morning stiffness >20mins, nocturnal awakening, limping, knee swelling, increased warmth, effusion	Not specified	Not specified	48 hours	Not specified	Regression analysis of cross-sectional data to validate proposed flare criteria
Murphy, 2015[70]	"Flare"	(1) Investigator definition: Inadequate pain relief for an episode of intense pain that is usually brought on by too much activity. (2) Participant definitions: Described in terms of pain quality, timing (onset and duration), antecedents and consequences. (3) **Pain magnitude:** increase in pain or 'intense' or 'severe' level of pain	**Pain:** ≥40 of 100mm or ≥4 of 10 on NRS	Patients described: 'Quick' or 'sudden'	Patients: 10 seconds to 15 minutes	Patients: Rest or take additional medication	For investigator definition: Battisti 2004, Pareek 2010. Plus researchers own experience.
Parry, 2017[86]	"Flare"	**Pain:** Recalled worst pain intensity in previous 6 months 0-10 NRS; ≥5	**Pain:** Recalled worse pain to be ≥2 points higher than recalled average pain (0-10 NRS) in previous 6 months	Not specified	Not specified	Not specified	Based on previous studies defining knee flares in OA and flares in diseases such as back

							pain and COPD.
Ricci 2005[55]	"Flare up"	**Pain:** Self-reported flare severity rating 0-10 NRS; increase ≥2 point over usual pain severity	Not specified	Not specified	Not specified	Not specified	Based on statistical analysis and clinical judgement
Wise 2010[71]	"Flare"	Not specified	**Pain:** WOMAC Pain subscale (0-10); score in highest 30% of all WOMAC scores	Not specified	Not specified	Not specified	None
Zhang 2009[72]	"Exacerbation or flare"	Not specified	(1) Pain: WOMAC pain subscale 0-10 (total score of 50 normalised to a 0-10 scale); score of ≥5, a score corresponding to highest 33% of all WOMAC scores	Not specified	Not specified	Not specified	None
Zhang 2011[73] (abstract)	"Exacerbation"	**Pain:** WOMAC Pain score VAS (0-500); increase ≥100 units	Not specified	Not specified	Not specified	Not specified	None
Zobel, 2016[93]	Exacerbation	**Pain:** 0-10 NRS; Increase ≥2	(1) Disabling pain	Not specified	8 hours	Not specified	None

Acronyms:
COPD- Chronic Obstructive Pulmonary Disease
KOFUS- Knee Osteoarthritis Flare-up Score
NRS-Numerical Rating scale
VAS- Visual Analogue Score
WOMAC- Western Ontario and McMaster Universities Osteoarthritis Index

262

One study used the term "flare-up"[53], two studies referred simply to "worsening of symptoms" [32, 51] and three studies used no specific label[35, 36, 74].

Coverage of key components

Onset/worsening of symptoms and signs beyond normal-day-to-day variability: Forty-four studies included onset or worsening of signs and symptoms as part of their definition[25, 27-33, 35-42, 44, 46-54, 56-65, 74-76, 78, 89-92]. All studies included increased pain intensity in their definition. A further two[53, 54] specified further signs and symptoms. These included swelling, inflammation, erythema, morning stiffness and nocturnal pain. No studies quantified day-to-day variability.

Twenty-six measurement tools were used to define onset/worsening of symptoms and signs. The most commonly used tools were the Western Ontario & McMaster Universities Arthritis index (WOMAC) Q1 (pain on walking on flat surface) 100mm Visual Analogue Scale (VAS) (n=9)[30, 31, 33, 39, 42, 46, 60, 74, 76] and the Investigator Assessment of Disease Status (n=11)[29-31, 39, 41, 46, 60, 74-76, 78] (Table 3).

Table 3: Summary of number and type of single and multi-item measurement tools used.

Single item scales:	
Pain on activity:	WOMAC Q1 3.0 VAS 'pain on walking on a flat surface' (0-100mm) [n=11] Pain on walking VAS (0-100mm) [n=5] Pain on movement VAS (0-100mm); Ambulatory pain (5-point Likert); Pain with physical activity VAS 11-point scale [n=2]
Pain (not further specified):	Pain VAS (0-100mm) [n=15] Patients Assessment of Pain Score (0-10); Pain Scale (0-3); Pain NRS (0-10) [n=11]
Standing knee pain	Item 5 WOMAC pain scale [n=1]
Global rating (physician/ investigator)	Investigator Assessment of Disease Status [n=11] Physicians Global Assessment of Arthritis [n=6] Physician Global Assessment of OA [n=2] Physician Global Assessment of Disease Status [n=2]; Investigator Assessed Pain Grade; (Physician) Overall Disease Activity (0-100); Physicians Pain Assessment (4-point LK) [n=3]
Global rating (patient)	Patients Global Assessment of Arthritis [n=7] Patient Global Assessment of OA [n=3] Patient Global Assessment of Disease Status [n=4]
Multiple-item scales:	
	Lequesne OA Severity Index [n=5] WOMAC LK3.1 (0-20) [n=3] WOMAC LK Pain subscale (0-25); WOMAC OA Index

	Questionnaire; WOMAC knee pain score (0-500) [n=7]; KOFUS (0-14)

N, number of included studies; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; VAS, visual analogue scale; OA, osteoarthritis; KOFUS, Knee Osteoarthritis Flare-up Score.

In addition, the format of global ratings appears to be variable as is use and reporting of the WOMAC[94]. However, despite the exact format of reporting being inconsistent, in general, studies used single items in 4 areas – pain on activity, pain (not necessarily on activity), physician/investigator global rating and patient global rating.

Temporal characteristics: None of the included drug withdrawal design studies reported a specific time for defining the speed of onset of symptoms. However, they did describe withdrawal or 'washout' periods whereby, after withdrawal of usual medication, participants were given a certain time frame in which to experience 'flare' symptoms in order that they were entered into the study. In total 30 of the studies specified a withdrawal period[28, 31, 32, 34-37, 39-41, 44, 46-53, 57, 59, 61, 62, 65, 74, 75, 77, 78, 89-91].For studies using a drug withdrawal design the duration of the washout period differed between studies, ranging from 2-15 days.

Four studies specified a time period for minimum duration of symptoms which ranged from 24 hours to 5 days[53, 54, 56, 58] .

Change in medication or healthcare usage: Only one study used increase in medication as part of their definition; 'pain requiring supplemental analgesic medication and/or an increase in NSAID dose'[58].

Additional domains: Thirty-six studies included a minimum threshold which was usually a minimum level of pain that was required before the participant was considered to have a flare[25, 27, 29-32, 34, 36-39, 41-44, 46-48, 52-54, 56, 57, 59-64, 74, 76, 77, 89-92]. There was general concordance with the minimum thresholds that different measurement tools used with a few exceptions. A threshold of 40mm on a 0-100mm scale was used in eight of ten studies using the WOMAC VAS 3.0 Q1 'pain on walking on a flat surface'[30, 31, 39, 42, 46, 60, 74, 76] and four of fourteen studies using the Patient Global Assessment of Disease Status[30, 46, 74, 76]. In studies using various forms of investigator/physician global assessment, the majority adopted a minimum threshold for a flare of 'fair, poor or very poor' [30, 31, 46, 74]. The minimum threshold on the Lequesne index (0-10) was either five[54] or seven[47, 52, 61].

Flare definitions in non-withdrawal flare/ discontinuation studies

Terminology used

"Flare" was the term most common used in non-withdrawal design studies[20, 26, 67, 68, 70, 71, 79-81, 86, 88](n=11) (Table 2). One study used the term "flare-up"[55], eight used "exacerbation"[45, 66, 69, 73, 82-85] (five publications were from the same team) and one referred to both "exacerbation" and "flare"[72]. None referred to "worsening of symptoms" or did not use any specific label.

Coverage of key components

Onset/worsening of symptoms and signs beyond normal-day-to-day variability:

Sixteen of twenty-two studies used onset or worsening of symptoms in their definition[26, 45, 55, 67, 69, 70, 73, 79, 82-88, 93]. Two studies did not use pain intensity as part of its definition[20, 81]. Three studies included symptoms other than pain in their definition[20, 67, 69]. These included nocturnal awakenings, effusion, morning stiffness, night pain, limping, and warmth.

The Murphy et al[70] study included an investigator definition of flare but also sought to describe patient experience of flares through face to face individual interviews. Both investigator and patient definitions included onset/worsening of symptoms and signs however there was no differentiation from day-to-day variability.

Seven studies used a measurement tool to define onset of signs and symptoms (Table 3). These included the Pain NRS (0-10)[26, 55, 66, 79, 86], WOMAC knee pain score VAS (0-500)[73], pain walking on a flat surface (WOMAC)[87, 88], Global Assessment of Disease Status (physician) (Likert 5-point scale)[87, 88], and knee pain VAS not further specified (0-100)[45, 82-85].

Temporal characteristics: Only one study set a definition for speed of onset, describing this only as 'sudden' with no further specification[67]. Patients in the Murphy et al study used the terms 'quick' and 'sudden' to describe flare onset[70]. Three studies specified a minimum duration of symptoms ranging from 8 to 48 hours[20, 66, 68]. In the Murphy et al study patients described duration of between 10 seconds to 15 minutes[70].

Change in medication/healthcare usage: No studies used change is medication or healthcare usage as part of their definition. However, in Murphy et al patients reported either taking rest or using additional medication[70].

Additional Domains: Two studies defined distribution-based minimum thresholds for flare as the highest 30%[72] or highest 33%[73] of WOMAC Pain Subscale scores among participants in the Longitudinal Examination of Arthritis Pain (LEAP) cohort (total score out of 50 was normalised to a 0-10 scale).

DISCUSSION

Flares in OA are recognised in existing clinical guidance[95] and reviews[96, 97] but typically merit little more than a passing mention. Only one recent study has sought

to define flare-ups in in hip and knee OA but this only yielded 23 studies and four of the included studies did not contain clear definitions for a flare-up[98]. Our review was motivated by an interest in seeking greater clarity on how these phenomena might be defined by undertaking a broad search strategy, noting that similar efforts have been pursued in other chronic diseases. While we found no current single, agreed definition of OA flare, our review of 69 published studies suggests a number of common domains which may capture cardinal features. These were: onset/worsening of symptoms and signs, attainment of a minimum symptom threshold during flare, speed of onset/worsening, and duration of elevated symptoms/signs. However, we found considerable variation in how these domains have been operationalised for measurement suggesting the need for further conceptual clarification and consensus.

Each potential cardinal feature of OA flare presents different challenges for achieving consensus and how these are resolved depends partly on whether the goal is a shared definition for reproducible and comparative research or for identifying these phenomena in routine practice. Most studies included in our review required an increase in pain over 'usual' or 'baseline' intensity. Although this was measured using a wide range of measurement instruments several studies selected an increase of 2 or more points on a 0-10 scale providing a possible starting point for consensus. Yet this possible 'signal' is arguably difficult to interpret without also considering the

amount of background 'noise', i.e. within-person diurnal[99] and day-to-day variability[100], and the absolute level ('minimum threshold') of pain during a flare. In the study by Marty et al an increase in pain was not independently associated with flare-up after adjusting for other potential features[20]. Further research on detecting flares over within-person 'normal' variability by collecting frequent repeated measures of pain intensity may be valuable but this approach would not be feasible when identifying flares presenting at the point of care in routine clinical practice. Instead, this may have to rely on the judgement of the patient and/or clinician, the approach used, for example, in defining exacerbations in COPD[1]. A similar consideration surrounds the speed of onset, which was not well defined by studies in our review. Drug withdrawal design studies specified washout periods between 2-15 days but this is unlikely to be synonymous with speed of onset. The remaining studies used terms such as 'sudden' and 'quick'. In COPD, for instance, a judgement around 'acute onset' or 'sudden onset' appears to be acceptable for clinical recommendations but we would add that the speed of onset of OA flares ought to be considered also in relation to underlying biologically plausible mechanisms. Indeed presumed aetiology has been argued as a useful feature in defining acute exacerbations in COPD[101]. Minimum duration ranged from 8 hours to 5 days in our review however this was not widely reported. COPD definitions refer to a 'sustained worsening' of symptoms[2] but does not appear to be a feature in other chronic diseases. A minimum duration in OA may help distinguish flares from day-to-day

variability. Increase in medication was not found to be a key component in this review despite it being a feature in other chronic diseases; AS[5], SLE[4, 102], Inflammatory Bowel Disease[103], COPD[1]. Interference with function did not emerge strongly from our review as a cardinal feature of OA flare. In other chronic musculoskeletal conditions, such as back pain, interference with function was not shown to be significantly associated with having a flare up[104] and this domain does not feature in the definitions of exacerbations or flares in diseases such as COPD[1, 2], asthma[3], AS[5] or SLE[4].

Our review has several strengths but also some weaknesses that deserve attention. We adopted a broad search strategy, covering a wide range of databases, and featuring bibliography checks, contact with authors, inclusion of conference abstracts, no language restrictions, and a minimal threshold (any description or definition of flare) for inclusion. Five studies that were included in the Cross et al[98] review were not included in this study; four did not contain a clear definition of flare-up, including one which gave a definition of knee OA progression and the final paper by Sands et al[105] was not in our search but the original study was[59]. We did not, however, search the grey literature and we did not include some potential synonyms as search terms ('attack', 'episode', 'fluctuations'). Nevertheless, we argue that our review provides a reasonably comprehensive summary of how 'flares' in OA have been described and defined in the medical literature. The majority of studies describe

experimental 'flare design' trials in which flares are induced by drug withdrawal prior to enrolment and randomisation. While intentional or unintentional reduction in usual analgesia may indeed be one trigger for flare, experimentally induced flares should not be assumed to represent 'naturally occurring' flares. Flare design trials, for example, are unlikely to capture change in management or healthcare usage that may be a common consequence of OA flares – something that is included in flare definitions in other conditions such as AS[5], SLE[4, 102], inflammatory bowel disease[103], and COPD[1].

A systematic review such as this cannot hope to resolve the need for a common conception and definition of flares in OA. Definitions for exacerbations of disease states are generally reached through a long process of consensus exercises involving key stakeholders, experts and patients in addition to appraisal of relevant literature from studies using multiple methods[6, 8, 106]. However, we believe that a consensus definition that is reliable, valid, and feasible and widely acceptable both clinically and for research purposes should now be sought.

CONCLUSION

A broad range of ad-hoc definitions currently exist in the medical literature. The majority are from drug-withdrawal or flare-induced trials rather than 'naturally'

occurring flares. The cardinal feature is pain intensity with minimum symptom threshold being another important feature. This review has identified the need to gain consensus on a common definition that can be used for research and clinical application.

Acknowledgements

We would like to thank you to Popay et al for allowing us to use their guidance on the conduct of narrative synthesis in systematic reviews. We would also like to thank Jo Jordan and Opeyemi Babatunde for their advice on conducting a systematic review.

Contributions

All authors were involved in conception and design of the study, analysis and interpretation of data, drafting the article, critical revision of the article for important intellectual content, final approval of the article. ELP and MJT extracted and synthesised data. ELP assembled the data.

GMP (g.m.peat@keele.ac.uk) takes responsibility for the integrity of the work as a whole from inception to finished article.

Data sharing statement

No unpublished data is available following this study

Role of the funding source

The paper presents research funded by Arthritis Research UK [Centre of Excellence award: 18171]. ELP received funding from a National Institute for Health Research (NIHR) In-Practice Fellowship and a NIHR Academic Clinical Fellowship. MJT received funding from a NIHR School for Primary Care Research Launching Fellowship and is currently funded by an Integrated Clinical Academic Programme Clinical Lectureship from the NIHR and Health Education England (HEE) (ICA-CL-2016-02-014). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, HEE or the Department of Health. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interest statement

GP received consultancy fees from InFirst plc and Good Relations plc.

Figure and Table Legends

Figure 1: PRISMA Flowchart

Table 1: Characteristics of all included studies

Table 2: Summary of number and type of single and multi-item measurement

Table 3: Definition, terminology and measurement instruments used in all included studies

Supplementary data: Database search strategy

REFERENCES

1 Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the diagnosis, management and prevention of COPD: GOLD 2016.

2 National Institute for Health and Care Excellence (NICE). Chronic obstructive pulmonary disease in over 16s: diagnosis and management (CG101). London: NICE 2010.

3 Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention: GINA 2015.

4 Ruperto N, Hanrahan L, Alarcón G, et al. International consensus for a definition of disease flare in lupus, *Lupus* 2011;20:453-62.

5 Stone MA, Pomeroy E, Keat A, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration, *Rheumatology* 2008;47:1213-8.

6 Bingham CO, Alten R, Bartlett SJ, et al. Identifying Preliminary Domains to Detect and Measure Rheumatoid Arthritis Flares: Report of the OMERACT 10 RA Flare Workshop, *The Journal of Rheumatology* 2011;38:1751-8.

7 Bykerk VP, Lie E, Bartlett SJ, et al. Establishing a Core Domain Set to Measure Rheumatoid Arthritis Flares: Report of the OMERACT 11 RA Flare Workshop, *The Journal of Rheumatology* 2014;41:799-809.

8 Bartlett SJ, Hewlett S, Bingham CO, et al. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus, *Annals of the Rheumatic Diseases* 2012;71:1855-60.

9 Taylor WJ, Shewchuk R, Saag KG, et al. Toward a valid definition of gout flare: Results of consensus exercises using delphi methodology and cognitive mapping, *Arthritis Care & Research* 2009;61:535-43.

10 Schmitt J, Spuls PI, Thomas KS, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials, *J Allergy Clin Immunol* 2014;134:800-7.

11 Holla JFM, van dL, Knol DL, et al. The association of body-mass index and depressed mood with knee pain and activity limitations in knee osteoarthritis: results from the Amsterdam osteoarthritis cohort, *BMC Musculoskeletal Disorders* 2013;14:296.

12 Collins JE, Katz JN, Dervan EE, et al. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative, *Osteoarthritis and Cartilage* 2014;22:622-30.

13 Leffondré K, Abrahamowicz M, Regeasse A, et al. Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators, *J Clin Epidemiol* 2004;57:1049-62.

14 Emrani PS, Katz JN, Kessler CL, et al. Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis, *Osteoarthritis and Cartilage* 2008;16:873-82.

15 Bartlett SJ, Ling SM, Mayo NE, et al. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis, *Arthritis Care & Research* 2011;63:1722-8.

16 Hawker GA, Stewart L, French MR, et al. Understanding the pain experience in hip and knee osteoarthritis – an OARSI/OMERACT initiative, *Osteoarthritis and Cartilage* 2008;16:415-22.

17 Arthritis Research UK. Osteoarthritis: Patient Information Booklet. 2012.

18 Bonnet CS, Walsh DA. Osteoarthritis, angiogenesis and inflammation, *Rheumatology* 2005;44:7-16.

19 Smith TO, Zou K, Abdullah N, et al. Does flare trial design affect the effect size of non-steroidal anti-inflammatory drugs in symptomatic osteoarthritis? A systematic review and meta-analysis, *Annals of the Rheumatic Diseases* 2016;75:1971-8.

20 Marty M, Hilliquin P, Rozenberg S, et al. Validation of the KOFUS (Knee Osteoarthritis Flare-Ups Score), *Joint Bone Spine* 2009;76:268-72.

21 Rutjes AS, Jüni P, Da Costa BR, et al. Viscosupplementation for osteoarthritis of the knee: A systematic review and meta-analysis, *Ann Intern Med* 2012;157:180-91.

22 Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions Version 5.1.0. Available from www.handbook.cochrane.org.: The Cochrane Colloboration 2011.

23 Popay J, Roberts H, S, A., et al. Guidance on the conduct of narrative synthesis in systematic reviews: A product of the ESRC methods programme Lancaster: ESRC Method Programme, 2006.

24 Thomas J, Harden A, Newman M. Synthesis: Combining results systematically and appropriately. In: Gough A, Oliver S, Thomas J, eds. An introduction to systematic reviews. London: Sage publications limited 2013:191-2.

25 Altman R, Hochberg M, Gibofsky A, et al. Efficacy and safety of low-dose SoluMatrix meloxicam in the treatment of osteoarthritis pain: A 12-week, phase 3 study, *Curr Med Res Opin* 2015;31:2331-43.

26 Atukorala I, Pathmeswaran A, Makovey J, et al. Is there a relationship between the intermittent and constant osteoarthritis pain score (ICOAP) and pain flares in knee osteoarthritis? (abstract) [abstract]. *Osteoarthritis and Cartilage* 2016;24:S429-30.

27 Baer PA, Thomas LM, Shainhouse Z. Treatment of osteoarthritis of the knee with a topical diclofenac solution: a randomised controlled, 6-week trial ISRCTN53366886], *BMC Musculoskeletal Disorders* 2005;6:44.

28 Baraf HSB, Gloth FM, Barthel HR, et al. Safety and Efficacy of Topical Diclofenac Sodium Gel for Knee Osteoarthritis in Elderly and Younger Patients, *Drugs Aging* 2011;28:27-40.

29 Battisti WP, Katz NP, Weaver AL, et al. Pain management in osteoarthritis: A focus on onset of efficacy—a comparison of rofecoxib, celecoxib, acetaminophen, and nabumetone across four clinical trials, *The Journal of Pain* 2004;5:511-20.

30 Bingham CO, Sebba AI, Rubin BR, et al. Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies, *Rheumatology* 2007;46:496-507.

31 Birbara C, Ruoff G, Sheldon E, et al. Efficacy and safety of rofecoxib 12.5 mg and celecoxib 200 mg in two similarly designed osteoarthritis studies, *Curr Med Res Opin* 2006;22:199-210.

32 Bocanegra T, Weaver A, Tindall E, et al. Diclofenac/misoprostol compared with diclofenac in the treatment of osteoarthritis of the knee or hip: a randomized,

placebo controlled trial. Arthrotec Osteoarthritis Study Group. *Journal of Rheumatology* 1998;25:1602-11.

33 Boswell DJ, Ostergaard K, Philipson RS, et al. Evaluation of GW406381 for Treatment of Osteoarthritis of the Knee: Two Randomized, Controlled Studies, *The Medscape Journal of Medicine* 2008;10:259.

34 Brandt KD, Mazzuca SA, Buckwalter KA. Acetaminophen, like conventional NSAIDs, may reduce synovitis in osteoarthritic knees, *Rheumatology* 2006;45:1389-94.

35 Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: A randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium, *Arch Intern Med* 2003;163:169-78.

36 Ehrich E, Schnitzer T, McIlwain H, et al. Effect of specific COX-2 inhibition in osteoarthritis of the knee: a 6 week double blind, placebo controlled pilot study of rofecoxib. Rofecoxib Osteoarthritis Pilot Study Group. *Journal of Rheumatology* 1999;26:2438-47.

37 Essex M, O'Connell M, Brown PB. Response to Nonsteroidal Anti-Inflammatory Drugs in African Americans with Osteoarthritis of the Knee, *Journal of International Medical Research* 2012;40:2251-66.

38 Gibofsky A, Hochberg MC, Jaros MJ, et al. Efficacy and safety of low-dose submicron diclofenac for the treatment of osteoarthritis pain: A 12 week, phase 3 study, *Curr Med Res Opin* 2014;30:1883-93.

39 Gineyts E, Mo JA, Ko A, et al. Effects of ibuprofen on molecular markers of cartilage and synovium turnover in patients with knee osteoarthritis, *Annals of the Rheumatic Diseases* 2004;63:857-61.

40 Goldberg M, McIlwain H, Poiley J, et al. Controlled-release naproxen in the treatment of osteoarthritis, *Current Therapeutic Research-Clinical and Experimental* 1988;44:51-60.

41 Gottesdiener K, Schnitzer T, Fisher C, et al. Results of a randomized, dose ranging trial of etoricoxib in patients with osteoarthritis, *Rheumatology* 2002;41:1052-61.

42 Hochberg MC, Fort JG, Svensson O, et al. Fixed-dose combination of enteric-coated naproxen and immediate-release esomeprazole has comparable efficacy to celecoxib for knee osteoarthritis: two randomized trials, *Curr Med Res Opin* 2011;27:1243-53.

43 Katz N, Sun S, Johnson F, et al. ALO-01 (Morphine Sulfate and Naltrexone Hydrochloride) Extended-Release Capsules in the Treatment of Chronic Pain of Osteoarthritis of the Hip or Knee: Pharmacokinetics, Efficacy, and Safety, *The Journal of Pain* 2010;11:303-11.

44 Kivitz AJ, Makarowski WS, Fiechtner JJ, et al. A Flexible Daily Dosage Regimen of Oxaprozin Potassium in Patients with Acute Knee Pain Associated with Osteoarthritis, *Clinical Drug Investigation* 2001;21:745-53.

45 Erfani T, Zhang Y, Makovey J, et al. Intermittent analgesic use and risk of pain exacerbation in knee osteoarthritis: A web based case-crossover study (abstract) [abstract]. *Arthritis and Rheumatology* 2014;66.

46 Leung AT, Malmstrom K, Gallacher AE, et al. Efficacy and Tolerability Profile of Etoricoxib in Patients with Osteoarthritis: A Randomized, Double-blind, Placebo and Active-comparator Controlled 12-Week Efficacy Trial, *Curr Med Res Opin* 2002;18:49-58.

47 Luyten FP, Geusens P, Malaise M, et al. A prospective randomised multicentre study comparing continuous and intermittent treatment with celecoxib in patients with osteoarthritis of the knee or hip, *Annals of the Rheumatic Diseases* 2007;66:99-106.

48 Manicourt D, Bevilacqua M, Righini V, et al. Comparative Effect of Nimesulide and Ibuprofen on the Urinary Levels of Collagen Type II C-Telopeptide Degradation Products and on the Serum Levels of Hyaluronan and Matrix Metalloproteinases-3 and -13 in Patients with Flare-Up of Osteoarthritis, *Drugs in R & D* 2005;6:261-71.

49 Mazzuca S, Brandt K, Lane K, et al. Knee pain reduces joint space width in conventional standing anteroposterior radiographs of osteoarthritic knees, *Arthritis Rheum* 2002;46:1223-7.

50 McIlwain H, Silverfield JC, Cheatum DE, et al. Intra-articular orgotein in osteoarthritis of the knee: A placebo-controlled efficacy, safety, and dosage comparison, *Am J Med* 1989;87:295-300.

51 Mendelsohn S. Clinical efficacy and tolerability of naproxen in osteoarthritis patients using twice-daily and once-daily regimens, *Clinical therapeutics* 1991;13:8-15.

52 Moskowitz RW, Sunshine A, Hooper M, et al. An analgesic model for assessment of acute pain response in osteoarthritis of the knee, *Osteoarthritis and Cartilage* 2006;14:1111-8.

53 Pareek A, Chandurkar N, Sharma VD, et al. A randomized, multicentric, comparative evaluation of aceclofenac-paracetamol combination with aceclofenac alone in Indian patients with osteoarthritis flare-up, *Expert Opin Pharmacother* 2009;10:727-35.

54 Pareek A, Chandurkar N, Ambade R, et al. Efficacy and Safety of Etodolac-Paracetamol Fixed Dose Combination in Patients With Knee Osteoarthritis Flare-up: A Randomized, Double-blind Comparative Evaluation, *Clin J Pain* 2010;26:561-6.

55 Ricci JA, Stewart WF, Chee E, et al. Pain Exacerbation as a Major Source of Lost Productive Time in US Workers With Arthritis, *Arthritis & Rheumatism: Arthritis Care & Research* 2005;53:673-81.

56 Schnitzer TJ, Fricke JR, Gitton X, et al. Lumiracoxib in the treatment of osteoarthritis, rheumatoid arthritis and acute postoperative dental pain: results of three dose-response studies, *Curr Med Res Opin* 2005;21:151-61.

57 Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, et al. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis, *Arthritis & Rheumatism* 2001;44:1599-607.

58 Silverfield JC, Kamin M, Wu S, et al. Tramadol/acetaminophen combination tablets for the treatment of osteoarthritis flare pain: a multicenter, outpatient, randomized, double-blind, placebo-controlled, parallel-group, add-on study, *Clin Ther* 2002;24:282-97 doi:http://dx.doi.org/10.1016/S0149-2918(02)85024-X [published Online First: February 2002].

59 Strand V, Simon LS, Dougados M, et al. Treatment of osteoarthritis with continuous versus intermittent celecoxib, *J Rheumatol* 2011;38:2625-34.

60 Wiesenhutter CW, Boice JA, Ko A, et al. Evaluation of the Comparative Efficacy of Etoricoxib and Ibuprofen for Treatment of Patients With Osteoarthritis: A Randomized, Double-Blind, Placebo-Controlled Trial, *Mayo Clin Proc* 2005;80:470-9.

61 Williams GW, Hubbard RC, Yu SS, et al. Comparison of once-daily and twice-daily administration of celecoxib for the treatment of osteoarthritis of the knee, *Clin Ther* 2001;23:213-27 doi:http://dx.doi.org/10.1016/S0149-2918(01)80004-7 [published Online First: February 2001].

62 Wittenberg RH, Schell E, Krehan G, et al. First-dose analgesic effect of the cyclo-oxygenase-2 selective inhibitor lumiracoxib in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled comparison with celecoxib NCT00267215], *Arthritis Research & Therapy* 2006;8:R35.

63 Yeasted R, McPherson J, Schnitzer T. Characterization of osteoarthritis pain variability (abstract) [abstract]. *Osteoarthritis and Cartilage* 2014;22:S390-1.

64 Young C, Parenti D, Hochberg M. Lower-dose diclofenac capsules developed using solumatrix fine particle technology result in clinically meaningful improvements in pain in a phase 3 study of patients with osteoarthritis (abstract) [abstract]. *Osteoarthritis and Cartilage* 2014;22.

65 Zhao SZ, McMillen JI, Markenson JA, et al. Evaluation of the Functional Status Aspects of Health-Related Quality of Life of Patients with Osteoarthritis Treated with Celecoxib, *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 1999;19:1269-78.

66 Zobel I, Erfani T, Bennell K, et al. Relationship of buckling and knee injury to pain exacerbation in knee osteoarthritis: A web-based case-crossover stud, *Interact J Med Res* 2014;66:S560-1.

67 Conrozier T, Mathieu P, Vignon E, et al. Differences in the osteoarthritic synovial fluid composition and rheology between patients with or without flare: a pilot study. *Clinical and experimental rheumatology* 2012;30:729-34.

68 D'Agostino MA, Conaghan P, Le Bars M, et al. EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1: Prevalence of inflammation in osteoarthritis, *Annals of the Rheumatic Diseases* 2005;64:1703-9.

69 Jawad ASM. Analgesics and osteoarthritis: are treatment guidelines reflected in clinical practice? *Am J Ther* 2005;12:98-104.

70 Murphy SL, Lyden AK, Kratz AL, et al. Characterizing pain flares from the perspective of individuals with symptomatic knee osteoarthritis, *Arthritis Care and Research* 2015;67:1103-11.

71 Wise BL, Niu J, Zhang Y, et al. Psychological factors and their relation to osteoarthritis pain, *Osteoarthritis and Cartilage* 2010;18:883-7.

72 Zhang Y, Zhang B, Wise B, et al. Statistical approaches to evaluating the effect of risk factors on the pain of knee osteoarthritis in longitudinal studies, *Curr Opin Rheumatol* 2009;21:513-9.

73 Zhang Y, Wheaton D, N, J., et al. Recent heavy physical activities trigger knee pain exacerbation in persons with symptomatic knee osteoarthritis (abstract) [abstract]. *Arthritis & Rheumatism* 2011;63(10).

74 Day R, Morrison B, Luza A, et al. A randomized trial of the efficacy and tolerability of the cox-2 inhibitor rofecoxib vs ibuprofen in patients with osteoarthritis, *Arch Intern Med* 2000;160:1781-7.

75 Kivitz AJ, Greenwald MW, Cohen SB, et al. Efficacy and Safety of Rofecoxib 12.5 mg Versus Nabumetone 1,000 mg in Patients with Osteoarthritis of the Knee: A Randomized Controlled Trial, *J Am Geriatr Soc* 2004;52:666-74.

76 Bingham CO, Smugar SS, Wang H, et al. Predictors of Response to Cyclo-Oxygenase-2 Inhibitors in Osteoarthritis: Pooled Results from Two Identical Trials Comparing Etoricoxib, Celecoxib, and Placebo, *Pain Medicine* 2011;12:352-61.

77 Essex MN, Behar R, O'Connell MA, et al. Efficacy and tolerability of celecoxib and naproxen vs placebo in hispanic patients with knee osteoarthritis, *Osteoarthritis and Cartilage* 2013;21.

78 Yocum D, Fleischmann R, Dalgin P, et al. Safety and efficacy of meloxicam in the treatment of osteoarthritis: A 12-week, double-blind, multiple-dose, placebo-controlled trial, *Arch Intern Med* 2000;160:2947-54.

79 Atukorala I, Pathmeswaran A, Chang T, et al. Do Traditional Risk Factors for Knee Osteoarthritis Predict Pain Flares in Knee Osteoarthritis? *Ann Rheum Dis* 2016;75:835.

80 Bartholdy C, Klokker L, Bandak E, et al. A Standardized "Rescue" Exercise Program for Symptomatic Flare-up of Knee Osteoarthritis: Description and Safety Considerations, *J Orthop Sports Phys Ther* 2016;46:942-6.

81 Bassiouni H. Detection of changes in the serum and synovial fluid levels of resistin during flare ups and remissions in primary knee osteoarthritis, *Arthritis and Rheumatology* 2015;67.

82 Erfani T, Makovey J, Bennell K, et al. Psychosocial Factors and Pain Exacerbation in Knee Osteoarthritis: a Web Based Case-Crossover Study, *Intern Med J* 2014;44:16-.

83 Ferreira ML, Zhang Y, Metcalf B, et al. The influence of weather on the risk of pain exacerbation in patients with knee osteoarthritis - a case-crossover study, *Osteoarthritis and cartilage* 2016;24:2042-7.

84 Hunter DJ, Bennell K, Makovey J, et al. Psychosocial Factors and Pain Exacerbation in Knee Osteoarthiritis: a Web Based Case-Crossover Study, *Osteoarthritis and Cartilage* 2014;22:S21-2.

85 Makovey J, Metcalf B, Zhang Y, et al. Web-Based Study of Risk Factors for Pain Exacerbation in Osteoarthritis of the Knee (SPARK-Web): Design and Rationale, *JMIR research protocols* 2015;4.

86 Parry E, Ogollah R, Peat G. Significant pain variability in persons with, or at high risk of, knee osteoarthritis: preliminary investigation based on secondary analysis of cohort data, *BMC musculoskeletal disorders* 2017;18:80.

87 Cibere J, Kopec JA, Thorne A, et al. Randomized, double-blind, placebo-controlled glucosamine discontinuation trial in knee osteoarthritis, *Arthritis Care and Research* 2004;51:738-45.

88 Cibere J, Kopec JA, Esdaile JM, et al. Glucosamine sulfate and cartilage type II collagen degradation in patients with knee osteoarthritis: randomized discontinuation trial results employing biomarkers. In: Anonymous . Journal of rheumatology 2005:896-902.

89 Roth SH, Shainhouse JZ. 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, *Arch Intern Med* 2004;164:2017-23.

90 Simon LS, Grierson LM, Naseer Z, et al. 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* 2009;143:238-45.

91 Weaver A, Rubin B, Caldwell J, et al. Comparison of the efficacy and safety of oxaprozin and nabumetone in the treatment of patients with osteoarthritis of the knee, *Clin Ther* 1995;17:735-45.

92 Rother M, Lavins BJ, Kneer W, et al. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral celecoxib and placebo in osteoarthritis of the knee: multicentre randomised controlled trial, *Annals of the Rheumatic Diseases* 2007;66:1178-83.

93 Zobel I, Erfani T, Bennell KL, et al. Relationship of Buckling and Knee Injury to Pain Exacerbation in Knee Osteoarthritis: A Web-Based Case-Crossover Study, *Interactive journal of medical research* 2016;5:e17.

94 Woolacott NF, Corbett MS, Rice SJC. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials, *Rheumatology* 2012;51:1440-6.

95 National Institute for Health and Care Excellence (NICE). Osteoarthritis: care and management (CG177). London: NICE 2014.

96 Buttgereit F, Burmester G, Bijlsma JWJ. Non-surgical management of knee osteoarthritis: where are we now and where do we need to go? *RMD Open* 2015;1.

97 Porcheret M, Healey E, Dziedzic K, et al. Ostoearthritis: a modern approach to diagnosis and management, *Arthritis Research UK* 2011;Series 6.

98 Cross M, Dubouis L, Mangin M, et al. Defining Flare in Osteoarthritis of the Hip and Knee: A Systematic Literature Review- OMERACT Virtual Special Interest Group, *J Rheumatol* 2017.

99 Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee, *J Rheumatol* 1990;17:364-72.

100 Allen KD, Coffman CJ, Golightly YM, et al. Daily pain variations among patients with hand, hip, and knee osteoarthritis, *Osteoarthritis and Cartilage* 2009;17:1275-82.

101 Makris D, Bouros D. COPD exacerbation: Lost in translation, *BMC Pulmonary Medicine* 2009;9:6.

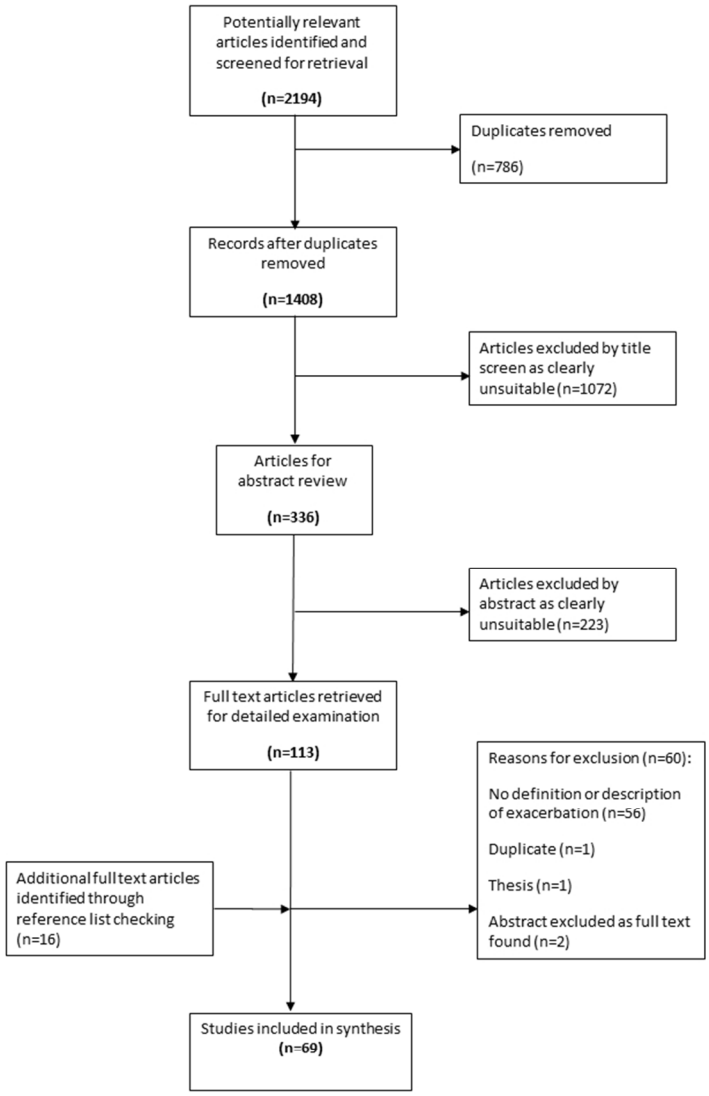
102 Fitzgerald JD, Grossman JM. Validity and reliability of retrospective assessment of disease activity and flare in observational cohorts of lupus patients, *Lupus* 1999;8:638-44.

103 Lewis JD, Aberra FN, Lichtenstein GR, et al. Seasonal variation in flares of inflammatory bowel disease, *Gastroenterology*;126:665-73.

104 Suri P, Saunders KW, Von Korff M. Prevalence and Characteristics of Flare-ups of Chronic Nonspecific Back Pain in Primary Care: A Telephone Survey, *Clin J Pain* 2012;28:573-80.

105 Sands GH, Brown PB, Essex MN. The Efficacy of Continuous Versus Intermittent Celecoxib Treatment in Osteoarthritis Patients with Body Mass Index â‰¥30 and *The Open Rheumatology Journal* 2013;7:32-7.

106 Berthelot J, De Bandt M, Morel J, et al. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: The 'FLARE' instrument, *Annals of the Rheumatic Diseases* 2012;71:1110-6.



190x275mm (96 x 96 DPI)

Online supplement: Example search strategy

Table 1: Key terms and MeSH headings used for EMBASE database search. The concepts were combined as follows: "KNEE JOINT" AND "ACUTE EVENTS"

Concepts	Search terms
KNEE JOINT	"knee adj3 (pain OR painful)" or "Knee osteoarthritis" or "knee adj3 (arthrosis)" or "knee adj3 (joint OR joints OR degenerative)" or "knee adj3 (osteoarthritis)"
ACUTE EVENTS	"exacerbation" or "flare" or "daily adj3 (pain)" or "pain AND (diary OR diaries)" or "pain adj3 (variab$)"

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6-7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Online supplement
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	n/a
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	9

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n/a
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10 (flowchart)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10, Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A but rationale on 18
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	n/a
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	n/a
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	19-35
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	36-37
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	39
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	40

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

PRISMA 2009 Checklist

BMJ Open

Defining Acute Flares in Knee Osteoarthritis: A Systematic Review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019804.R1
Article Type:	Research
Date Submitted by the Author:	12-Feb-2018
Complete List of Authors:	Parry, Emma; Keele University, Research Institute for Primary Care and Health Sciences Thomas, M; Keele University, Research Institute for Primary Care and Health Sciences Peat, George; Keele University, Research Institute for Primary Care & Health Sciences
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	General practice / Family practice
Keywords:	Osteoarthritis, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Flare

DEFINING ACUTE FLARES IN KNEE OSTEOARTHRITIS: A SYSTEMATIC REVIEW

Emma L. Parry[1] , e.parry@keele.ac.uk

Martin J. Thomas[1], m.thomas@keele.ac.uk

George Peat[1*], g.m.peat@keele.ac.uk

[1] *Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, UK*

*Correspondence: Professor George Peat, Professor of Clinial Epidemiology, Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG. Tel: +44 (0) 1782 732929. Fax:+44 (0) 1782 734719. Email: g.m.peat @keele.ac.uk

Keywords: knee osteoarthritis, flare, systematic review

ABSTRACT

Objective: To identify and critically synthesise definitions of acute flares in knee osteoarthritis (OA) reported in the medical literature.

Design: Systematic review and narrative synthesis. We searched MEDLINE, EMBASE, Web of science and 6 other electronic databases (inception to July 2017) for original articles and conference abstracts reporting a definition of acute flare (or synonym) in humans with knee OA. There were no restrictions by language or study design (apart from iatrogenic induced flare-ups e.g. injection-induced). Data extraction comprised: definition, pain scale used, flare duration or withdrawal period, associated symptoms, definition rationale, terminology (e.g. exacerbation or flare), baseline OA severity, age, gender, sample size and study design.

Results: Sixty-nine articles were included (46 flare-design trials, 17 observational studies, 6 other designs; sample sizes: 15-6085). Domains used to define flares included: worsening of signs and symptoms (61 studies, 27 different measurement tools), specifically increased pain intensity; minimum pain threshold at baseline (44 studies); minimum duration (7 studies, range 8-48 hours); speed of onset (2 studies, defined as 'sudden' or 'quick'); requirement for increased medication (2 studies). No definitions included activity interference.

Conclusions: The concept of OA flare appears in the medical literature but most often in the context of flare design trials (pain increases observed after stopping usual treatment). Key domains, used to define acute events in other chronic

conditions, appear relevant to OA flare and could provide the basis for consensus on a single, agreed definition of 'naturally occurring' OA flares for research and clinical application.

PROSPERO registration: CRD42014010169

Strengths and limitations of this study

Strengths

- Identified key domains that are used to define acute events by undertaking a comprehensive synthesis of definitions used in the medical literature.
- Broad search strategy covering a wide range of databases including bibliography checks and conference abstracts.
- Prospectively registered with Prospero

Limitations

- Did not include potential synonyms as search terms ('attack', 'episode', 'fluctuations')
- Data extraction was performed by only a single reviewer.

INTRODUCTION

Recurrent acute events or episodes feature in the natural history of many chronic health conditions. The extent to which they characterise the condition varies, as do the presumed pathophysiological mechanisms, and scientific and lay terms used to describe them (e.g. an acute exacerbation of chronic obstructive pulmonary disease (COPD) or asthma, an attack of gout or a rheumatoid arthritis flare). With recognition of their importance has come concerted effort to define these phenomena. Definitions for exacerbations or flares currently exist for COPD[1, 2] , asthma[3], systemic lupus erythematosus (SLE)[4], and ankylosing spondylitis (AS)[5] and there are working groups currently trying to define these for rheumatoid arthritis[6-8], gout[9], and atopic dermatitis/eczema[10]. Despite the different language used, these definitions share some common, core domains: the onset or worsening of symptoms and signs above normal day-to-day variability; speed of onset; duration of sustained worsening; and change in medication/healthcare usage.

Osteoarthritis (OA) appears to comprise multiple disease trajectories[11-15] and symptom variability over time and the presence of intermittent pain is well-recognised[16]. Although OA does not typically have the same very obvious acute events as conditions like gout, flares in OA joints are encountered in practice, these

phenomena appear in patient literature[17], have been discussed in expert reviews[18], and are mentioned in 'flare design' trials in OA[19]**.** These studies invoke acute episodes of pain or flare-ups by asking patients to withdraw their usual medication.

In 2009 Marty et al proposed scoring criteria for knee OA flares based on nocturnal awakening, knee effusion, morning stiffness and limping[20] but it is unclear whether this has contributed to a common understanding, shared terminology and criteria. A common definition of OA flare could be important for a number of reasons; (i) to facilitate communication between researchers, (ii) to allow more direct comparisons between studies on frequencies, determinants and course of events, (iii) to facilitate new insights into novel pathophysiological mechanisms and treatments through valid and homogenous case definitions, and (iv) to help clinicians with prompt diagnosis and management.

The aim of this systematic review was to explore the extent to which a concept of OA flare is reported in the medical literature and the prospects for a common, shared definition of these for research and clinical application. A review addressing similar aims but not registered on PROSPERO came to our attention when it was published while we were drafting our manuscript[21]. In principle and upon comparing the

respective findings of both reviews, we felt our review could justify making an original contribution.

METHODS

This systematic review was registered with PROSPERO registration number CRD42014010169. The review protocol has not been published.

Literature sources and study selection

We searched electronic databases from inception to July 2017; ASSIA, EMBASE, Web of Science, Health Management Information Consortium (HMIC), SPORTDiscus, Medline, CINAHL, PsycINFO, AMED, Ageline, Cochrane Database of Systematic Reviews and Cochrane Controlled Clinical Trials (CENTRAL). The search was developed using previously piloted terms for knee OA and a literature search for common terms used to describe acute events. Searches used combined and/or truncated key terms including: ("KNEE OSTEOARTHRITIS" OR (knee N3 pain) OR (knee N3 arthrosis) OR (knee N3 joint) OR (knee N3 osteoarthritis)) AND (exacerbation OR flare OR (pain AND (diary OR diaries)) OR (pain N3 variab*) OR (pain N3 pattern$) OR (daily N3 pain)). A database search strategy is included in the online supplement . Reference lists of all included full text articles retrieved for detailed examination were manually searched.

Studies were included in the final full text peer review if they contained a description or definition of an acute exacerbation or flare-up of knee OA in human adults (18 years or over) in the general population, primary care or hospital settings. Studies were included even if their description was not based on clear measurement criteria (e.g. stating a 'significant increase in pain' but not the amount of change on a pain score this would equate to). Studies that included a mixed OA population (e.g. knee or hip OA) and did not separately report knee-specific findings were included. There were no restrictions on study dates or design. All non-English language articles were translated to identify a flare definition. Theses, dissertations, book chapters and guidelines, and animal studies were excluded. Conference abstracts were included if they contained a definition for an OA flare-up. Studies were excluded if the flare was induced by an iatrogenic source, for example, injection-induced flares[22]. As these may have been caused by a different pathophysiological process. Abstracts were included in this study as the main outcome of interest was the definition of flare used and it was decided that including abstracts would ensure a more comprehensive review. For each abstract a search was conducted to identify a corresponding full text paper. Where one was found only the full paper was included in the review.

The search and article retrieval was conducted by the first reviewer (ELP). Articles were downloaded into RefWorks© bibliography and database manager (RefWorks Copyright 2009). Duplicates were removed and all titles were screened by ELP

against inclusion criteria, with the first 20 titles checked by two reviewers (ELP and MJT) for consistency. For qualitative studies, all identified potentially eligible full text articles were obtained.

All abstracts and then full text articles were screened by two reviewers (ELP and MJT). with disagreements resolved by consensus adjudicated by a third reviewer (GP). Where articles could not be retrieved or if the flare definition used was not included in the text, contact with authors was made.

The final included articles were checked to ensure results were not duplicated, for example, where different authors were reporting on the same dataset, to reduce bias [23] . For articles containing pooled studies, the original studies were sought and included in the main analysis, where available.. No full text articles were required to be translated.

Data extraction

The following data pertaining to flares were extracted from full text articles by the first reviewer: definition used for change in pain, pain scale used, duration of flare (for flare design trials we extracted the duration of the withdrawal period for

comparison), associated symptoms, rationale behind definition used, terminology used (e.g. exacerbation or flare), baseline OA severity, age range, gender, geographical location, number of participants and study design. Missing data was described in the data extraction tables. Extraction for every tenth article was independently checked (MJT).

Quality assessment of included studies

Our aim was to identify and contrast definitions of flare-ups used in the literature. We were not concerned with the methodological rigour of the studies deriving, evaluating or applying those definitions. However, for studies presenting definitions we sought supporting statements that gave the rationale for the definition.

Data analysis

A narrative synthesis was undertaken, guided by Popay et al's[24] four stage process to develop a conceptual framework[25]. This approach was chosen as it allowed the words and text in the definitions to be synthesised to summarise findings[24]. The initial data extracted was grouped into drug withdrawal studies ('flare design') and other studies, and frequencies of components included in definitions was tabulated,

these included; terminology used, onset/worsening of symptoms; signs/symptoms above day-to-day variability/minimum threshold; speed of onset of symptoms; duration of worsening and change in medication/healthcare usage.

This initial tabulation helped identify similarities and differences and allowed themes to emerge. This was done with an inductive type approach, where possible i.e. without an *a priori* assumption, but also deductively acknowledging that the reviewers were clinicians i.e. they had some background knowledge of the topic of interest. This allowed further examination of the differences of definitions used in drug withdrawal and non-drug withdrawal study designs, and examination of key components of definitions used.

RESULTS

Study selection

The literature search yielded 2194 articles of which 786 were duplicates (Figure 1). After title screening 336 abstracts were reviewed, 223 were not relevant for the study purpose. 113 articles were examined in full which resulted in a further 60 being excluded. The main reason for exclusion was no definition of flare-up reported in text

(n=56). At this stage a further 16 articles were identified from the reference lists of the retrieved full text articles resulting in 69 included studies for synthesis.

Study characteristics

Characteristics of the included studies are described in Table 1[20, 26-93]**.** The number of participants in each study ranged from 15-6085[20, 50]. Knee OA was defined by clinical and/or radiological criteria.

214 **Table 1: Characteristics of all included studies**

DRUG WITHDRAWAL DESIGN STUDIES					
First author, year of publication	**Setting, geographic location**	**Participants**	**Joint**	**Severity**	**Study design**
Altman, 2015[26]	Multi-centre, recruitment not specified, USA	403 males & females, ≥40y	Knee and hip	KL grade 2-3	RCT, flare design
Baer, 2005[28]	17 medical centres recruiting from community and physician private practice; Canada	216 males & females, 40-85y	Knee	Radiographic evidence of OA (severity not defined)	RCT, flare design
Baraf, 2011[29]	Primary care, internal medicine, orthopaedic, rheumatology; USA	602 males & females, ≥25y	Knee	Radiographically mild to moderate (KL grade 1-3)	RCT, flare design
Battisti, 2004[30]	Clinical centres, out patients; USA	3980 males & females, ≥40y (age unavailable for Geba 2003 and Weaver 2003)	Knee	ACR functional class rating of I,II or III	RCT, pooled 4 trials, flare design
Bingham, 2007[31] Bingham 2011[77]	2x74 outpatient clinics; USA	1207 males & females, ≥40y	Knee and hip	ARA Functional capacity classification I-III	RCT, flare design

Birbara, 2006[32]	Investigative sites; USA	808 males & females, ≥40y	Knee	ARA functional class, I, II, or III	RCT, flare design
Bocanegra, 1998[33]	Clinic; USA	572 males & females, 28-88y (mean 61-62)	Knee and hip	ARA Functional capacity classification I-III	RCT, flare design
Boswell, 2008[34]	50 centres (Europe & Australia) + 187 centres (Europe & USA)	1908 males & females, ≥40y	Knee	KL scale 2 or 3 and ARA class rating of I,II or III	Pooled RCTs (2; one flare design, one non-flare), flare design
Brandt, 2006[35] (pilot studies)	Community; USA	30 males & females, mean age 62y	Knee	KL ≥2	Cohort design, flare design
Case, 2003[36]	Hospital-rheumatology centre; Chicago, USA	82 males & females, 40-75y	Knee	KL ≥1, and clinical criteria (pre-enrolment ambulatory pain; moderate pain by a 5-point Likert scale or increased pain.	RCT, flare design
Day, 2000[75]	49 investigative sites in 26 countries	809 males & females, mean age range 62-65y	Knee and hip	ARA functional class I-III, symptomatic for at least 6 months	RCT, flare design
Ehrich, 1999[37]	Clinical centres; USA	219 males & females, >40y	Knee	ARA functional class, I, II, or III	RCT, flare design
Essex, 2012[38]	Clinical centre; African-American, USA	322 males & females, ≥45y	Knee	ARA Functional capacity classification I-III	RCT, flare design
Essex 2013[78]	Hispanic population, 31 US centers	≥45y	Knee	ACR criteria, Functional capacity classification I-III	RCT, flare design
Gibofsky, 2014[39]	Not specified, USA	305 males & females, 41-90 y	Knee and hip	KL 2-3	RCT, flare design
Gineyts, 2004[40]	Subset of larger study; France	201 males & females, mean age 61-62y	Knee and hip	ARA I-III	RCT, flare design

Goldberg, 1988[41]	Investigative sites; USA	214 males & females, 40-85y (mean 64)	Knee and hip	Radiographic evidence of knee OA-not further defined	RCT, flare design
Gottesdiener, 2002[42]	Investigative sites; USA	617 males & females, ≥40y	Knee	ARA functional class I,II,III	RCT, flare design
Hochberg, 2011[43]	Centres; USA	1234 males & females, ≥50y	Knee	ACR functional class I-III	Pooled RCTs (2), flare design
Katz, 2010[44]	Clinical sites; USA	113 males & females, 28-83y (median 57))	Knee and hip	OA of hip and knee as diagnosed using ACR criteria-no definition of severity	RCT, flare design
Kivitz, 2001[45]	Investigative sites; USA	491 males & females, 28-91y (mean 58-61)	Knee	Confirmation of OA on weight bearing radiograph- no definition of severity	RCT, flare design
Kivitz, 2004[76]	Outpatient sites; USA	1042 males & females, ≥40y	Knee	ACR rating of I, II, III.	RCT, flare design
Leung, 2002[47]	Clinic; USA	677 males & females, ≥40y	Knee and hip	ARA functional class, I, II, or III	RCT, flare design
Luyten, 2007[47]	Centres; Belgium	181 males & females, ≥40y	Knee and hip	ACR Functional capacity classification I-III	RCT, flare design
Manicourt, 2005[49]	Outpatient clinic; Belgium	90 males & females, 50-81y (mean 63-67)	Knee and hip	Clinical and radiographic evidence of OA-severity not defined.	RCT, flare design
Mazzuca, 2002[50]	Not specified, USA	15 males & females, ≥45y	Knee	KL 2-3	Observational, flare design
McIlwain, 1989[51]	Investigative sites; USA	139 males & females, mean 65y	Knee	Radiological evidence of moderate or severe osteoarthritis- not further defined	RCT, flare design
Mendelsohn, 1991[52]	Investigative sites; USA	139 males & females, 21-88y (mean age 63.3y)	Knee	Radiological evidence of moderate or severe osteoarthritis- not further defined	RCT, flare design

Moskowitz, 2006[53]	Investigative sites; USA	530 males & females, ≥45y	Knee	ACR Functional capacity classification I-III	RCT, flare design
Pareek, 2009[54]	Multi-centre study, India	199 males & females, 40-70y	Knee	Lequesne criteria-score of 5 and above	RCT, flare design
Pareek, 2010[55]	Hospital; India	220 males & females, 40-70y	Knee	Clinical and radiological evidence of OA- severity not defined.	RCT, flare design
Roth, 2004[90]	Physicians private practice or community; USA	326 males & females, 40-85y	Knee	Radiological evidence of OA- severity not defined.	RCT, flare design
Rother, 2007[93]	Outpatient units; Germany	397 males & females, ≥40y	Knee	KL 2-3	RCT, flare design
Schnitzer, 2005[57]	Investigative sites; International (7 countries)	583 males & females, 18-75y	Knee and hip	Diagnosis based on ACR criteria- severity not defined.	RCT, flare design
Scott-Lennox, 2001[58]	Investigative sites; USA	182 males & females, mean 61y	Knee	Not defined	RCT, flare design
Silverfield, 2002[59]	Centres; USA	308 males & females, 35-75y	Knee and hip	Clinical evidence of OA- severity not defined	RCT, flare design
Simon, 2009[91]	Outpatient centres; Canada, USA	775 males & females, 40-85y	Knee	Clinical and radiological evidence of OA- severity not defined	RCT, flare design

Strand, 2011[60]	Investigative sites; Multinational-not specified including USA	875 males & females, 18-80y	Knee and hip	OA according to ACR criteria and requiring NSAID treatment to control symptoms in the month preceding screening	RCT, flare design
Weaver, 1995[92]	Investigative sites; USA	328 males & females, >50y	Knee	ACR clinical criteria-diagnostic	RCT, flare design
Wiesenhutter, 2005[61]	Medical Centres; USA	528 males & females, 40-89y	Knee and hip	ARA functional class, I, II, or III	RCT, flare design
Williams, 2001[62]	Clinical sites; USA	718 males & females, mean 61-62y	Knee	ACR clinical and radiographic criteria I-III	RCT, flare design
Wittenberg, 2006[63]	Centres (not specified) ; Germany	364 males & females, 50y	Knee	Moderate to severe symptomatic OA of the knee according to ACR criteria.	RCT, flare design
Yeasted, 2014[64] (Pooled, abstract)	USA	219 (merged observational), 137 (merged trial)>40y	Not specified	ACR criteria-diagnostic	2 longitudinal observational studies, placebo arms of 2 clinical trials
Yocum, 2000[79]	USA, 62 study centres	774 males & females, ≥40y	Knee or hip	Diagnosis confirmed by XR and clinical symptoms (not further specified)	RCT, flare design
Young, 2014[65] (abstract)	Multicenter,	305 males & females, >40y	Knee or hip	KL 2-3	RCT, flare design
Zhao, 1999[66]	Centre (not specified); USA, Canada	1004 males & females, ≥18y	Knee	ACR Functional capacity classification I-III	RCT, flare design
NON-DRUG WITHDRAWAL DESIGN STUDIES					

Atukorala, 2016[80] (abstract) Atukorala, 2016[27] (abstract)	Not specified, USA + Australia + Sri Lanka	213 males & females, mean age 62y 345 males & females, mean age 62y	Knee	Not specified	3-month, web based longitudinal follow up study
Bartholdy, 2016[81]	OA out-patient clinic, Denmark	131 males & females, ≥40y	Knee	Radiographic evidence of OA (severity no defined) and BMI between 20-35 kg/m^2	RCT
Bassiouni 2015[82] (abstract)	Not specified, Egypt	60 participants not further specified	Knee	Not specified	Observational
Cibere, 2004[88] Cibere, 2005[89]	Community, Canada	137 males & females, mean age 65y (43-88) for placebo and 64y (40-83) for glucosamine group	Knee	KL ≥2 on anteroposterior radiograph	RCT
Conrozier 2012[68]	Hospital-rheumatology unit, France	44 males & females, mean age 67.6y	Knee	Radiographic evidence of knee OA-not further defined	Observational
D'Agostino 2005[69]	Hospital-European multicentre	600 males & females, ≥18y	Knee	KL grade 1-4	Observational
Erfani, 2014[46] abstract) Erfani, 2014[83] (abstract) Ferreira[84] 2016	Australia	268 males & females, mean age 62y 345 males & females, ≥40y	Knee	ACR criteria- meet at least one, KL ≥2	Web based cross over

Hunter 2014[85] (abstract) Makovey 2015[86] (Protocol)					
Jawad, 2005[70]	GPs in France	3000 (for GP study) males & females	Knee	Not defined	n/a, review of surveys. Definition relates to survey of 3000 French GPs
Marty 2009[20]	Community and hospital, France	6085+641males & females, mean age 66.4y (10.9) for flare group, 66.2y (10.2) no flare group	Knee	OA diagnosis based on ACR criteria- severity not defined	Observational
Murphy, 2015[71]	Community based, pain clinics; USA	45 males & females, 37-83y	Knee	ACR criteria- severity not defined	Qualitative
Parry, 2017[87]	Community, UK	719 males & females, ≥50y	Knee	Self-reported knee pain in previous 12 months	Observational
Ricci 2005[56]	Community, USA	329 males & females, 40-65y	Knee and hip	Clinical evidence of OA- severity not defined	Nested case control
Wise 2010[72]	Primary care, hospital, USA	303 males & females, ≥50y	Knee and hip	Signal joint pain in a hip or knee on at least 15 out of the 30 days prior to enrolment- not further defined	Observational

Zhang 2009[73]	Primary care, hospital, USA	303 males & females, ≥50y	Knee and hip	Signal joint pain in a hip or knee on at least 15 out of the 30 days prior to enrolment-not further defined	Observational
Zhang 2011[74] (abstract)	Not specified	52 males & females, median age 63, (50-72y)	Knee	KL>2	Case-crossover
Zobel, 2016[94]	Hospital databases, Australia	297 males & females, >40y	Knee	ACR criteria, KL ≥2, or patellofemoral OA on radiograph	Web based case-cross over
Acronyms: KL- Kellgran and Lawrence RCT- Randomised Controlled Trial USA- United States of AmericaACR- Arthritis Center Research ARA- American Rheumatism Association GP- General Practitioner					

215

216

Twenty-one included mixed knee and hip OA groups[26, 31, 33, 39-41, 44, 47-49, 56, 57, 59-61, 65, 73, 75, 77, 79]. In total, 46 publications used a drug withdrawal RCT design[26, 28-34, 36-45, 47-55, 57-66, 75-79, 90-93], four of which were pooled studies[30, 34, 43, 64] and one used a cohort drug withdrawal design[35] (Table 1). The remaining 22 publications included seventeen observational studies[20, 27, 46, 56, 67-69, 72-74, 80, 82-87], three RCTs[81, 88, 89], one survey[70] and one qualitative interview study[71]. Nine of the included studies were abstracts[27, 46, 64, 65, 74, 80, 82, 83, 85]. Two abstracts were removed as the corresponding full text article was available[71, 94]. Studies using pooled data or the same dataset were included if they used different definitions of OA flare[30, 46, 54, 55, 64, 67, 72, 73, 76].

Rationale given for flare definitions

Six of the included studies gave rationale for the definition used[20, 56, 58, 71, 87, 88]. None of the definitions were based on a consensus procedure. Marty et al[20] and Scott-Lennox et al[58] were the only studies that undertook empirical investigation of flare definitions. The study by Marty et al[20] was the only study specifically designed to validate a diagnostic tool for knee OA flares. Potential factors associated with flare-ups were identified, for example, knee swelling and the authors used a logistic regression analysis to assign a weight to each of the items identified. A flare up score

was determined using a general practitioner database and this was then validated using a rheumatologist database. Pain was not included in the final model.

Scott-Lennox et al[58] sought to test whether four measures for flare intensity (patient's self-assessment of pain scores, physician's assessment of pain scores, patient's global OA assessment and physician's global OA assessment) could be combined to form a reliable and valid index using data from an RCT using a confirmatory factor analysis. The authors produced three flare intensity groups (low, moderate and severe) and highlighted how these could be used to examine treatment effects.

Cibere[88] outlined face validity checks. It was specified that the flare definition had been determined by study rheumatologists to be a clinically important change in the WOMAC score. The definition used by Murphy et al[71] was informed by two studies[30, 55] which used a drug withdrawal design and from the research team's own experience. Ricci et al[56] used a combination of data-driven and clinical judgement approaches to establish an agreed cut point. Parry et al based their definition on OA flare design studies and flare definitions used in other chronic disease such as back pain and COPD.

Flare definitions in drug withdrawal studies

Terminology used

The majority of publications using a drug withdrawal design used the term “flare” in their description[26-32, 34, 35, 38-45, 47-51, 53, 55, 57-66, 76-79, 90-93] (n=42; Table 2).

Table 2: Definition, terminology and measurement instruments used in all included studies

DRUG WITHDRAWAL STUDY DESIGN							
First author	**Terminology used**	**Amount of change in symptoms/signs (symptom/sign: measurement instrument; operational definition)**	**Minimum absolute level of symptoms/signs (symptom/sign: measurement instrument; operational definition)**	**Speed of onset**	**Duration**	**Change in medication/healthcare use**	**Reference / rationale**
Altman, 2015[26]	"Flare"	**Pain**: WOMAC Pain subscale (0-100); increase ≥15mm	**Pain:** WOMAC Pain subscale; ≥40mm	Not specified	Not specified	Not specified	None
Baer, 2005[28]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale (0-20); increase ≥2 points and ≥25%	**Pain:** WOMAC Pain score (0-20); ≥6 and ≥1 item rated 'moderate, severe, or extreme'	Interval between screening and baseline re-measurement unclear	Not specified	Not specified	None
Baraf, 2011[29]	"Flare"	**Pain on movement:** VAS (0-100mm); increase ≥5mm	Not specified	1 week washout	Not specified	Not specified	None
Battisti, 2004[30]	"Flare"	**Global assessment (investigator):** single item, 5-point LK; Worsening ≥1 point	**Pain:** VAS (0-100mm); ≥40mm	Not specified	Not specified	Not specified	None

Bingham, 2007[31] Bingham 2011[77]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); increase ≥15mm (2) **Global assessment of disease status (investigator):** single item, 5-point LK; Worsening ≥1 point	(1) **Pain walking on flat surface:** ≥40mm on WOMAC VAS3.0 Q1 (0-100) (2) **Global assessment (investigator)**: single item, 5-point LK; fair, poor, very poor (*acetaminophen users only*) (3) **Global assessment of disease status (patient):** VAS 0-100mm; ≥40mm (*acetaminophen users only*)	Not specified	Not specified	Not specified	None
Birbara, 2006[32]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global assessment (investigator):** single item, 5-point LK; Worsening ≥1 point	(1) **Pain walking on flat surface:** WOMAC VAS3.0 Q1 (0-100); ≥40mm (2) **Global assessment (investigator):** single item, 5-point LK; Fair, poor or very poor (*paracetamol arm only*)	4-15 day washout	Not specified	Not specified	None
Bocanegra, 1998[33]	"Worsening of symptoms"	Two out of the following three: (1) **Global assessment (physician):** single item, 5-point LK; Increase ≥1 grade (2) **Global assessment (patient):** Patients global assessment (current symptoms and limitation of activity) 5-point LK; Increase ≥1 grade (3) **Composite index:** Lequesne OA Severity Index (0-24); Increase ≥2 points	(1) **Global assessment (physician):** single item, 5-point LK; 'poor/very poor' (2) **Global assessment (patient):** Patients global assessment (current symptoms and limitation of activity) 5-point LK; 'poor/very poor' (3) **Composite index:** Lequesne OA Severity Index (0-24); ≥7	3-14d washout	Not specified	Not specified	None

Boswell, 2008[34]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global assessment (patient):** Patient Global Assessment of Arthritis Condition (PGAC) (unspecified); Worsening ≥1 point	Not specified	Not specified	Not specified	Not specified	None
Brandt, 2006[35] (pilot studies)	"Flare"	Not specified	**Pain:** WOMAC LK Pain subscale (5-25); ≥15 points	5 half-lives of NSAID washout	Not specified	Not specified	None
Case, 2003[36]	Not used	(1) **Pain walking on flat surface:** VAS (0-100mm); Increase ≥10mm (2) **Ambulatory pain**; 5-point LK; worsening ≥1 point	Not specified	14d washout	Not specified	Not specified	None
Day, 2000[75]	Not used	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global Assessment (investigator):** single item, 5-point LK; worsening ≥1 point (3) **Global assessment (patient):** VAS (0-100mm); increase ≥15mm (*acetaminophen users only*)	(1) **Pain walking on a flat surface**: WOMAC VAS Q1 (0-100mm); ≥40mm; (2) **Global Assessment (investigator):** single item, 5-point LK; 'Fair, poor, or very poor'; (3) **Global assessment (patient):** VAS (0-100mm); ≥40mm	Longer than 5 plasma half-lives washout	Not specified	Not specified	None
Ehrich, 1999[37]	Not used	**Pain:** VAS (0-100mm); increase ≥15mm	**Pain:** VAS (0-100mm); ≥40mm	Longer than 5 plasma half-lives washout of NSAID	Not specified	Not specified	None

Essex, 2012[38]	"Flare"	(1) **Global Assessment (Physician):** 5-point LK; increase ≥1 grade (2) **Global Assessment (patient):** 5-point LK; increase ≥1 grade	(1) **Global Assessment (Physician):** 5-point LK; 'Fair, poor or very poor' (2) **Global Assessment (patient):** 5-point LK; 'Fair, poor or very poor' (3) **Pain**: VAS (0-100mm); 40-90mm	48 hour withdrawal	Not specified	Not specified	None
Essex 2013[78]	"Flare"	Not specified	(1) **Global Assessment of arthritis (Physician):** Minimum rating of 3 (2) **Global Assessment of arthritis (patient):** Minimum rating of 3 (3) **Pain**: VAS (0-100mm); 40-90mm	48 hour withdrawal	Not specified	Not specified	None
Gibofksy, 2014[39]	"Flare"	**Pain:** WOMAC Pain VAS; increase ≥15mm	**Pain:** WOMAC Pain VAS; ≥40mm	Not specified	Not specified	Not specified	None
Gineyts, 2004[40]	"Flare"	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2)**Global Assessment (investigator):** 5-point scale: worsening ≥1 point	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); ≥40mm	5 half-lives of NSAID washout	Not specified	Not specified	None
Goldberg, 1988[41]	"Flare"	(1) **Pain:** Investigator assessed pain grade (None/mild/mod/severe): (i) at rest, (ii) on passive motion, (iii) on palpation, (iv) weight bearing; increase ≥1 grade in two items OR increase ≥2 grade in one item	Not specified	2-14 day washout until flare	Not specified	Not specified	None
Gottesdiener, 2002[42]	"Flare"	(1) **Pain on walking**: VAS (0-100mm); increase ≥15mm (2)**Global Assessment (Investigator):** 5-point LK; Increase ≥1 point	(1) **Pain on walking**: VAS (0-100mm); ≥40mm	3-15 day washout	Not specified	Not specified	None

Hochberg, 2011[43]	"Flare"	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm)**;** Increase ≥15mm (2) **Global Assessment (patient):** 5-point LK; worsening ≥1 point	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm)**;** ≥40mm	Not specified	Not specified	Not specified	None
Katz, 2010[44]	"Flare"	Not specified	**Pain**: Pain score (0-10); ≥5	Not specified-washout until flare occurred	Not specified	Not specified	None
Kivitz, 2001[45]	"Flare"	**Pain:** Patients Assessment of Pain Score (0-10) (unspecified); increase ≥2 points	**Pain:** Patients Assessment of Pain Score (0-10) (unspecified); ≥5	5 drug half-lives or 48 hours	Not specified	Not specified	None
Kivitz, 2004[76]	"Flare"	(1) **Pain on walking:** VAS (0-100mm); worsening ≥15mm (2) **Global Assessment (investigator):** 5-point LK; worsening ≥1 point	Not specified	NSAID dependent half-life washout	Not specified	Not specified	None
Leung, 2002[47]	"Flare"	(1) **Pain on walking on a flat surface:** WOMAC VAS Q1 (0-100mm); Increase ≥15mm (2) **Global Assessment (investigator):** 5-point LK; worsening ≥1 point	(1)**Pain on walking on a flat surface:** WOMAC VAS Q1 (0-100mm); ≥40mm (2) **Global Assessment (patient):** **(**0-100mm); ≥40mm (*acetaminophen users only*) (3) **Global Assessment (investigator):** 5-point LK; 'Fair, poor, or very poor' (*acetaminophen users only*)	Determined by drug half-life washout	Not specified	Not specified	None

Luyten, 2007[48]	"Flare"	(1) **Global Assessment (Patient):** 5-point LK; Increase ≥1 grade (2) **Global Assessment (physician):** 5-point LK; increase ≥1 grade (3) **Composite definition:** Lequesne Osteoarthritis Severity Index (0-24); increase ≥2 points	(1) **Global Assessment (Patient):** 5-point LK; 'Fair, poor or very poor' (*Not on treatment – 'Poor or very poor'*) (2) **Global Assessment (physician):** 5-point LK; 'Fair, poor or very poor' (*Not on treatment – 'Poor or very poor'*) (3) **Composite definition:** Lequesne Osteoarthritis Severity Index (0-24); ≥7 (4) **Pain**: VAS (0-100mm); ≥40mm	2-14 day washout	Not specified	Not specified	None
Manicourt, 2005[49]	"Flare"	**Pain when walking on a flat surface:** VAS (0-100mm) ; ≥10mm	Not specified	7-10 days washout	Not specified	Not specified	None
Mazzuca, 2002[50]	"Flare"	**Pain on standing**: WOMAC LK Pain Q5 'severe or extreme' after the washout AND decreased after resumption of usual analgesic drugs and/or NSAIDs	Not specified	Drug washout 5 half lives	Not specified	Not specified	None
McIlwain, 1989[51]	"Flare"	No measurement instrument: Increase in pain on motion, swelling, tenderness, redness and/or heat (unspecified if patient/physician/investigator reported)	Not specified	2-14 day washout	Not specified	Not specified	None
Mendelsohn, 1991[52]	"Worsening of arthritis condition"	(1) **Pain:** Pain scale (0-3) (0=none, 3=severe); worsening score (2) **Global (physician):** (0-100); worsening score	Not specified	Up to 14 days washout	Not specified	Not specified	None

Moskowitz, 2006[53]	"Flare"	(1) **Global assessment (patient)**: 5-point LK; increase ≥1 grade (2) **Global Assessment (physician)**: 5-point LK; ≥ 1 grade increase (3) **Composite index:** Lequesne OA Severity Index (0-24); increase ≥2 points	(1) **Global assessment (patient):** 5-point LK; '(Fair), poor, or very poor' (2) **Global Assessment (physician):** 5-point LK; '(Fair), poor or very poor' (3) **Composite index**: Lequesne OA Severity Index (0-24); Minimum ≥7 (4) **Pain walking on a flat surface:** VAS (0-100mm); ≥40mm	NSAID washout of 5 half-lives or at least 2 days	Not specified	Not specified	None
Pareek, 2009[54]	"Flare-up"	(1) **Pain:** 11-point NRS; increase ≥ 2 points during previous 2-5 days (2) Signs and symptoms suggestive of inflammation, morning stiffness and nocturnal pain interfering with sleep	**Pain**: Pain intensity of at least 4 on a 11-point NRS during physical activity for past 24 hours	Placebo washout for 24-48 hours	2-5 days	Not specified	None
Pareek, 2010[55]	"Flare"	Flare symptoms noted but not part of definition: morning stiffness, erythema, nocturnal pain, and swelling/inflammation	(1) **Pain with physical activity**: VAS 0-10; ≥6 (2) **Composite index:** WOMAC Total LK; ≥25. (3) **Composite index**: Lequesne OA Severity Index (0-24); ≥5	Not specified	2-5 days	Not specified	None
Roth, 2004[90]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale (0-20); increase ≥2 points and ≥25%	**Pain:** WOMAC LK3.1 Pain subscale (0-20); Score ≥'moderate' on at least 1 of the 5 items, (ii) Pain score ≥6	Washout period of at least 3 days per week past month	Not specified	Not specified	None

Rother, 2007[93]	"Flare"	(1) **Pain on walking:** VAS (0-100mm); Increase ≥15mm (2) **Global Assessment (patient):** 5-point LK; increase ≥1 grade	(1) **Pain on walking:** VAS (0-100mm); ≥40mm (2) **Global Assessment (patient):** 5-point LK; 3-5	Not specified	Not specified	Not specified	None
Schnitzer, 2005[57]	"Flare"	No tool: increase in pain	**Pain:** VAS (0-100mm); ≥40mm	Not specified	24 hours	Not specified	None
Scott-Lennox, 2001[58]	"Flare"	(1) **Pain:** VAS (0-100mm)**;** ≥20mm (2) **Pain (physician):** 4-point LK; worsening ≥1 point (3) **Global Assessment (patient):** 4-point LK; worsening ≥1 point (4) **Global Assessment (physician)**:4 point LK; worsening ≥1 point	(1) **Pain:** VAS (0-100mm)**;** ≥40mm at baseline) (2) **Pain (physician):** 4-point LK; ≥2 (3) **Global Assessment (patient):** 4-point LK; ≥2 (4) **Global Assessment (physician)**: 4 point LK; worsening ≥2	14 day washout	Not specified	Not specified	Confirmatory Factor Analysis
Simon, 2009[91]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale; increase ≥2 and ≥25%	**Pain:** WOMAC LK3.1 Pain subscale; ≥'moderate' on ≥1 item	14 day washout	Not specified	Not specified	None
Silverfield, 2002[59]	"Flare"	**Pain:** No measurement tool; significant increase	Not specified	Not specified	Not specified	Pain requiring supplemental analgesic medication and/or an increase in NSAID dose	None
Strand, 2011[60]	"Flare"	**Global Assessment (patient):** 5-point LK; Increase ≥1	(1) **Global Assessment (patient):** 5-point LK; 'Fair, poor or very poor' (2) **Pain**: (0-10 NRS); ≥4 but <9 (3) **Global Assessment (physician)**: 5-point LK; 'Fair, poor or very poor'	14 day washout	Not specified	Not specified	None

Weaver, 1995[92]	"Flare"	(1) **Global Assessment (Physician)**: 5-point Likert; increase ≥1 grade (2) **Global Assessment (patient)**: 5-point LK; increase ≥1 grade (3) **Pain:** Worsening pain on motion and weight bearing	(1) **Global Assessment (Physician)**: 5-point Likert; ≥2 (2) **Global Assessment (patient)**: 5-point LK; ≥2	2-14 day washout	Not specified	Not specified	None
Wiesenhutter, 2005[61]	"Flare"	(1) **Pain on walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); increase ≥15mm (2) **Global Assessment (Investigator):** 5-point LK; worsening ≥1 unit	(1) **Pain on walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); ≥40mm	Not specified	Not specified	Not specified	None
Williams, 2001[62]	"Flare"	(1)**Global Assessment (patient):** 5-point LK; Increase ≥1 point (2) **Global Assessment (physician)**: 5-point LK; increase ≥1 point(3) **Composite Index**: Lequesne OA Severity Index (0-24); Increase ≥2 points	(1) **Global Assessment (patient):** 5-point LK; '(Fair), poor or very poor' (2) **Global Assessment (physician)**: 5-point LK; '(Fair), poor or very poor' (3) **Composite Index**: Lequesne OA Severity Index (0-24); ≥7 (4) **Pain:** VAS (0-100mm); ≥40mm	2-14 days	Not specified	Not specified	None
Wittenberg, 2006[63]	"Flare"	**Pain:** VAS (0-100mm); Increase ≥10mm	**Pain:** VAS (0-100mm); ≥40mm	2-7 day washout	Not specified	Not specified	None
Yeasted, 2014[64] (Pooled, abstract)	"Flare"	**Pain:** 0-10 NRS; Increase ≥2 points over the mean pain score from the previous 3 days	**Pain:** Average daily 0-10 NRS; 4-9	Not specified	Not specified	Not specified	None
Yocum 2000[79]	"Flare"	**Disease activity** (1) **Global (Investigator)**: Reduction of ≥ 1 grade (2) **Global Assessment (Patient)**: 100-mm VAS; Increase of ≥10mm	Not specified	≥3 days washout	Not specified	Not specified	None

		(3) **Pain: Overall assessment (patient):** 100-mm VAS; ≥35mm					
Young, 2014[65]	"Flare"	(3) **Pain:** WOMAC pain subscale; increase >15mm	**Pain:** WOMAC Pain subscale >40mm	Not specified	Not specified	Not specified	None
Zhao, 1999[66]	"Flare"	No measurement tool: Worsening of signs and symptoms after discontinuation of NSAIDs of analgesics	Not specified	2-7 day washout	Not specified	Not specified	None
NON-DRUG WITHDRAWAL STUDY DESIGN							
Atukorala, 2016[80] (abstract) Atukorala, 2016[27] (abstract)	"Flare"	**Pain:** (10-point NRS); increase >2 points from the mildest knee OA pain intensity reported at day 0	Not specified	Not specified	Not specified	Not specified	None
Bartholdy, 2016[81]	"Flare"	Not specified	**Pain:** (10-point NRS): Pain >5	Not specified	Not specified	Not specified	None
Bassiouni 2015[82] (abstract)	"Flare"	Not specified	**Global Assessment (physician):** KOFUS ≥7	Not specified	Not specified	Not specified	None
Cibere, 2004[88] Cibere, 2005[89]	"Flare"	(1) Patients perception of worsening of symptoms (2) **Pain walking on flat surface**: WOMAC VAS3.0 Q1 (0-100mm); increase ≥20mm (3) **Global Assessment (physician):** 5-point LK; worsening ≥1 grade	Not specified	Not specified	Not specified	Not specified	Definition determined by study rheumatologists to be a clinically important change in WOMAC-Ehrich2000/

							Bellamy 1998
Conrozier 2012[68]	"Flare"	Fulfilled 4 following criteria: (1) **Pain:** No measurement tool; 'sudden aggravation of knee pain' (2) causing nocturnal awakenings, (3) clinical evidence of effusion.	Not specified	Sudden aggravation of knee pain, whose beginning was identifiable	Not specified	Not specified	None
D'Agostino 2005[69]	"Flare"	Not specified	**Pain intensity during physical activity:** VAS-(0-100mm); ≥40mm	Not specified	48 hours	Not specified	None
Erfani, 2014[46] abstract) Erfani, 2014[83] (abstract) Ferreira[84] 2016 Hunter 2014[85] (abstract) Makovey 2015[86] (Protocol)	Exacerbation	**Pain:** VAS (0-100mm); Increase ≥20mm from mildest pain score reported at baseline	Not specified	Not specified	Not specified	Not specified	None

Jawad, 2005[70]	Exacerbation	**Pain symptoms:** Increased morning stiffness, night pain and synovial fluid effusion	Not specified	Not specified	Not specified	Not specified	None
Marty 2009[20]	"Flare"	No measurement tool: Morning stiffness >20mins, nocturnal awakening, limping, knee swelling, increased warmth, effusion	Not specified	Not specified	48 hours	Not specified	Regression analysis of cross-sectional data to validate proposed flare criteria
Murphy, 2015[71]	"Flare"	(1) Investigator definition: Inadequate pain relief for an episode of intense pain that is usually brought on by too much activity. (2) Participant definitions: Described in terms of pain quality, timing (onset and duration), antecedents and consequences. (3) **Pain magnitude:** increase in pain or 'intense' or 'severe' level of pain	**Pain:** ≥40 of 100mm or ≥4 of 10 on NRS	Patients described: 'Quick' or 'sudden'	Patients: 10 seconds to 15 minutes	Patients: Rest or take additional medication	For investigator definition: Battisti 2004, Pareek 2010. Plus researchers own experience.
Parry, 2017[87]	"Flare"	**Pain:** Recalled worst pain intensity in previous 6 months 0-10 NRS; ≥5	**Pain:** Recalled worse pain to be ≥2 points higher than recalled average pain (0-10 NRS) in previous 6 months	Not specified	Not specified	Not specified	Based on previous studies defining knee flares in OA and flares in diseases such as back

							pain and COPD.
Ricci 2005[56]	"Flare up"	**Pain:** Self-reported flare severity rating 0-10 NRS; increase ≥2 point over usual pain severity	Not specified	Not specified	Not specified	Not specified	Based on statistical analysis and clinical judgement
Wise 2010[72]	"Flare"	Not specified	**Pain:** WOMAC Pain subscale (0-10); score in highest 30% of all WOMAC scores	Not specified	Not specified	Not specified	None
Zhang 2009[73]	"Exacerbation or flare"	Not specified	(1) Pain: WOMAC pain subscale 0-10 (total score of 50 normalised to a 0-10 scale); score of ≥5, a score corresponding to highest 33% of all WOMAC scores	Not specified	Not specified	Not specified	None
Zhang 2011[74] (abstract)	"Exacerbation"	**Pain:** WOMAC Pain score VAS (0-500); increase ≥100 units	Not specified	Not specified	Not specified	Not specified	None
Zobel, 2016[94]	Exacerbation	**Pain:** 0-10 NRS; Increase ≥2	(1) Disabling pain	Not specified	8 hours	Not specified	None

Acronyms:
COPD- Chronic Obstructive Pulmonary Disease
KOFUS- Knee Osteoarthritis Flare-up Score
NRS-Numerical Rating scale
VAS- Visual Analogue Score
WOMAC- Western Ontario and McMaster Universities Osteoarthritis Index
LK-Likert scale

267

One study used the term "flare-up"[54], two studies referred simply to "worsening of symptoms" [33, 52] and three studies used no specific label[36, 37, 75].

Coverage of key components

Onset/worsening of symptoms and signs beyond normal-day-to-day variability: Forty-four studies included onset or worsening of signs and symptoms as part of their definition[26, 28-34, 36-43, 45, 47-55, 57-66, 75-77, 79, 90-93]. All studies included increased pain intensity in their definition. A further two[54, 55] specified further signs and symptoms. These included swelling, inflammation, erythema, morning stiffness and nocturnal pain. No studies quantified day-to-day variability.

Twenty-six measurement tools were used to define onset/worsening of symptoms and signs. The most commonly used tools were the Western Ontario & McMaster Universities Arthritis index (WOMAC) Q1 (pain on walking on flat surface) 100mm Visual Analogue Scale (VAS) (n=9)[31, 32, 34, 40, 43, 47, 61, 75, 77] and the Investigator Assessment of Disease Status (n=11)[30-32, 40, 42, 47, 61, 75-77, 79] (Table 3). Thirty-four studies used only single item measurement tools[29-32, 34, 36-45, 47, 49, 50, 52, 54, 57, 58, 60, 61, 63-65, 75-79, 92, 93], 5 used multi-item[33, 48, 53, 55, 62] and 5 used both single and multi-item tools[26, 28, 35, 91, 95].

Table 3: Summary of number and type of single and multi-item measurement tools used.

Single item scales:	
Pain on activity:	WOMAC Q1 3.0 VAS 'pain on walking on a flat surface' (0-100mm) [n=11] Pain on walking VAS (0-100mm) [n=5] Pain on movement VAS (0-100mm); Ambulatory pain (5-point Likert); Pain with physical activity VAS 11-point scale [n=2]
Pain (not further specified):	Pain VAS (0-100mm) [n=15] Patients Assessment of Pain Score (0-10); Pain Scale (0-3); Pain NRS (0-10) [n=11]
Standing knee pain	Item 5 WOMAC pain scale [n=1]
Global rating (physician/ investigator)	Investigator Assessment of Disease Status [n=11] Physicians Global Assessment of Arthritis [n=6] Physician Global Assessment of OA [n=2] Physician Global Assessment of Disease Status [n=2]; Investigator Assessed Pain Grade; (Physician) Overall Disease Activity (0-100); Physicians Pain Assessment (4-point LK) [n=3]
Global rating (patient)	Patients Global Assessment of Arthritis [n=7] Patient Global Assessment of OA [n=3] Patient Global Assessment of Disease Status [n=4]
Multiple-item scales:	
	Lequesne OA Severity Index [n=5] WOMAC LK3.1 (0-20) [n=3] WOMAC LK Pain subscale (0-25); WOMAC OA Index

	Questionnaire [n=1]; WOMAC knee pain score (0-500) [n=7]; KOFUS (0-14) [n=1]

N, number of included studies; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; VAS, visual analogue scale; OA, osteoarthritis; KOFUS, Knee Osteoarthritis Flare-up Score.

In addition, the format of global ratings appears to be variable as is use and reporting of the WOMAC[96]. However, despite the exact format of reporting being inconsistent, in general, studies used single items in 4 areas – pain on activity, pain (not necessarily on activity), physician/investigator global rating and patient global rating.

Temporal characteristics: None of the included drug withdrawal design studies reported a specific time for defining the speed of onset of symptoms. However, they did describe withdrawal or 'washout' periods whereby, after withdrawal of usual medication, participants were given a certain time frame in which to experience 'flare' symptoms in order that they were entered into the study. In total 30 of the studies specified a withdrawal period[29, 32, 33, 35-38, 40-42, 45, 47-54, 58, 60, 62, 63, 66, 75, 76, 78, 79, 90-92].

Four studies specified a time period for minimum duration of symptoms which ranged from 24 hours to 5 days[54, 55, 57, 59] .

Change in medication or healthcare usage: Only one study used increase in medication as part of their definition; 'pain requiring supplemental analgesic medication and/or an increase in NSAID dose'[59].

Additional domains: Thirty-six studies included a minimum threshold which was usually a minimum level of pain that was required before the participant was considered to have a flare[26, 28, 30-33, 35, 37-40, 42-45, 47-49, 53-55, 57, 58, 60-65, 75, 77, 78, 90-93]. There was general concordance with the minimum thresholds that different measurement tools used with a few exceptions. A threshold of 40mm on a 0-100mm scale was used in eight of ten studies using the WOMAC VAS 3.0 Q1 'pain on walking on a flat surface'[31, 32, 40, 43, 47, 61, 75, 77] and four of fourteen studies using the Patient Global Assessment of Disease Status[31, 47, 75, 77]. In studies using various forms of investigator/physician global assessment, the majority adopted a minimum threshold for a flare of 'fair, poor or very poor' [31, 32, 47, 75]. The minimum threshold on the Lequesne index (0-10) was either five[55] or seven[48, 53, 62].

Flare definitions in non-withdrawal flare/ discontinuation studies

Terminology used

"Flare" was the term most common used in non-withdrawal design studies[20, 27, 68, 69, 71, 72, 80-82, 87, 89](n=11) (Table 2). One study used the term "flare-up"[56], eight used "exacerbation"[46, 67, 70, 74, 83-86] (five publications were from the same team) and one referred to both "exacerbation" and "flare"[73]. None referred to "worsening of symptoms" or did not use any specific label.

Coverage of key components

Onset/worsening of symptoms and signs beyond normal-day-to-day variability:

Sixteen of twenty-two studies used onset or worsening of symptoms in their definition[27, 46, 56, 68, 70, 71, 74, 80, 83-89, 94]. Two studies did not use pain intensity as part of its definition[20, 82]. Three studies included symptoms other than pain in their definition[20, 68, 70]. These included nocturnal awakenings, effusion, morning stiffness, night pain, limping, and warmth.

The Murphy et al[71] study included an investigator definition of flare but also sought to describe patient experience of flares through face to face individual interviews. Both investigator and patient definitions included onset/worsening of symptoms and signs however there was no differentiation from day-to-day variability.

Seven studies used a measurement tool to define onset of signs and symptoms (Table 3). These included the Pain NRS (0-10)[27, 56, 67, 80, 87], WOMAC knee pain score VAS (0-500)[74], pain walking on a flat surface (WOMAC)[88, 89], Global Assessment of Disease Status (physician) (Likert 5-point scale)[88, 89], and knee pain VAS not further specified (0-100)[46, 83-86].

Temporal characteristics: Only one study set a definition for speed of onset, describing this only as 'sudden' with no further specification[68]. Patients in the Murphy et al study used the terms 'quick' and 'sudden' to describe flare onset[71]. Three studies specified a minimum duration of symptoms ranging from 8 to 48 hours[20, 67, 69]. In the Murphy et al study patients described duration of between 10 seconds to 15 minutes[71].

Change in medication/healthcare usage: No studies used change is medication or healthcare usage as part of their definition. However, in Murphy et al patients reported either taking rest or using additional medication[71].

Additional Domains: Two studies defined distribution-based minimum thresholds for flare as the highest 30%[72] or highest 33%[73] of WOMAC Pain Subscale scores among participants in the Longitudinal Examination of Arthritis Pain (LEAP) cohort (total score out of 50 was normalised to a 0-10 scale).

DISCUSSION

Flares in OA are recognised in existing clinical guidance[97] and reviews[98, 99] but typically merit little more than a passing mention. The recently published review that

sought to define flare-ups in in hip and knee OA only yielded 23 studies and four of the included studies did not contain clear definitions for a flare-up[21]. Furthermore, our analysis of the definitions has resulted in the findings of common core domains which will be useful for developing an agreed consensus definition for OA flare. From a clinical perspective, a unified definition of a flare could enable clinicians to provide prompt, rationalised and focussed treatment. This could also have implications for delivery of self-management strategies involving patients and how episodic management is advocated by clinical guidelines. Our review was motivated by an interest in seeking greater clarity on how these phenomena might be defined by undertaking a broad search strategy, noting that similar efforts have been pursued in other chronic diseases. While we found no current single, agreed definition of OA flare, our review of 69 published studies suggests a number of common domains which may capture cardinal features. These were: onset/worsening of symptoms and signs, attainment of a minimum symptom threshold during flare, speed of onset/worsening, and duration of elevated symptoms/signs. However, we found considerable variation in how these domains have been operationalised for measurement suggesting the need for further conceptual clarification and consensus.

Each potential cardinal feature of OA flare presents different challenges for achieving consensus. The goal of an agreed composite definition is to facilitate both reproducible and comparable research, whilst enabling more consistent recognition

and identification of these phenomena in routine practice. The heterogeneity of OA should also be considered in any definition of a flare-up Most studies included in our review required an increase in pain over 'usual' or 'baseline' intensity. Although this was measured using a wide range of measurement instruments several studies selected an increase of 2 or more points on a 0-10 scale providing a possible starting point for consensus. Yet this possible 'signal' is arguably difficult to interpret without also considering the amount of background 'noise', i.e. within-person diurnal[100] and day-to-day variability[101], and the absolute level ('minimum threshold') of pain during a flare. There was general concurrence with the minimum threshold that was adopted, for example, 40mm on a 0-100mm scale and this may indicate the potential level of minimally important clinical difference. In the study by Marty et al an increase in pain was not independently associated with flare-up after adjusting for other potential features[20]. However, the study by Marty et al[20] and Scott-Lennox et al[58] were the only two studies we found that had attempted to derive and/or validate a prediction model for OA flares. Interestingly their approaches have not been widely adopted which suggests the complexity of reaching a widely accepted model. Further research on detecting flares over within-person 'normal' variability by collecting frequent repeated measures of pain intensity may be valuable but this approach would not be feasible when identifying flares presenting at the point of care in routine clinical practice. Instead, this may have to rely on the judgement of the patient and/or clinician, the approach used, for example, in defining

exacerbations in COPD[1]. A similar consideration surrounds the speed of onset, which was not well defined by studies in our review. Drug withdrawal design studies specified washout periods between 2-15 days but this is unlikely to be synonymous with speed of onset. The remaining studies used terms such as 'sudden' and 'quick'. In COPD, for instance, a judgement around 'acute onset' or 'sudden onset' appears to be acceptable for clinical recommendations but we would add that the speed of onset of OA flares ought to be considered also in relation to underlying biologically plausible mechanisms. Indeed presumed aetiology has been argued as a useful feature in defining acute exacerbations in COPD[102]. Minimum duration ranged from 8 hours to 5 days in our review however this was not widely reported. COPD definitions refer to a 'sustained worsening' of symptoms[2] but does not appear to be a feature in other chronic diseases. A minimum duration in OA may help distinguish flares from day-to-day variability. Increase in medication was not found to be a key component in this review despite it being a feature in other chronic diseases; AS[5], SLE[4, 103], Inflammatory Bowel Disease[104], COPD[1]. Interference with function did not emerge strongly from our review as a cardinal feature of OA flare. In other chronic musculoskeletal conditions, such as back pain, interference with function was not shown to be significantly associated with having a flare up[105] and this domain does not feature in the definitions of exacerbations or flares in diseases such as COPD[1, 2], asthma[3], AS[5] or SLE[4].

Our review has several strengths but also some weaknesses that deserve attention. We adopted a broad search strategy, covering a wide range of databases, and featuring bibliography checks, contact with authors, inclusion of conference abstracts, no language restrictions, and a minimal threshold (any description or definition of flare) for inclusion. Five studies that were included in the Cross et al[21] review were not included in this study; four did not contain a clear definition of flare-up, including one which gave a definition of knee OA progression and the final paper by Sands et al[106] was not in our search but the original study was[60]. We did not, however, search the grey literature and we did not include some potential synonyms as search terms ('attack', 'episode', 'fluctuations') although these terms appeared often to relate to comorbidities and other phenomena (e.g. episodes of care) and would therefore have been a less efficient search strategy than relying on snowball references. Data extraction was performed by only a single reviewer. Nevertheless, we argue that our review provides a reasonably comprehensive summary of how 'flares' in OA have been described and defined in the medical literature. In comparison with Cross et al[21] our search strategy appeared comprehensive yet efficient – returning 69 included articles compared with 23. The majority of studies describe experimental 'flare design' trials in which flares are induced by drug withdrawal prior to enrolment and randomisation. While intentional or unintentional reduction in usual analgesia may indeed be one trigger for flare, experimentally induced flares should not be assumed to represent 'naturally occurring' flares. Flare

design trials, for example, are unlikely to capture change in management or healthcare usage that may be a common consequence of OA flares – something that is included in flare definitions in other conditions such as AS[5], SLE[4, 103], inflammatory bowel disease[104], and COPD[1].

A systematic review such as this cannot hope to resolve the need for a common conception and definition of flares in OA. Definitions for exacerbations of disease states are generally reached through a long process of consensus exercises involving key stakeholders, experts and patients in addition to appraisal of relevant literature from studies using multiple methods[6, 8, 107]. However, we believe that a consensus definition that is reliable, valid, and feasible and widely acceptable both clinically and for research purposes should now be sought. The cardinal features described in this review; onset/worsening of symptoms and signs, attainment of a minimum symptom threshold during flare, speed of onset/worsening, and duration of elevated symptoms/signs could help start this discussion.

CONCLUSION

A broad range of ad-hoc definitions currently exist in the medical literature. The majority are from drug-withdrawal or flare-induced trials rather than 'naturally' occurring flares. The cardinal feature is pain intensity with minimum symptom

threshold being another important feature. This review has identified the need to gain consensus on a common definition that can be used for research and clinical application.

Acknowledgements

We would like to thank you to Popay et al for allowing us to use their guidance on the conduct of narrative synthesis in systematic reviews. We would also like to thank Jo Jordan and Opeyemi Babatunde for their advice on conducting a systematic review.

Contributions

All authors were involved in conception and design of the study, analysis and interpretation of data, drafting the article, critical revision of the article for important intellectual content, final approval of the article. ELP and MJT extracted and synthesised data. ELP assembled the data.

GMP (g.m.peat@keele.ac.uk) takes responsibility for the integrity of the work as a whole from inception to finished article.

Data sharing statement

No unpublished data is available following this study

Role of the funding source

ELP is funded by a National Institute for Health Research (NIHR) In Practice Fellowship (IPF-2014-08-03). MJT received funding from a NIHR School for Primary Care Research Launching Fellowship and is currently funded by an Integrated Clinical Academic Programme Clinical Lectureship from the NIHR and Health Education England (HEE) (ICA-CL-2016-02-014).

The views expressed in this paper are those of the author(s) and not necessarily those of the NHS, the NIHR, HEE or the Department of Health.

Competing interest statement

GP received consultancy fees from InFirst plc and Good Relations plc.

Figure and Table Legends

Figure 1: PRISMA Flowchart

Table 1: Characteristics of all included studies

Table 2: Summary of number and type of single and multi-item measurement

Table 3: Definition, terminology and measurement instruments used in all included studies

Supplementary data: Database search strategy

REFERENCES

1 Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the diagnosis, management and prevention of COPD: GOLD 2016.

2 National Institute for Health and Care Excellence (NICE). Chronic obstructive pulmonary disease in over 16s: diagnosis and management (CG101). London: NICE 2010.

3 Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention: GINA 2015.

4 Ruperto N, Hanrahan L, Alarcón G, et al. International consensus for a definition of disease flare in lupus, *Lupus* 2011;20:453-62.

5 Stone MA, Pomeroy E, Keat A, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration, *Rheumatology* 2008;47:1213-8.

6 Bingham CO, Alten R, Bartlett SJ, et al. Identifying Preliminary Domains to Detect and Measure Rheumatoid Arthritis Flares: Report of the OMERACT 10 RA Flare Workshop, *The Journal of Rheumatology* 2011;38:1751-8.

7 Bykerk VP, Lie E, Bartlett SJ, et al. Establishing a Core Domain Set to Measure Rheumatoid Arthritis Flares: Report of the OMERACT 11 RA Flare Workshop, *The Journal of Rheumatology* 2014;41:799-809.

8 Bartlett SJ, Hewlett S, Bingham CO, et al. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus, *Annals of the Rheumatic Diseases* 2012;71:1855-60.

9 Taylor WJ, Shewchuk R, Saag KG, et al. Toward a valid definition of gout flare: Results of consensus exercises using delphi methodology and cognitive mapping, *Arthritis Care & Research* 2009;61:535-43.

10 Schmitt J, Spuls PI, Thomas KS, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials, *J Allergy Clin Immunol* 2014;134:800-7.

11 Holla JFM, van dL, Knol DL, et al. The association of body-mass index and depressed mood with knee pain and activity limitations in knee osteoarthritis: results from the Amsterdam osteoarthritis cohort, *BMC Musculoskeletal Disorders* 2013;14:296.

12 Collins JE, Katz JN, Dervan EE, et al. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative, *Osteoarthritis and Cartilage* 2014;22:622-30.

13 Leffondré K, Abrahamowicz M, Regeasse A, et al. Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators, *J Clin Epidemiol* 2004;57:1049-62.

14 Emrani PS, Katz JN, Kessler CL, et al. Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis, *Osteoarthritis and Cartilage* 2008;16:873-82.

15 Bartlett SJ, Ling SM, Mayo NE, et al. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis, *Arthritis Care & Research* 2011;63:1722-8.

16 Hawker GA, Stewart L, French MR, et al. Understanding the pain experience in hip and knee osteoarthritis – an OARSI/OMERACT initiative, *Osteoarthritis and Cartilage* 2008;16:415-22.

17 Arthritis Research UK. Osteoarthritis: Patient Information Booklet. 2012.

18 Bonnet CS, Walsh DA. Osteoarthritis, angiogenesis and inflammation, *Rheumatology* 2005;44:7-16.

19 Smith TO, Zou K, Abdullah N, et al. Does flare trial design affect the effect size of non-steroidal anti-inflammatory drugs in symptomatic osteoarthritis? A systematic review and meta-analysis, *Annals of the Rheumatic Diseases* 2016;75:1971-8.

20 Marty M, Hilliquin P, Rozenberg S, et al. Validation of the KOFUS (Knee Osteoarthritis Flare-Ups Score), *Joint Bone Spine* 2009;76:268-72.

21 Cross M, Dubouis L, Mangin M, et al. Defining Flare in Osteoarthritis of the Hip and Knee: A Systematic Literature Review- OMERACT Virtual Special Interest Group, *J Rheumatol* 2017;44(12):1920-7.

22 Rutjes AS, Jüni P, Da Costa BR, et al. Viscosupplementation for osteoarthritis of the knee: A systematic review and meta-analysis, *Ann Intern Med* 2012;157:180-91.

23 Higgins J, Green S, eds. Cochrane handbook for systematic reviews of interventions Version 5.1.0. Available from www.handbook.cochrane.org.: The Cochrane Colloboration 2011.

24 Popay J, Roberts H, S, A., et al. Guidance on the conduct of narrative synthesis in systematic reviews: A product of the ESRC methods programme Lancaster: ESRC Method Programme, 2006.

25 Thomas J, Harden A, Newman M. Synthesis: Combining results systematically and appropriately. In: Gough A, Oliver S, Thomas J, eds. An introduction to systematic reviews. London: Sage publications limited 2013:191-2.

26 Altman R, Hochberg M, Gibofsky A, et al. Efficacy and safety of low-dose SoluMatrix meloxicam in the treatment of osteoarthritis pain: A 12-week, phase 3 study, *Curr Med Res Opin* 2015;31:2331-43.

27 Atukorala I, Pathmeswaran A, Makovey J, et al. Is there a relationship between the intermittent and constant osteoarthritis pain score (ICOAP) and pain flares in knee osteoarthritis? (abstract) [abstract]. *Osteoarthritis and Cartilage* 2016;24:S429-30.

28 Baer PA, Thomas LM, Shainhouse Z. Treatment of osteoarthritis of the knee with a topical diclofenac solution: a randomised controlled, 6-week trial ISRCTN53366886], *BMC Musculoskeletal Disorders* 2005;6:44.

29 Baraf HSB, Gloth FM, Barthel HR, et al. Safety and Efficacy of Topical Diclofenac Sodium Gel for Knee Osteoarthritis in Elderly and Younger Patients, *Drugs Aging* 2011;28:27-40.

30 Battisti WP, Katz NP, Weaver AL, et al. Pain management in osteoarthritis: A focus on onset of efficacy—a comparison of rofecoxib, celecoxib, acetaminophen, and nabumetone across four clinical trials, *The Journal of Pain* 2004;5:511-20.

31 Bingham CO, Sebba AI, Rubin BR, et al. Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies, *Rheumatology* 2007;46:496-507.

32 Birbara C, Ruoff G, Sheldon E, et al. Efficacy and safety of rofecoxib 12.5 mg and celecoxib 200 mg in two similarly designed osteoarthritis studies, *Curr Med Res Opin* 2006;22:199-210.

33 Bocanegra T, Weaver A, Tindall E, et al. Diclofenac/misoprostol compared with diclofenac in the treatment of osteoarthritis of the knee or hip: a randomized, placebo controlled trial. Arthrotec Osteoarthritis Study Group. *Journal of Rheumatology* 1998;25:1602-11.

34 Boswell DJ, Ostergaard K, Philipson RS, et al. Evaluation of GW406381 for Treatment of Osteoarthritis of the Knee: Two Randomized, Controlled Studies, *The Medscape Journal of Medicine* 2008;10:259.

35 Brandt KD, Mazzuca SA, Buckwalter KA. Acetaminophen, like conventional NSAIDs, may reduce synovitis in osteoarthritic knees, *Rheumatology* 2006;45:1389-94.

36 Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: A randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium, *Arch Intern Med* 2003;163:169-78.

37 Ehrich E, Schnitzer T, McIlwain H, et al. Effect of specific COX-2 inhibition in osteoarthritis of the knee: a 6 week double blind, placebo controlled pilot study of rofecoxib. Rofecoxib Osteoarthritis Pilot Study Group. *Journal of Rheumatology* 1999;26:2438-47.

38 Essex M, O'Connell M, Brown PB. Response to Nonsteroidal Anti-Inflammatory Drugs in African Americans with Osteoarthritis of the Knee, *Journal of International Medical Research* 2012;40:2251-66.

39 Gibofsky A, Hochberg MC, Jaros MJ, et al. Efficacy and safety of low-dose submicron diclofenac for the treatment of osteoarthritis pain: A 12 week, phase 3 study, *Curr Med Res Opin* 2014;30:1883-93.

40 Gineyts E, Mo JA, Ko A, et al. Effects of ibuprofen on molecular markers of cartilage and synovium turnover in patients with knee osteoarthritis, *Annals of the Rheumatic Diseases* 2004;63:857-61.

41 Goldberg M, McIlwain H, Poiley J, et al. Controlled-release naproxen in the treatment of osteoarthritis, *Current Therapeutic Research-Clinical and Experimental* 1988;44:51-60.

42 Gottesdiener K, Schnitzer T, Fisher C, et al. Results of a randomized, dose ranging trial of etoricoxib in patients with osteoarthritis, *Rheumatology* 2002;41:1052-61.

43 Hochberg MC, Fort JG, Svensson O, et al. Fixed-dose combination of enteric-coated naproxen and immediate-release esomeprazole has comparable efficacy to celecoxib for knee osteoarthritis: two randomized trials, *Curr Med Res Opin* 2011;27:1243-53.

44 Katz N, Sun S, Johnson F, et al. ALO-01 (Morphine Sulfate and Naltrexone Hydrochloride) Extended-Release Capsules in the Treatment of Chronic Pain of Osteoarthritis of the Hip or Knee: Pharmacokinetics, Efficacy, and Safety, *The Journal of Pain* 2010;11:303-11.

45 Kivitz AJ, Makarowski WS, Fiechtner JJ, et al. A Flexible Daily Dosage Regimen of Oxaprozin Potassium in Patients with Acute Knee Pain Associated with Osteoarthritis, *Clinical Drug Investigation* 2001;21:745-53.

46 Erfani T, Zhang Y, Makovey J, et al. Intermittent analgesic use and risk of pain exacerbation in knee osteoarthritis: A web based case-crossover study (abstract) [abstract]. *Arthritis and Rheumatology* 2014;66.

47 Leung AT, Malmstrom K, Gallacher AE, et al. Efficacy and Tolerability Profile of Etoricoxib in Patients with Osteoarthritis: A Randomized, Double-blind, Placebo and Active-comparator Controlled 12-Week Efficacy Trial, *Curr Med Res Opin* 2002;18:49-58.

48 Luyten FP, Geusens P, Malaise M, et al. A prospective randomised multicentre study comparing continuous and intermittent treatment with celecoxib in patients with osteoarthritis of the knee or hip, *Annals of the Rheumatic Diseases* 2007;66:99-106.

49 Manicourt D, Bevilacqua M, Righini V, et al. Comparative Effect of Nimesulide and Ibuprofen on the Urinary Levels of Collagen Type II C-Telopeptide Degradation Products and on the Serum Levels of Hyaluronan and Matrix Metalloproteinases-3 and -13 in Patients with Flare-Up of Osteoarthritis, *Drugs in R & D* 2005;6:261-71.

50 Mazzuca S, Brandt K, Lane K, et al. Knee pain reduces joint space width in conventional standing anteroposterior radiographs of osteoarthritic knees, *Arthritis Rheum* 2002;46:1223-7.

51 McIlwain H, Silverfield JC, Cheatum DE, et al. Intra-articular orgotein in osteoarthritis of the knee: A placebo-controlled efficacy, safety, and dosage comparison, *Am J Med* 1989;87:295-300.

52 Mendelsohn S. Clinical efficacy and tolerability of naproxen in osteoarthritis patients using twice-daily and once-daily regimens, *Clinical therapeutics* 1991;13:8-15.

53 Moskowitz RW, Sunshine A, Hooper M, et al. An analgesic model for assessment of acute pain response in osteoarthritis of the knee, *Osteoarthritis and Cartilage* 2006;14:1111-8.

54 Pareek A, Chandurkar N, Sharma VD, et al. A randomized, multicentric, comparative evaluation of aceclofenac-paracetamol combination with aceclofenac alone in Indian patients with osteoarthritis flare-up, *Expert Opin Pharmacother* 2009;10:727-35.

55 Pareek A, Chandurkar N, Ambade R, et al. Efficacy and Safety of Etodolac-Paracetamol Fixed Dose Combination in Patients With Knee Osteoarthritis Flare-up: A Randomized, Double-blind Comparative Evaluation, *Clin J Pain* 2010;26:561-6.

56 Ricci JA, Stewart WF, Chee E, et al. Pain Exacerbation as a Major Source of Lost Productive Time in US Workers With Arthritis, *Arthritis & Rheumatism: Arthritis Care & Research* 2005;53:673-81.

57 Schnitzer TJ, Fricke JR, Gitton X, et al. Lumiracoxib in the treatment of osteoarthritis, rheumatoid arthritis and acute postoperative dental pain: results of three dose-response studies, *Curr Med Res Opin* 2005;21:151-61.

58 Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, et al. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis, *Arthritis & Rheumatism* 2001;44:1599-607.

59 Silverfield JC, Kamin M, Wu S, et al. Tramadol/acetaminophen combination tablets for the treatment of osteoarthritis flare pain: a multicenter, outpatient, randomized, double-blind, placebo-controlled, parallel-group, add-on study, *Clin Ther* 2002;24:282-97 doi:http://dx.doi.org/10.1016/S0149-2918(02)85024-X [published Online First: February 2002].

60 Strand V, Simon LS, Dougados M, et al. Treatment of osteoarthritis with continuous versus intermittent celecoxib, *J Rheumatol* 2011;38:2625-34.

61 Wiesenhutter CW, Boice JA, Ko A, et al. Evaluation of the Comparative Efficacy of Etoricoxib and Ibuprofen for Treatment of Patients With Osteoarthritis: A Randomized, Double-Blind, Placebo-Controlled Trial, *Mayo Clin Proc* 2005;80:470-9.

62 Williams GW, Hubbard RC, Yu SS, et al. Comparison of once-daily and twice-daily administration of celecoxib for the treatment of osteoarthritis of the knee, *Clin Ther* 2001;23:213-27 doi:http://dx.doi.org/10.1016/S0149-2918(01)80004-7 [published Online First: February 2001].

63 Wittenberg RH, Schell E, Krehan G, et al. First-dose analgesic effect of the cyclo-oxygenase-2 selective inhibitor lumiracoxib in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled comparison with celecoxib NCT00267215], *Arthritis Research & Therapy* 2006;8:R35.

64 Yeasted R, McPherson J, Schnitzer T. Characterization of osteoarthritis pain variability (abstract) [abstract]. *Osteoarthritis and Cartilage* 2014;22:S390-1.

65 Young C, Parenti D, Hochberg M. Lower-dose diclofenac capsules developed using solumatrix fine particle technology result in clinically meaningful improvements in pain in a phase 3 study of patients with osteoarthritis (abstract) [abstract]. *Osteoarthritis and Cartilage* 2014;22.

66 Zhao SZ, McMillen JI, Markenson JA, et al. Evaluation of the Functional Status Aspects of Health-Related Quality of Life of Patients with Osteoarthritis Treated with Celecoxib, *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 1999;19:1269-78.

67 Zobel I, Erfani T, Bennell K, et al. Relationship of buckling and knee injury to pain exacerbation in knee osteoarthritis: A web-based case-crossover stud, *Interact J Med Res* 2014;66:S560-1.

68 Conrozier T, Mathieu P, Vignon E, et al. Differences in the osteoarthritic synovial fluid composition and rheology between patients with or without flare: a pilot study. *Clinical and experimental rheumatology* 2012;30:729-34.

69 D'Agostino MA, Conaghan P, Le Bars M, et al. EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1: Prevalence of inflammation in osteoarthritis, *Annals of the Rheumatic Diseases* 2005;64:1703-9.

70 Jawad ASM. Analgesics and osteoarthritis: are treatment guidelines reflected in clinical practice? *Am J Ther* 2005;12:98-104.

71 Murphy SL, Lyden AK, Kratz AL, et al. Characterizing pain flares from the perspective of individuals with symptomatic knee osteoarthritis, *Arthritis Care and Research* 2015;67:1103-11.

72 Wise BL, Niu J, Zhang Y, et al. Psychological factors and their relation to osteoarthritis pain, *Osteoarthritis and Cartilage* 2010;18:883-7.

73 Zhang Y, Zhang B, Wise B, et al. Statistical approaches to evaluating the effect of risk factors on the pain of knee osteoarthritis in longitudinal studies, *Curr Opin Rheumatol* 2009;21:513-9.

74 Zhang Y, Wheaton D, N, J., et al. Recent heavy physical activities trigger knee pain exacerbation in persons with symptomatic knee osteoarthritis (abstract) [abstract]. *Arthritis & Rheumatism* 2011;63(10).

75 Day R, Morrison B, Luza A, et al. A randomized trial of the efficacy and tolerability of the cox-2 inhibitor rofecoxib vs ibuprofen in patients with osteoarthritis, *Arch Intern Med* 2000;160:1781-7.

76 Kivitz AJ, Greenwald MW, Cohen SB, et al. Efficacy and Safety of Rofecoxib 12.5 mg Versus Nabumetone 1,000 mg in Patients with Osteoarthritis of the Knee: A Randomized Controlled Trial, *J Am Geriatr Soc* 2004;52:666-74.

77 Bingham CO, Smugar SS, Wang H, et al. Predictors of Response to Cyclo-Oxygenase-2 Inhibitors in Osteoarthritis: Pooled Results from Two Identical Trials Comparing Etoricoxib, Celecoxib, and Placebo, *Pain Medicine* 2011;12:352-61.

78 Essex MN, Behar R, O'Connell MA, et al. Efficacy and tolerability of celecoxib and naproxen vs placebo in hispanic patients with knee osteoarthritis, *Osteoarthritis and Cartilage* 2013;21.

79 Yocum D, Fleischmann R, Dalgin P, et al. Safety and efficacy of meloxicam in the treatment of osteoarthritis: A 12-week, double-blind, multiple-dose, placebo-controlled trial, *Arch Intern Med* 2000;160:2947-54.

80 Atukorala I, Pathmeswaran A, Chang T, et al. Do Traditional Risk Factors for Knee Osteoarthritis Predict Pain Flares in Knee Osteoarthritis? *Ann Rheum Dis* 2016;75:835.

81 Bartholdy C, Klokker L, Bandak E, et al. A Standardized "Rescue" Exercise Program for Symptomatic Flare-up of Knee Osteoarthritis: Description and Safety Considerations, *J Orthop Sports Phys Ther* 2016;46:942-6.

82 Bassiouni H. Detection of changes in the serum and synovial fluid levels of resistin during flare ups and remissions in primary knee osteoarthritis, *Arthritis and Rheumatology* 2015;67.

83 Erfani T, Makovey J, Bennell K, et al. Psychosocial Factors and Pain Exacerbation in Knee Osteoarthritis: a Web Based Case-Crossover Study, *Intern Med J* 2014;44:16-.

84 Ferreira ML, Zhang Y, Metcalf B, et al. The influence of weather on the risk of pain exacerbation in patients with knee osteoarthritis - a case-crossover study, *Osteoarthritis and cartilage* 2016;24:2042-7.

85 Hunter DJ, Bennell K, Makovey J, et al. Psychosocial Factors and Pain Exacerbation in Knee Osteoarthiritis: a Web Based Case-Crossover Study, *Osteoarthritis and Cartilage* 2014;22:S21-2.

86 Makovey J, Metcalf B, Zhang Y, et al. Web-Based Study of Risk Factors for Pain Exacerbation in Osteoarthritis of the Knee (SPARK-Web): Design and Rationale, *JMIR research protocols* 2015;4.

87 Parry E, Ogollah R, Peat G. Significant pain variability in persons with, or at high risk of, knee osteoarthritis: preliminary investigation based on secondary analysis of cohort data, *BMC musculoskeletal disorders* 2017;18:80.

88 Cibere J, Kopec JA, Thorne A, et al. Randomized, double-blind, placebo-controlled glucosamine discontinuation trial in knee osteoarthritis, *Arthritis Care and Research* 2004;51:738-45.

89 Cibere J, Kopec JA, Esdaile JM, et al. Glucosamine sulfate and cartilage type II collagen degradation in patients with knee osteoarthritis: randomized discontinuation trial results employing biomarkers. In: Anonymous . Journal of rheumatology 2005:896-902.

90 Roth SH, Shainhouse JZ. 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, *Arch Intern Med* 2004;164:2017-23.

91 Simon LS, Grierson LM, Naseer Z, et al. 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* 2009;143:238-45.

92 Weaver A, Rubin B, Caldwell J, et al. Comparison of the efficacy and safety of oxaprozin and nabumetone in the treatment of patients with osteoarthritis of the knee, *Clin Ther* 1995;17:735-45.

93 Rother M, Lavins BJ, Kneer W, et al. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral celecoxib and placebo in osteoarthritis of the knee: multicentre randomised controlled trial, *Annals of the Rheumatic Diseases* 2007;66:1178-83.

94 Zobel I, Erfani T, Bennell KL, et al. Relationship of Buckling and Knee Injury to Pain Exacerbation in Knee Osteoarthritis: A Web-Based Case-Crossover Study, *Interactive journal of medical research* 2016;5:e17.

95 Roth ML, Tripp DA, Harrison MH, et al. Demographic and psychosocial predictors of acute perioperative pain for total knee arthroplasty, *Pain Research & Management* 2007;12:185-94.

96 Woolacott NF, Corbett MS, Rice SJC. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials, *Rheumatology* 2012;51:1440-6.

97 National Institute for Health and Care Excellence (NICE). Osteoarthritis: care and management (CG177). London: NICE 2014.

98 Buttgereit F, Burmester G, Bijlsma JWJ. Non-surgical management of knee osteoarthritis: where are we now and where do we need to go? *RMD Open* 2015;1.

99 Porcheret M, Healey E, Dziedzic K, et al. Ostoearthritis: a modern approach to diagnosis and management, *Arthritis Research UK* 2011;Series 6.

100 Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee, *J Rheumatol* 1990;17:364-72.

101 Allen KD, Coffman CJ, Golightly YM, et al. Daily pain variations among patients with hand, hip, and knee osteoarthritis, *Osteoarthritis and Cartilage* 2009;17:1275-82.

102 Makris D, Bouros D. COPD exacerbation: Lost in translation, *BMC Pulmonary Medicine* 2009;9:6.

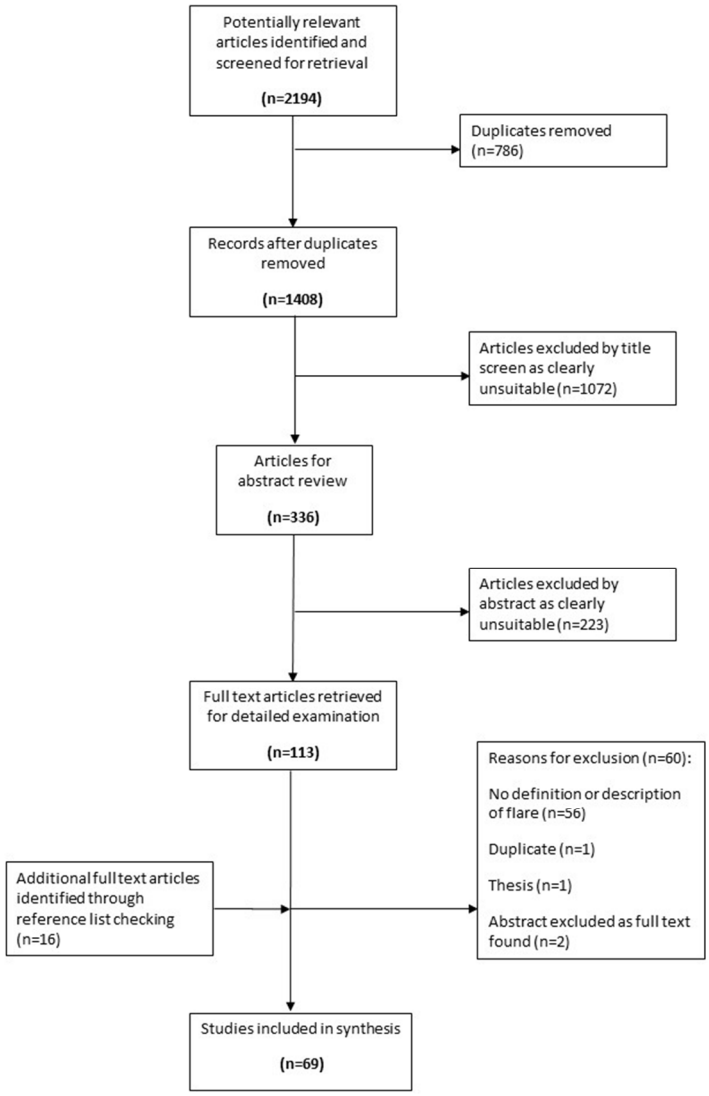
103 Fitzgerald JD, Grossman JM. Validity and reliability of retrospective assessment of disease activity and flare in observational cohorts of lupus patients, *Lupus* 1999;8:638-44.

104 Lewis JD, Aberra FN, Lichtenstein GR, et al. Seasonal variation in flares of inflammatory bowel disease, *Gastroenterology*;126:665-73.

105 Suri P, Saunders KW, Von Korff M. Prevalence and Characteristics of Flare-ups of Chronic Nonspecific Back Pain in Primary Care: A Telephone Survey, *Clin J Pain* 2012;28:573-80.

106 Sands GH, Brown PB, Essex MN. The Efficacy of Continuous Versus Intermittent Celecoxib Treatment in Osteoarthritis Patients with Body Mass Index â‰¥30 and *The Open Rheumatology Journal* 2013;7:32-7.

107 Berthelot J, De Bandt M, Morel J, et al. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: The 'FLARE' instrument, *Annals of the Rheumatic Diseases* 2012;71:1110-6.



190x275mm (96 x 96 DPI)

Online supplement: Example search strategy

Table 1: Key terms and MeSH headings used for EMBASE database search. The concepts were combined as follows: "KNEE JOINT" AND "ACUTE EVENTS"

Concepts	Search terms
KNEE JOINT	"knee adj3 (pain OR painful)" or "Knee osteoarthritis" or "knee adj3 (arthrosis)" or "knee adj3 (joint OR joints OR degenerative)" or "knee adj3 (osteoarthritis)"
ACUTE EVENTS	"exacerbation" or "flare" or "daily adj3 (pain)" or "pain AND (diary OR diaries)" or "pain adj3 (variab$)"

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6-7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Online supplement
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	n/a
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	9

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n/a
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10 (flowchart)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10, Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A but rationale on 18
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	n/a
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	n/a
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	19-35
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	36-37
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	39
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	40

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

PRISMA 2009 Checklist

BMJ Open

Defining Acute Flares in Knee Osteoarthritis: A Systematic Review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019804.R2
Article Type:	Research
Date Submitted by the Author:	10-Apr-2018
Complete List of Authors:	Parry, Emma; Keele University, Research Institute for Primary Care and Health Sciences Thomas, M; Keele University, Research Institute for Primary Care and Health Sciences Peat, George; Keele University, Research Institute for Primary Care & Health Sciences
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	General practice / Family practice
Keywords:	Osteoarthritis, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Flare

DEFINING ACUTE FLARES IN KNEE OSTEOARTHRITIS: A SYSTEMATIC REVIEW

Emma L. Parry[1] , e.parry@keele.ac.uk

Martin J. Thomas[1], m.thomas@keele.ac.uk

George Peat[1*], g.m.peat@keele.ac.uk

[1] *Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, UK*

*Correspondence: Professor George Peat, Professor of Clinial Epidemiology, Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG. Tel: +44 (0) 1782 732929. Fax:+44 (0) 1782 734719. Email: g.m.peat @keele.ac.uk

Keywords: knee osteoarthritis, flare, systematic review

ABSTRACT

Objective: To identify and critically synthesise definitions of acute flares in knee osteoarthritis (OA) reported in the medical literature.

Design: Systematic review and narrative synthesis. We searched MEDLINE, EMBASE, Web of science and 6 other electronic databases (inception to July 2017) for original articles and conference abstracts reporting a definition of acute flare (or synonym) in humans with knee OA. There were no restrictions by language or study design (apart from iatrogenic induced flare-ups e.g. injection-induced). Data extraction comprised: definition, pain scale used, flare duration or withdrawal period, associated symptoms, definition rationale, terminology (e.g. exacerbation or flare), baseline OA severity, age, gender, sample size and study design.

Results: Sixty-nine articles were included (46 flare-design trials, 17 observational studies, 6 other designs; sample sizes: 15-6085). Domains used to define flares included: worsening of signs and symptoms (61 studies, 27 different measurement tools), specifically increased pain intensity; minimum pain threshold at baseline (44 studies); minimum duration (7 studies, range 8-48 hours); speed of onset (2 studies, defined as 'sudden' or 'quick'); requirement for increased medication (2 studies). No definitions included activity interference.

Conclusions: The concept of OA flare appears in the medical literature but most often in the context of flare design trials (pain increases observed after stopping usual treatment). Key domains, used to define acute events in other chronic

conditions, appear relevant to OA flare and could provide the basis for consensus on a single, agreed definition of 'naturally occurring' OA flares for research and clinical application.

PROSPERO registration: CRD42014010169

Strengths and limitations of this study

Strengths

- Identified key domains that are used to define acute events by undertaking a comprehensive synthesis of definitions used in the medical literature.
- Broad search strategy covering a wide range of databases including bibliography checks and conference abstracts.
- Prospectively registered with Prospero

Limitations

- Did not include potential synonyms as search terms ('attack', 'episode', 'fluctuations')
- Data extraction was performed by only a single reviewer.

INTRODUCTION

Recurrent acute events or episodes feature in the natural history of many chronic health conditions. The extent to which they characterise the condition varies, as do the presumed pathophysiological mechanisms, and scientific and lay terms used to describe them (e.g. an acute exacerbation of chronic obstructive pulmonary disease (COPD) or asthma, an attack of gout or a rheumatoid arthritis flare). With recognition of their importance has come concerted effort to define these phenomena. Definitions for exacerbations or flares currently exist for COPD[1, 2] , asthma[3], systemic lupus erythematosus (SLE)[4], and ankylosing spondylitis (AS)[5] and there are working groups currently trying to define these for rheumatoid arthritis[6-8], gout[9], and atopic dermatitis/eczema[10]. Despite the different language used, these definitions share some common, core domains: the onset or worsening of symptoms and signs above normal day-to-day variability; speed of onset; duration of sustained worsening; and change in medication/healthcare usage.

Osteoarthritis (OA) appears to comprise multiple disease trajectories[11-15] and symptom variability over time and the presence of intermittent pain is well-recognised[16]. Although OA does not typically have the same very obvious acute events as conditions like gout, flares in OA joints are encountered in practice, these

phenomena appear in patient literature[17], have been discussed in expert reviews[18], and are mentioned in 'flare design' trials in OA[19]**.** These studies invoke acute episodes of pain or flare-ups by asking patients to withdraw their usual medication.

In 2009 Marty et al proposed scoring criteria for knee OA flares based on nocturnal awakening, knee effusion, morning stiffness and limping[20] but it is unclear whether this has contributed to a common understanding, shared terminology and criteria. A common definition of OA flare could be important for a number of reasons; (i) to facilitate communication between researchers, (ii) to allow more direct comparisons between studies on frequencies, determinants and course of events, (iii) to facilitate new insights into novel pathophysiological mechanisms and treatments through valid and homogenous case definitions, and (iv) to help clinicians with prompt diagnosis and management.

The aim of this systematic review was to explore the extent to which a concept of OA flare is reported in the medical literature and the prospects for a common, shared definition of these for research and clinical application.

94

95

96

97

98

METHODS

This systematic review was registered with PROSPERO registration number CRD42014010169. The review protocol has not been published.

Literature sources and study selection

We searched electronic databases from inception to July 2017; ASSIA, EMBASE, Web of Science, Health Management Information Consortium (HMIC), SPORTDiscus, Medline, CINAHL, PsycINFO, AMED, Ageline, Cochrane Database of Systematic Reviews and Cochrane Controlled Clinical Trials (CENTRAL). The search was developed using previously piloted terms for knee OA and a literature search for common terms used to describe acute events. Searches used combined and/or truncated key terms including: ("KNEE OSTEOARTHRITIS" OR (knee N3 pain) OR (knee N3 arthrosis) OR (knee N3 joint) OR (knee N3 osteoarthritis)) AND (exacerbation OR flare OR (pain AND (diary OR diaries)) OR (pain N3 variab*) OR (pain N3 pattern$) OR (daily N3 pain)). A database search strategy is included in the online supplement . Reference lists of all included full text articles retrieved for detailed examination were manually searched.

Studies were included in the final full text peer review if they contained a description or definition of an acute exacerbation or flare-up of knee OA in human adults (18 years or over) in the general population, primary care or hospital settings. Studies were included even if their description was not based on clear measurement criteria (e.g. stating a 'significant increase in pain' but not the amount of change on a pain score this would equate to). Studies that included a mixed OA population (e.g. knee or hip OA) and did not separately report knee-specific findings were included. There were no restrictions on study dates or design. All non-English language articles were translated to identify a flare definition. Theses, dissertations, book chapters and guidelines, and animal studies were excluded. Conference abstracts were included if they contained a definition for an OA flare-up. Studies were excluded if the flare was induced by an iatrogenic source, for example, injection-induced flares[21]. As these may have been caused by a different pathophysiological process. Abstracts were included in this study as the main outcome of interest was the definition of flare used and it was decided that including abstracts would ensure a more comprehensive review. For each abstract a search was conducted to identify a corresponding full text paper. Where one was found only the full paper was included in the review.

The search and article retrieval was conducted by the first reviewer (ELP). Articles were downloaded into RefWorks© bibliography and database manager (RefWorks Copyright 2009). Duplicates were removed and all titles were screened by ELP

against inclusion criteria, with the first 20 titles checked by two reviewers (ELP and MJT) for consistency. For qualitative studies, all identified potentially eligible full text articles were obtained.

All abstracts and then full text articles were screened by two reviewers (ELP and MJT). with disagreements resolved by consensus adjudicated by a third reviewer (GP). Where articles could not be retrieved or if the flare definition used was not included in the text, contact with authors was made.

The final included articles were checked to ensure results were not duplicated, for example, where different authors were reporting on the same dataset, to reduce bias[22] . For articles containing pooled studies, the original studies were sought and included in the main analysis, where available.. No full text articles were required to be translated.

Data extraction

The following data pertaining to flares were extracted from full text articles by the first reviewer: definition used for change in pain, pain scale used, duration of flare (for flare design trials we extracted the duration of the withdrawal period for

comparison), associated symptoms, rationale behind definition used, terminology used (e.g. exacerbation or flare), baseline OA severity, age range, gender, geographical location, number of participants and study design. Missing data was described in the data extraction tables.

Quality assessment of included studies

Our aim was to identify and contrast definitions of flare-ups used in the literature. We were not concerned with the methodological rigour of the studies deriving, evaluating or applying those definitions. However, for studies presenting definitions we sought supporting statements that gave the rationale for the definition.

Data analysis

A narrative synthesis was undertaken, guided by Popay et al's[23] four stage process to develop a conceptual framework[22]. This approach was chosen as it allowed the words and text in the definitions to be synthesised to summarise findings[23]. The initial data extracted was grouped into drug withdrawal studies ('flare design') and other studies. Frequencies of components included in definitions was tabulated, these included; terminology used, onset/worsening of symptoms; signs/symptoms

above day-to-day variability/minimum threshold; speed of onset of symptoms; duration of worsening and change in medication/healthcare usage.

This initial tabulation helped identify similarities and differences and allowed themes to emerge. This was done with an inductive type approach, where possible i.e. without an *a priori* assumption, but also deductively acknowledging that the reviewers were clinicians i.e. they had some background knowledge of the topic of interest. This allowed further examination of the differences of definitions used in drug withdrawal and non-drug withdrawal study designs, and examination of key components of definitions used.

Patient and public involvement

There was no patient or public involvement in this study.

RESULTS

Study selection

The literature search yielded 2194 articles of which 786 were duplicates (Figure 1). After title screening 336 abstracts were reviewed, 223 were not relevant for the study purpose. 113 articles were examined in full which resulted in a further 60 being excluded. The main reason for exclusion was no definition of flare-up reported in text (n=56). At this stage a further 16 articles were identified from the reference lists of the retrieved full text articles resulting in 69 included studies for synthesis.

Study characteristics

Characteristics of the included studies are described in Table 1[20, 24-91]**.** The number of participants in each study ranged from 15-6085[20, 48]. Knee OA was defined by clinical and/or radiological criteria.

213 **Table 1: Characteristics of all included studies**

DRUG WITHDRAWAL DESIGN STUDIES					
First author, year of publication	**Setting, geographic location**	**Participants**	**Joint**	**Severity**	**Study design**
Altman, 2015[24]	Multi-centre, recruitment not specified, USA	403 males & females, ≥40y	Knee and hip	KL grade 2-3	RCT, flare design
Baer, 2005[26]	17 medical centres recruiting from community and physician private practice; Canada	216 males & females, 40-85y	Knee	Radiographic evidence of OA (severity not defined)	RCT, flare design
Baraf, 2011[27]	Primary care, internal medicine, orthopaedic, rheumatology; USA	602 males & females, ≥25y	Knee	Radiographically mild to moderate (KL grade 1-3)	RCT, flare design
Battisti, 2004[28]	Clinical centres, out patients; USA	3980 males & females, ≥40y (age unavailable for Geba 2003 and Weaver 2003)	Knee	ACR functional class rating of I,II or III	RCT, pooled 4 trials, flare design
Bingham, 2007[29] Bingham 2011[75]	2x74 outpatient clinics; USA	1207 males & females, ≥40y	Knee and hip	ARA Functional capacity classification I-III	RCT, flare design
Birbara, 2006[30]	Investigative	808 males &	Knee	ARA functional class, I, II, or III	RCT, flare design

	sites; USA	females, ≥40y			
Bocanegra, 1998[31]	Clinic; USA	572 males & females, 28-88y (mean 61-62)	Knee and hip	ARA Functional capacity classification I-III	RCT, flare design
Boswell, 2008[32]	50 centres (Europe & Australia) + 187 centres (Europe & USA)	1908 males & females, ≥40y	Knee	KL scale 2 or 3 and ARA class rating of I,II or III	Pooled RCTs (2; one flare design, one non-flare), flare design
Brandt, 2006[33] (pilot studies)	Community; USA	30 males & females, mean age 62y	Knee	KL ≥2	Cohort design, flare design
Case, 2003[34]	Hospital-rheumatology centre; Chicago, USA	82 males & females, 40-75y	Knee	KL ≥1, and clinical criteria (pre-enrolment ambulatory pain; moderate pain by a 5-point Likert scale or increased pain.	RCT, flare design
Day, 2000[73]	49 investigative sites in 26 countries	809 males & females, mean age range 62-65y	Knee and hip	ARA functional class I-III, symptomatic for at least 6 months	RCT, flare design
Ehrich, 1999[35]	Clinical centres; USA	219 males & females, >40y	Knee	ARA functional class, I, II, or III	RCT, flare design
Essex, 2012[36]	Clinical centre; African-American, USA	322 males & females, ≥45y	Knee	ARA Functional capacity classification I-III	RCT, flare design
Essex 2013[76]	Hispanic population, 31 US centers	≥45y	Knee	ACR criteria, Functional capacity classification I-III	RCT, flare design
Gibofsky, 2014[37]	Not specified, USA	305 males & females, 41-90 y	Knee and hip	KL 2-3	RCT, flare design
Gineyts, 2004[38]	Subset of larger study; France	201 males & females, mean age 61-62y	Knee and hip	ARA I-III	RCT, flare design

Goldberg, 1988[39]	Investigative sites; USA	214 males & females, 40-85y (mean 64)	Knee and hip	Radiographic evidence of knee OA-not further defined	RCT, flare design
Gottesdiener, 2002[40]	Investigative sites; USA	617 males & females, ≥40y	Knee	ARA functional class I,II,III	RCT, flare design
Hochberg, 2011[41]	Centres; USA	1234 males & females, ≥50y	Knee	ACR functional class I-III	Pooled RCTs (2), flare design
Katz, 2010[42]	Clinical sites; USA	113 males & females, 28-83y (median 57))	Knee and hip	OA of hip and knee as diagnosed using ACR criteria-no definition of severity	RCT, flare design
Kivitz, 2001[43]	Investigative sites; USA	491 males & females, 28-91y (mean 58-61)	Knee	Confirmation of OA on weight bearing radiograph- no definition of severity	RCT, flare design
Kivitz, 2004[74]	Outpatient sites; USA	1042 males & females, ≥40y	Knee	ACR rating of I, II, III.	RCT, flare design
Leung, 2002[45]	Clinic; USA	677 males & females, ≥40y	Knee and hip	ARA functional class, I, II, or III	RCT, flare design
Luyten, 2007[45]	Centres; Belgium	181 males & females, ≥40y	Knee and hip	ACR Functional capacity classification I-III	RCT, flare design
Manicourt, 2005[47]	Outpatient clinic; Belgium	90 males & females, 50-81y (mean 63-67)	Knee and hip	Clinical and radiographic evidence of OA-severity not defined.	RCT, flare design
Mazzuca, 2002[48]	Not specified, USA	15 males & females, ≥45y	Knee	KL 2-3	Observational, flare design
McIlwain, 1989[49]	Investigative sites; USA	139 males & females, mean 65y	Knee	Radiological evidence of moderate or severe osteoarthritis- not further defined	RCT, flare design
Mendelsohn, 1991[50]	Investigative sites; USA	139 males & females, 21-88y (mean age 63.3y)	Knee	Radiological evidence of moderate or severe osteoarthritis- not further defined	RCT, flare design

Moskowitz, 2006[51]	Investigative sites; USA	530 males & females, ≥45y	Knee	ACR Functional capacity classification I-III	RCT, flare design
Pareek, 2009[52]	Multi-centre study, India	199 males & females, 40-70y	Knee	Lequesne criteria-score of 5 and above	RCT, flare design
Pareek, 2010[53]	Hospital; India	220 males & females, 40-70y	Knee	Clinical and radiological evidence of OA- severity not defined.	RCT, flare design
Roth, 2004[88]	Physicians private practice or community; USA	326 males & females, 40-85y	Knee	Radiological evidence of OA- severity not defined.	RCT, flare design
Rother, 2007[91]	Outpatient units; Germany	397 males & females, ≥40y	Knee	KL 2-3	RCT, flare design
Schnitzer, 2005[55]	Investigative sites; International (7 countries)	583 males & females, 18-75y	Knee and hip	Diagnosis based on ACR criteria- severity not defined.	RCT, flare design
Scott-Lennox, 2001[56]	Investigative sites; USA	182 males & females, mean 61y	Knee	Not defined	RCT, flare design
Silverfield, 2002[57]	Centres; USA	308 males & females, 35-75y	Knee and hip	Clinical evidence of OA- severity not defined	RCT, flare design
Simon, 2009[89]	Outpatient centres; Canada, USA	775 males & females, 40-85y	Knee	Clinical and radiological evidence of OA- severity not defined	RCT, flare design

Strand, 2011[58]	Investigative sites; Multinational-not specified including USA	875 males & females, 18-80y	Knee and hip	OA according to ACR criteria and requiring NSAID treatment to control symptoms in the month preceding screening	RCT, flare design
Weaver, 1995[90]	Investigative sites; USA	328 males & females, >50y	Knee	ACR clinical criteria-diagnostic	RCT, flare design
Wiesenhutter, 2005[59]	Medical Centres; USA	528 males & females, 40-89y	Knee and hip	ARA functional class, I, II, or III	RCT, flare design
Williams, 2001[60]	Clinical sites; USA	718 males & females, mean 61-62y	Knee	ACR clinical and radiographic criteria I-III	RCT, flare design
Wittenberg, 2006[61]	Centres (not specified) ; Germany	364 males & females, 50y	Knee	Moderate to severe symptomatic OA of the knee according to ACR criteria.	RCT, flare design
Yeasted, 2014[62] (Pooled, abstract)	USA	219 (merged observational), 137 (merged trial)>40y	Not specified	ACR criteria-diagnostic	2 longitudinal observational studies, placebo arms of 2 clinical trials
Yocum, 2000[77]	USA, 62 study centres	774 males & females, ≥40y	Knee or hip	Diagnosis confirmed by XR and clinical symptoms (not further specified)	RCT, flare design
Young, 2014[63] (abstract)	Multicenter,	305 males & females, >40y	Knee or hip	KL 2-3	RCT, flare design
Zhao, 1999[64]	Centre (not specified); USA, Canada	1004 males & females, ≥18y	Knee	ACR Functional capacity classification I-III	RCT, flare design
NON-DRUG WITHDRAWAL DESIGN STUDIES					

Atukorala, 2016[78] (abstract) Atukorala, 2016[25] (abstract)	Not specified, USA + Australia + Sri Lanka	213 males & females, mean age 62y 345 males & females, mean age 62y	Knee	Not specified	3-month, web based longitudinal follow up study
Bartholdy, 2016[79]	OA out-patient clinic, Denmark	131 males & females, ≥40y	Knee	Radiographic evidence of OA (severity no defined) and BMI between 20-35 kg/m^2	RCT
Bassiouni 2015[80] (abstract)	Not specified, Egypt	60 participants not further specified	Knee	Not specified	Observational
Cibere, 2004[86] Cibere, 2005[87]	Community, Canada	137 males & females, mean age 65y (43-88) for placebo and 64y (40-83) for glucosamine group	Knee	KL ≥2 on anteroposterior radiograph	RCT
Conrozier 2012[66]	Hospital-rheumatology unit, France	44 males & females, mean age 67.6y	Knee	Radiographic evidence of knee OA-not further defined	Observational
D'Agostino 2005[67]	Hospital-European multicentre	600 males & females, ≥18y	Knee	KL grade 1-4	Observational
Erfani, 2014[44] abstract) Erfani, 2014[81] (abstract) Ferreira[82] 2016	Australia	268 males & females, mean age 62y 345 males & females, ≥40y	Knee	ACR criteria- meet at least one, KL ≥2	Web based cross over

Hunter 2014[83] (abstract) Makovey 2015[84] (Protocol)					
Jawad, 2005[68]	GPs in France	3000 (for GP study) males & females	Knee	Not defined	n/a, review of surveys. Definition relates to survey of 3000 French GPs
Marty 2009[20]	Community and hospital, France	6085+641males & females, mean age 66.4y (10.9) for flare group, 66.2y (10.2) no flare group	Knee	OA diagnosis based on ACR criteria- severity not defined	Observational
Murphy, 2015[69]	Community based, pain clinics; USA	45 males & females, 37-83y	Knee	ACR criteria- severity not defined	Qualitative
Parry, 2017[85]	Community, UK	719 males & females, ≥50y	Knee	Self-reported knee pain in previous 12 months	Observational
Ricci 2005[54]	Community, USA	329 males & females, 40-65y	Knee and hip	Clinical evidence of OA- severity not defined	Nested case control
Wise 2010[70]	Primary care, hospital, USA	303 males & females, ≥50y	Knee and hip	Signal joint pain in a hip or knee on at least 15 out of the 30 days prior to enrolment- not further defined	Observational

Zhang 2009[71]	Primary care, hospital, USA	303 males & females, ≥50y	Knee and hip	Signal joint pain in a hip or knee on at least 15 out of the 30 days prior to enrolment-not further defined	Observational
Zhang 2011[72] (abstract)	Not specified	52 males & females, median age 63, (50-72y)	Knee	KL>2	Case-crossover
Zobel, 2016[92]	Hospital databases, Australia	297 males & females, >40y	Knee	ACR criteria, KL ≥2, or patellofemoral OA on radiograph	Web based case-cross over
Acronyms: KL- Kellgran and Lawrence RCT- Randomised Controlled Trial USA- United States of AmericaACR- Arthritis Center Research ARA- American Rheumatism Association GP- General Practitioner					

214

215

Twenty-one included mixed knee and hip OA groups[24, 29, 31, 37-39, 42, 45-47, 54, 55, 57-59, 63, 71, 73, 75, 77]. In total, 46 publications used a drug withdrawal RCT design[24, 26-32, 34-43, 45-53, 55-64, 73-77, 88-91], four of which were pooled studies[28, 32, 41, 62] and one used a cohort drug withdrawal design[33] (Table 1). The remaining 22 publications included seventeen observational studies[20, 25, 44, 54, 65-67, 70-72, 78, 80-85], three RCTs[79, 86, 87], one survey[68] and one qualitative interview study[69]. Nine of the included studies were abstracts[25, 44, 62, 63, 72, 78, 80, 81, 83]. Two abstracts were removed as the corresponding full text article was available[69, 92]. Studies using pooled data or the same dataset were included if they used different definitions of OA flare[28, 44, 52, 53, 62, 65, 70, 71, 74].

Rationale given for flare definitions

Six of the included studies gave rationale for the definition used[20, 54, 56, 69, 85, 86]. None of the definitions were based on a consensus procedure. Marty et al[20] and Scott-Lennox et al[56] were the only studies that undertook empirical investigation of flare definitions. The study by Marty et al[20] was the only study specifically designed to validate a diagnostic tool for knee OA flares. Potential factors associated with flare-ups were identified, for example, knee swelling and the authors used a logistic regression analysis to assign a weight to each of the items identified. A flare up score was determined using a general practitioner database and this was

then validated using a rheumatologist database. Pain was not included in the final model.

Scott-Lennox et al[56] sought to test whether four measures for flare intensity (patient's self-assessment of pain scores, physician's assessment of pain scores, patient's global OA assessment and physician's global OA assessment) could be combined to form a reliable and valid index using data from an RCT using a confirmatory factor analysis. The authors produced three flare intensity groups (low, moderate and severe) and highlighted how these could be used to examine treatment effects.

Cibere[86] outlined face validity checks. It was specified that the flare definition had been determined by study rheumatologists to be a clinically important change in the WOMAC score. The definition used by Murphy et al[69] was informed by two studies[28, 53] which used a drug withdrawal design and from the research team's own experience. Ricci et al[54] used a combination of data-driven and clinical judgement approaches to establish an agreed cut point. Parry et al based their definition on OA flare design studies and flare definitions used in other chronic disease such as back pain and COPD.

Flare definitions in drug withdrawal studies

Terminology used

The majority of publications using a drug withdrawal design used the term “flare” in their description[24-30, 32, 33, 36-43, 45-49, 51, 53, 55-64, 74-77, 88-91] (n=42; Table 2).

Table 2: Definition, terminology and measurement instruments used in all included studies

DRUG WITHDRAWAL STUDY DESIGN							
First author	**Terminology used**	**Amount of change in symptoms/signs (symptom/sign: measurement instrument; operational definition)**	**Minimum absolute level of symptoms/signs (symptom/sign: measurement instrument; operational definition)**	**Speed of onset**	**Duration**	**Change in medication/healthcare use**	**Reference / rationale**
Altman, 2015[24]	"Flare"	**Pain**: WOMAC Pain subscale (0-100); increase ≥15mm	**Pain:** WOMAC Pain subscale; ≥40mm	Not specified	Not specified	Not specified	None
Baer, 2005[26]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale (0-20); increase ≥2 points and ≥25%	**Pain:** WOMAC Pain score (0-20); ≥6 and ≥1 item rated 'moderate, severe, or extreme'	Interval between screening and baseline re-measurement unclear	Not specified	Not specified	None
Baraf, 2011[27]	"Flare"	**Pain on movement:** VAS (0-100mm); increase ≥5mm	Not specified	1 week washout	Not specified	Not specified	None
Battisti, 2004[28]	"Flare"	**Global assessment (investigator):** single item, 5-point LK; Worsening ≥1 point	**Pain:** VAS (0-100mm); ≥40mm	Not specified	Not specified	Not specified	None

Bingham, 2007[29] Bingham 2011[75]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); increase ≥15mm (2) **Global assessment of disease status (investigator):** single item, 5-point LK; Worsening ≥1 point	(1) **Pain walking on flat surface:** ≥40mm on WOMAC VAS3.0 Q1 (0-100) (2) **Global assessment (investigator)**: single item, 5-point LK; fair, poor, very poor (*acetaminophen users only*) (3) **Global assessment of disease status (patient):** VAS 0-100mm; ≥40mm (*acetaminophen users only*)	Not specified	Not specified	Not specified	None
Birbara, 2006[30]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global assessment (investigator):** single item, 5-point LK; Worsening ≥1 point	(1) **Pain walking on flat surface:** WOMAC VAS3.0 Q1 (0-100); ≥40mm (2) **Global assessment (investigator):** single item, 5-point LK; Fair, poor or very poor (*paracetamol arm only*)	4-15 day washout	Not specified	Not specified	None
Bocanegra, 1998[31]	"Worsening of symptoms"	Two out of the following three: (1) **Global assessment (physician):** single item, 5-point LK; Increase ≥1 grade (2) **Global assessment (patient):** Patients global assessment (current symptoms and limitation of activity) 5-point LK; Increase ≥1 grade (3) **Composite index:** Lequesne OA Severity Index (0-24); Increase ≥2 points	(1) **Global assessment (physician):** single item, 5-point LK; 'poor/very poor' (2) **Global assessment (patient):** Patients global assessment (current symptoms and limitation of activity) 5-point LK; 'poor/very poor' (3) **Composite index:** Lequesne OA Severity Index (0-24); ≥7	3-14d washout	Not specified	Not specified	None

Boswell, 2008[32]	"Flare"	(1) **Pain walking on flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global assessment (patient):** Patient Global Assessment of Arthritis Condition (PGAC) (unspecified); Worsening ≥1 point	Not specified	Not specified	Not specified	Not specified	None
Brandt, 2006[33] (pilot studies)	"Flare"	Not specified	**Pain:** WOMAC LK Pain subscale (5-25); ≥15 points	5 half-lives of NSAID washout	Not specified	Not specified	None
Case, 2003[34]	Not used	(1) **Pain walking on flat surface:** VAS (0-100mm); Increase ≥10mm (2) **Ambulatory pain**; 5-point LK; worsening ≥1 point	Not specified	14d washout	Not specified	Not specified	None
Day, 2000[73]	Not used	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2) **Global Assessment (investigator):** single item, 5-point LK; worsening ≥1 point (3) **Global assessment (patient):** VAS (0-100mm); increase ≥15mm (*acetaminophen users only*)	(1) **Pain walking on a flat surface**: WOMAC VAS Q1 (0-100mm); ≥40mm; (2) **Global Assessment (investigator):** single item, 5-point LK; 'Fair, poor, or very poor'; (3) **Global assessment (patient):** VAS (0-100mm); ≥40mm	Longer than 5 plasma half-lives washout	Not specified	Not specified	None
Ehrich, 1999[35]	Not used	**Pain:** VAS (0-100mm); increase ≥15mm	**Pain:** VAS (0-100mm); ≥40mm	Longer than 5 plasma half-lives washout of NSAID	Not specified	Not specified	None

Essex, 2012[36]	"Flare"	(1) **Global Assessment (Physician):** 5-point LK; increase ≥1 grade (2) **Global Assessment (patient):** 5-point LK; increase ≥1 grade	(1) **Global Assessment (Physician):** 5-point LK; 'Fair, poor or very poor' (2) **Global Assessment (patient):** 5-point LK; 'Fair, poor or very poor' (3) **Pain**: VAS (0-100mm); 40-90mm	48 hour withdrawal	Not specified	Not specified	None
Essex 2013[76]	"Flare"	Not specified	(1) **Global Assessment of arthritis (Physician):** Minimum rating of 3 (2) **Global Assessment of arthritis (patient):** Minimum rating of 3 (3) **Pain**: VAS (0-100mm); 40-90mm	48 hour withdrawal	Not specified	Not specified	None
Gibofksy, 2014[37]	"Flare"	**Pain:** WOMAC Pain VAS; increase ≥15mm	**Pain:** WOMAC Pain VAS; ≥40mm	Not specified	Not specified	Not specified	None
Gineyts, 2004[38]	"Flare"	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); increase ≥15mm (2)**Global Assessment (investigator):** 5-point scale: worsening ≥1 point	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm); ≥40mm	5 half-lives of NSAID washout	Not specified	Not specified	None
Goldberg, 1988[39]	"Flare"	(1) **Pain:** Investigator assessed pain grade (None/mild/mod/severe): (i) at rest, (ii) on passive motion, (iii) on palpation, (iv) weight bearing; increase ≥1 grade in two items OR increase ≥2 grade in one item	Not specified	2-14 day washout until flare	Not specified	Not specified	None
Gottesdiener, 2002[40]	"Flare"	(1) **Pain on walking**: VAS (0-100mm); increase ≥15mm (2)**Global Assessment (Investigator):** 5-point LK; Increase ≥1 point	(1) **Pain on walking**: VAS (0-100mm); ≥40mm	3-15 day washout	Not specified	Not specified	None

Hochberg, 2011[41]	"Flare"	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm)**;** Increase ≥15mm (2) **Global Assessment (patient):** 5-point LK; worsening ≥1 point	(1) **Pain walking on a flat surface:** WOMAC VAS Q1 (0-100mm)**;** ≥40mm	Not specified	Not specified	Not specified	None
Katz, 2010[42]	"Flare"	Not specified	**Pain**: Pain score (0-10); ≥5	Not specified-washout until flare occurred	Not specified	Not specified	None
Kivitz, 2001[43]	"Flare"	**Pain:** Patients Assessment of Pain Score (0-10) (unspecified); increase ≥2 points	**Pain:** Patients Assessment of Pain Score (0-10) (unspecified); ≥5	5 drug half-lives or 48 hours	Not specified	Not specified	None
Kivitz, 2004[74]	"Flare"	(1) **Pain on walking:** VAS (0-100mm); worsening ≥15mm (2) **Global Assessment (investigator):** 5-point LK; worsening ≥1 point	Not specified	NSAID dependent half-life washout	Not specified	Not specified	None
Leung, 2002[45]	"Flare"	(1) **Pain on walking on a flat surface:** WOMAC VAS Q1 (0-100mm); Increase ≥15mm (2) **Global Assessment (investigator):** 5-point LK; worsening ≥1 point	(1)**Pain on walking on a flat surface:** WOMAC VAS Q1 (0-100mm); ≥40mm (2) **Global Assessment (patient):** **(**0-100mm); ≥40mm (*acetaminophen users only*) (3) **Global Assessment (investigator):** 5-point LK; 'Fair, poor, or very poor' (*acetaminophen users only*)	Determined by drug half-life washout	Not specified	Not specified	None

Luyten, 2007[46]	"Flare"	(1) **Global Assessment (Patient):** 5-point LK; Increase ≥1 grade (2) **Global Assessment (physician):** 5-point LK; increase ≥1 grade (3) **Composite definition:** Lequesne Osteoarthritis Severity Index (0-24); increase ≥2 points	(1) **Global Assessment (Patient):** 5-point LK; 'Fair, poor or very poor' (*Not on treatment – 'Poor or very poor'*) (2) **Global Assessment (physician):** 5-point LK; 'Fair, poor or very poor' (*Not on treatment – 'Poor or very poor'*) (3) **Composite definition:** Lequesne Osteoarthritis Severity Index (0-24); ≥7 (4) **Pain**: VAS (0-100mm); ≥40mm	2-14 day washout	Not specified	Not specified	None
Manicourt, 2005[47]	"Flare"	**Pain when walking on a flat surface:** VAS (0-100mm) ; ≥10mm	Not specified	7-10 days washout	Not specified	Not specified	None
Mazzuca, 2002[48]	"Flare"	**Pain on standing**: WOMAC LK Pain Q5 'severe or extreme' after the washout AND decreased after resumption of usual analgesic drugs and/or NSAIDs	Not specified	Drug washout 5 half lives	Not specified	Not specified	None
McIlwain, 1989[49]	"Flare"	No measurement instrument: Increase in pain on motion, swelling, tenderness, redness and/or heat (unspecified if patient/physician/investigator reported)	Not specified	2-14 day washout	Not specified	Not specified	None
Mendelsohn, 1991[50]	"Worsening of arthritis condition"	(1) **Pain:** Pain scale (0-3) (0=none, 3=severe); worsening score (2) **Global (physician):** (0-100); worsening score	Not specified	Up to 14 days washout	Not specified	Not specified	None

Moskowitz, 2006[51]	"Flare"	(1) **Global assessment (patient)**: 5-point LK; increase ≥1 grade (2) **Global Assessment (physician)**: 5-point LK; ≥ 1 grade increase (3) **Composite index:** Lequesne OA Severity Index (0-24); increase ≥2 points	(1) **Global assessment (patient):** 5-point LK; '(Fair), poor, or very poor' (2) **Global Assessment (physician):** 5-point LK; '(Fair), poor or very poor' (3) **Composite index**: Lequesne OA Severity Index (0-24); Minimum ≥7 (4) **Pain walking on a flat surface:** VAS (0-100mm); ≥40mm	NSAID washout of 5 half-lives or at least 2 days	Not specified	Not specified	None
Pareek, 2009[52]	"Flare-up"	(1) **Pain:** 11-point NRS; increase ≥ 2 points during previous 2-5 days (2) Signs and symptoms suggestive of inflammation, morning stiffness and nocturnal pain interfering with sleep	**Pain**: Pain intensity of at least 4 on a 11-point NRS during physical activity for past 24 hours	Placebo washout for 24-48 hours	2-5 days	Not specified	None
Pareek, 2010[53]	"Flare"	Flare symptoms noted but not part of definition: morning stiffness, erythema, nocturnal pain, and swelling/inflammation	(1) **Pain with physical activity**: VAS 0-10; ≥6 (2) **Composite index:** WOMAC Total LK; ≥25. (3) **Composite index**: Lequesne OA Severity Index (0-24); ≥5	Not specified	2-5 days	Not specified	None
Roth, 2004[88]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale (0-20); increase ≥2 points and ≥25%	**Pain:** WOMAC LK3.1 Pain subscale (0-20); Score ≥'moderate' on at least 1 of the 5 items, (ii) Pain score ≥6	Washout period of at least 3 days per week past month	Not specified	Not specified	None

Rother, 2007[91]	"Flare"	(1) **Pain on walking:** VAS (0-100mm); Increase ≥15mm (2) **Global Assessment (patient):** 5-point LK; increase ≥1 grade	(1) **Pain on walking:** VAS (0-100mm); ≥40mm (2) **Global Assessment (patient):** 5-point LK; 3-5	Not specified	Not specified	Not specified	None
Schnitzer, 2005[55]	"Flare"	No tool: increase in pain	**Pain:** VAS (0-100mm); ≥40mm	Not specified	24 hours	Not specified	None
Scott-Lennox, 2001[56]	"Flare"	(1) **Pain:** VAS (0-100mm)**;** ≥20mm (2) **Pain (physician):** 4-point LK; worsening ≥1 point (3) **Global Assessment (patient):** 4-point LK; worsening ≥1 point (4) **Global Assessment (physician)**:4 point LK; worsening ≥1 point	(1) **Pain:** VAS (0-100mm)**;** ≥40mm at baseline) (2) **Pain (physician):** 4-point LK; ≥2 (3) **Global Assessment (patient):** 4-point LK; ≥2 (4) **Global Assessment (physician)**: 4 point LK; worsening ≥2	14 day washout	Not specified	Not specified	Confirmatory Factor Analysis
Simon, 2009[89]	"Flare"	**Pain:** WOMAC LK3.1 Pain subscale; increase ≥2 and ≥25%	**Pain:** WOMAC LK3.1 Pain subscale; ≥'moderate' on ≥1 item	14 day washout	Not specified	Not specified	None
Silverfield, 2002[57]	"Flare"	**Pain:** No measurement tool; significant increase	Not specified	Not specified	Not specified	Pain requiring supplemental analgesic medication and/or an increase in NSAID dose	None
Strand, 2011[58]	"Flare"	**Global Assessment (patient):** 5-point LK; Increase ≥1	(1) **Global Assessment (patient):** 5-point LK; 'Fair, poor or very poor' (2) **Pain**: (0-10 NRS); ≥4 but <9 (3) **Global Assessment (physician)**: 5-point LK; 'Fair, poor or very poor'	14 day washout	Not specified	Not specified	None

Weaver, 1995[90]	"Flare"	(1) **Global Assessment (Physician)**: 5-point Likert; increase ≥1 grade (2) **Global Assessment (patient)**: 5-point LK; increase ≥1 grade (3) **Pain:** Worsening pain on motion and weight bearing	(1) **Global Assessment (Physician)**: 5-point Likert; ≥2 (2) **Global Assessment (patient)**: 5-point LK; ≥2	2-14 day washout	Not specified	Not specified	None
Wiesenhutter, 2005[59]	"Flare"	(1) **Pain on walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); increase ≥15mm (2) **Global Assessment (Investigator):** 5-point LK; worsening ≥1 unit	(1) **Pain on walking on flat surface:** WOMAC VAS3.0 Q1 (0-100mm); ≥40mm	Not specified	Not specified	Not specified	None
Williams, 2001[60]	"Flare"	(1)**Global Assessment (patient):** 5-point LK; Increase ≥1 point (2) **Global Assessment (physician)**: 5-point LK; increase ≥1 point(3) **Composite Index**: Lequesne OA Severity Index (0-24); Increase ≥2 points	(1) **Global Assessment (patient):** 5-point LK; '(Fair), poor or very poor' (2) **Global Assessment (physician)**: 5-point LK; '(Fair), poor or very poor' (3) **Composite Index**: Lequesne OA Severity Index (0-24); ≥7 (4) **Pain:** VAS (0-100mm); ≥40mm	2-14 days	Not specified	Not specified	None
Wittenberg, 2006[61]	"Flare"	**Pain:** VAS (0-100mm); Increase ≥10mm	**Pain:** VAS (0-100mm); ≥40mm	2-7 day washout	Not specified	Not specified	None
Yeasted, 2014[62] (Pooled, abstract)	"Flare"	**Pain:** 0-10 NRS; Increase ≥2 points over the mean pain score from the previous 3 days	**Pain:** Average daily 0-10 NRS; 4-9	Not specified	Not specified	Not specified	None
Yocum 2000[77]	"Flare"	**Disease activity** (1) **Global (Investigator)**: Reduction of ≥ 1 grade (2) **Global Assessment (Patient)**: 100-mm VAS; Increase of ≥10mm	Not specified	≥3 days washout	Not specified	Not specified	None

		(3) **Pain: Overall assessment (patient):** 100-mm VAS; ≥35mm					
Young, 2014[63]	"Flare"	(3) **Pain:** WOMAC pain subscale; increase >15mm	**Pain:** WOMAC Pain subscale >40mm	Not specified	Not specified	Not specified	None
Zhao, 1999[64]	"Flare"	No measurement tool: Worsening of signs and symptoms after discontinuation of NSAIDs of analgesics	Not specified	2-7 day washout	Not specified	Not specified	None
NON-DRUG WITHDRAWAL STUDY DESIGN							
Atukorala, 2016[78] (abstract) Atukorala, 2016[25] (abstract)	"Flare"	**Pain:** (10-point NRS); increase >2 points from the mildest knee OA pain intensity reported at day 0	Not specified	Not specified	Not specified	Not specified	None
Bartholdy, 2016[79]	"Flare"	Not specified	**Pain:** (10-point NRS): Pain >5	Not specified	Not specified	Not specified	None
Bassiouni 2015[80] (abstract)	"Flare"	Not specified	**Global Assessment (physician):** KOFUS ≥7	Not specified	Not specified	Not specified	None
Cibere, 2004[86] Cibere, 2005[87]	"Flare"	(1) Patients perception of worsening of symptoms (2) **Pain walking on flat surface**: WOMAC VAS3.0 Q1 (0-100mm); increase ≥20mm (3) **Global Assessment (physician):** 5-point LK; worsening ≥1 grade	Not specified	Not specified	Not specified	Not specified	Definition determined by study rheumatologists to be a clinically important change in WOMAC-Ehrich2000/

							Bellamy 1998
Conrozier 2012[66]	"Flare"	Fulfilled 4 following criteria: (1) **Pain:** No measurement tool; 'sudden aggravation of knee pain' (2) causing nocturnal awakenings, (3) clinical evidence of effusion.	Not specified	Sudden aggravation of knee pain, whose beginning was identifiable	Not specified	Not specified	None
D'Agostino 2005[67]	"Flare"	Not specified	**Pain intensity during physical activity:** VAS-(0-100mm); ≥40mm	Not specified	48 hours	Not specified	None
Erfani, 2014[44] abstract) Erfani, 2014[81] (abstract) Ferreira[82] 2016 Hunter 2014[83] (abstract) Makovey 2015[84] (Protocol)	Exacerbation	**Pain:** VAS (0-100mm); Increase ≥20mm from mildest pain score reported at baseline	Not specified	Not specified	Not specified	Not specified	None

Jawad, 2005[68]	Exacerbation	**Pain symptoms:** Increased morning stiffness, night pain and synovial fluid effusion	Not specified	Not specified	Not specified	Not specified	None
Marty 2009[20]	"Flare"	No measurement tool: Morning stiffness >20mins, nocturnal awakening, limping, knee swelling, increased warmth, effusion	Not specified	Not specified	48 hours	Not specified	Regression analysis of cross-sectional data to validate proposed flare criteria
Murphy, 2015[69]	"Flare"	(1) Investigator definition: Inadequate pain relief for an episode of intense pain that is usually brought on by too much activity. (2) Participant definitions: Described in terms of pain quality, timing (onset and duration), antecedents and consequences. (3) **Pain magnitude:** increase in pain or 'intense' or 'severe' level of pain	**Pain:** ≥40 of 100mm or ≥4 of 10 on NRS	Patients described: 'Quick' or 'sudden'	Patients: 10 seconds to 15 minutes	Patients: Rest or take additional medication	For investigator definition: Battisti 2004, Pareek 2010. Plus researchers own experience.
Parry, 2017[85]	"Flare"	**Pain:** Recalled worst pain intensity in previous 6 months 0-10 NRS; ≥5	**Pain:** Recalled worse pain to be ≥2 points higher than recalled average pain (0-10 NRS) in previous 6 months	Not specified	Not specified	Not specified	Based on previous studies defining knee flares in OA and flares in diseases such as back

							pain and COPD.
Ricci 2005[54]	"Flare up"	**Pain:** Self-reported flare severity rating 0-10 NRS; increase ≥2 point over usual pain severity	Not specified	Not specified	Not specified	Not specified	Based on statistical analysis and clinical judgement
Wise 2010[70]	"Flare"	Not specified	**Pain:** WOMAC Pain subscale (0-10); score in highest 30% of all WOMAC scores	Not specified	Not specified	Not specified	None
Zhang 2009[71]	"Exacerbation or flare"	Not specified	(1) Pain: WOMAC pain subscale 0-10 (total score of 50 normalised to a 0-10 scale); score of ≥5, a score corresponding to highest 33% of all WOMAC scores	Not specified	Not specified	Not specified	None
Zhang 2011[72] (abstract)	"Exacerbation"	**Pain:** WOMAC Pain score VAS (0-500); increase ≥100 units	Not specified	Not specified	Not specified	Not specified	None
Zobel, 2016[92]	Exacerbation	**Pain:** 0-10 NRS; Increase ≥2	(1) Disabling pain	Not specified	8 hours	Not specified	None

Acronyms:
COPD- Chronic Obstructive Pulmonary Disease
KOFUS- Knee Osteoarthritis Flare-up Score
NRS-Numerical Rating scale
VAS- Visual Analogue Score
WOMAC- Western Ontario and McMaster Universities Osteoarthritis Index
LK-Likert scale

268

One study used the term “flare-up”[52], two studies referred simply to “worsening of symptoms” [31, 50] and three studies used no specific label[34, 35, 73].

Coverage of key components

Onset/worsening of symptoms and signs beyond normal-day-to-day variability: Forty-four studies included onset or worsening of signs and symptoms as part of their definition[24, 26-32, 34-41, 43, 45-53, 55-64, 73-75, 77, 88-91]. All studies included increased pain intensity in their definition. A further two[52, 53] specified further signs and symptoms. These included swelling, inflammation, erythema, morning stiffness and nocturnal pain. No studies quantified day-to-day variability.

Twenty-six measurement tools were used to define onset/worsening of symptoms and signs. The most commonly used tools were the Western Ontario & McMaster Universities Arthritis index (WOMAC) Q1 (pain on walking on flat surface) 100mm Visual Analogue Scale (VAS) (n=9)[29, 30, 32, 38, 41, 45, 59, 73, 75] and the Investigator Assessment of Disease Status (n=11)[28-30, 38, 40, 45, 59, 73-75, 77] (Table 3). Thirty-four studies used only single item measurement tools[27-30, 32, 34-43, 45, 47, 48, 50, 52, 55, 56, 58, 59, 61-63, 73-77, 90, 91], 5 used multi-item[31, 46, 51, 53, 60] and 5 used both single and multi-item tools[24, 26, 33, 88, 89].

Table 3: Summary of number and type of single and multi-item measurement tools used.

Single item scales:	
Pain on activity:	WOMAC Q1 3.0 VAS 'pain on walking on a flat surface' (0-100mm) [n=11] Pain on walking VAS (0-100mm) [n=5] Pain on movement VAS (0-100mm); Ambulatory pain (5-point Likert); Pain with physical activity VAS 11-point scale [n=2]
Pain (not further specified):	Pain VAS (0-100mm) [n=15] Patients Assessment of Pain Score (0-10); Pain Scale (0-3); Pain NRS (0-10) [n=11]
Standing knee pain	Item 5 WOMAC pain scale [n=1]
Global rating (physician/ investigator)	Investigator Assessment of Disease Status [n=11] Physicians Global Assessment of Arthritis [n=6] Physician Global Assessment of OA [n=2] Physician Global Assessment of Disease Status [n=2]; Investigator Assessed Pain Grade; (Physician) Overall Disease Activity (0-100); Physicians Pain Assessment (4-point LK) [n=3]
Global rating (patient)	Patients Global Assessment of Arthritis [n=7] Patient Global Assessment of OA [n=3] Patient Global Assessment of Disease Status [n=4]
Multiple-item scales:	
	Lequesne OA Severity Index [n=5] WOMAC LK3.1 (0-20) [n=3] WOMAC LK Pain subscale (0-25); WOMAC OA Index

	Questionnaire [n=1]; WOMAC knee pain score (0-500) [n=7]; KOFUS (0-14) [n=1]

N, number of included studies; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; VAS, visual analogue scale; OA, osteoarthritis; KOFUS, Knee Osteoarthritis Flare-up Score.

In addition, the format of global ratings appears to be variable as is use and reporting of the WOMAC[93]. However, despite the exact format of reporting being inconsistent, in general, studies used single items in 4 areas – pain on activity, pain (not necessarily on activity), physician/investigator global rating and patient global rating.

Temporal characteristics: None of the included drug withdrawal design studies reported a specific time for defining the speed of onset of symptoms. However, they did describe withdrawal or 'washout' periods whereby, after withdrawal of usual medication, participants were given a certain time frame in which to experience 'flare' symptoms in order that they were entered into the study. In total 30 of the studies specified a withdrawal period[27, 30, 31, 33-36, 38-40, 43, 45-52, 56, 58, 60, 61, 64, 73, 74, 76, 77, 88-90].

Four studies specified a time period for minimum duration of symptoms which ranged from 24 hours to 5 days[52, 53, 55, 57].

Change in medication or healthcare usage: Only one study used increase in medication as part of their definition; 'pain requiring supplemental analgesic medication and/or an increase in NSAID dose'[57].

Additional domains: Thirty-six studies included a minimum threshold which was usually a minimum level of pain that was required before the participant was considered to have a flare[24, 26, 28-31, 33, 35-38, 40-43, 45-47, 51-53, 55, 56, 58-63, 73, 75, 76, 88-91]. There was general concordance with the minimum thresholds that different measurement tools used with a few exceptions. A threshold of 40mm on a 0-100mm scale was used in eight of ten studies using the WOMAC VAS 3.0 Q1 'pain on walking on a flat surface'[29, 30, 38, 41, 45, 59, 73, 75] and four of fourteen studies using the Patient Global Assessment of Disease Status[29, 45, 73, 75]. In studies using various forms of investigator/physician global assessment, the majority adopted a minimum threshold for a flare of 'fair, poor or very poor' [29, 30, 45, 73]. The minimum threshold on the Lequesne index (0-10) was either five[53] or seven[46, 51, 60].

Flare definitions in non-withdrawal flare/ discontinuation studies

Terminology used

"Flare" was the term most common used in non-withdrawal design studies[20, 25, 66, 67, 69, 70, 78-80, 85, 87](n=11) (Table 2). One study used the term "flare-up"[54], eight used "exacerbation"[44, 65, 68, 72, 81-84] (five publications were from the same

team) and one referred to both “exacerbation” and “flare”[71]. None referred to “worsening of symptoms” or did not use any specific label.

Coverage of key components

Onset/worsening of symptoms and signs beyond normal-day-to-day variability:

Sixteen of twenty-two studies used onset or worsening of symptoms in their definition[25, 44, 54, 66, 68, 69, 72, 78, 81-87, 92]. Two studies did not use pain intensity as part of its definition[20, 80]. Three studies included symptoms other than pain in their definition[20, 66, 68]. These included nocturnal awakenings, effusion, morning stiffness, night pain, limping, and warmth.

The Murphy et al[69] study included an investigator definition of flare but also sought to describe patient experience of flares through face to face individual interviews. Both investigator and patient definitions included onset/worsening of symptoms and signs however there was no differentiation from day-to-day variability.

Seven studies used a measurement tool to define onset of signs and symptoms (Table 3). These included the Pain NRS (0-10)[25, 54, 65, 78, 85], WOMAC knee pain score VAS (0-500)[72], pain walking on a flat surface (WOMAC)[86, 87], Global

Assessment of Disease Status (physician) (Likert 5-point scale)[86, 87], and knee pain VAS not further specified (0-100)[44, 81-84].

Temporal characteristics: Only one study set a definition for speed of onset, describing this only as 'sudden' with no further specification[66]. Patients in the Murphy et al study used the terms 'quick' and 'sudden' to describe flare onset[69]. Three studies specified a minimum duration of symptoms ranging from 8 to 48 hours[20, 65, 67]. In the Murphy et al study patients described duration of between 10 seconds to 15 minutes[69].

Change in medication/healthcare usage: No studies used change is medication or healthcare usage as part of their definition. However, in Murphy et al patients reported either taking rest or using additional medication[69].

Additional Domains: Two studies defined distribution-based minimum thresholds for flare as the highest 30%[72] or highest 33%[73] of WOMAC Pain Subscale scores among participants in the Longitudinal Examination of Arthritis Pain (LEAP) cohort (total score out of 50 was normalised to a 0-10 scale).

DISCUSSION

Flares in OA are recognised in existing clinical guidance[94] and reviews[95, 96] but typically merit little more than a passing mention. Our analysis of the definitions has resulted in the findings of common core domains which will be useful for developing an agreed consensus definition for OA flare. From a clinical perspective, a unified definition of a flare could enable clinicians to provide prompt, rationalised and focussed treatment. This could also have implications for delivery of self-management strategies involving patients and how episodic management is advocated by clinical guidelines. Our review was motivated by an interest in seeking greater clarity on how these phenomena might be defined by undertaking a broad search strategy, noting that similar efforts have been pursued in other chronic diseases. While we found no current single, agreed definition of OA flare, our review of 69 published studies suggests a number of common domains which may capture cardinal features. These were: onset/worsening of symptoms and signs, attainment of a minimum symptom threshold during flare, speed of onset/worsening, and duration of elevated symptoms/signs. However, we found considerable variation in how these domains have been operationalised for measurement suggesting the need for further conceptual clarification and consensus.

Each potential cardinal feature of OA flare presents different challenges for achieving consensus. The goal of an agreed composite definition is to facilitate both

reproducible and comparable research, whilst enabling more consistent recognition and identification of these phenomena in routine practice. The heterogeneity of OA should also be considered in any definition of a flare-up. Most studies included in our review required an increase in pain over 'usual' or 'baseline' intensity. Although this was measured using a wide range of measurement instruments several studies selected an increase of 2 or more points on a 0-10 scale providing a possible starting point for consensus. Yet this possible 'signal' is arguably difficult to interpret without also considering the amount of background 'noise', i.e. within-person diurnal[97] and day-to-day variability[98], and the absolute level ('minimum threshold') of pain during a flare. There was general concurrence with the minimum threshold that was adopted, for example, 40mm on a 0-100mm scale and this may indicate the potential level of minimally important clinical difference. In the study by Marty et al an increase in pain was not independently associated with flare-up after adjusting for other potential features[20]. However, the study by Marty et al[20] and Scott-Lennox et al[56] were the only two studies we found that had attempted to derive and/or validate a prediction model for OA flares. Interestingly their approaches have not been widely adopted which suggests the complexity of reaching a widely accepted model. Further research on detecting flares over within-person 'normal' variability by collecting frequent repeated measures of pain intensity may be valuable but this approach would not be feasible when identifying flares presenting at the point of care in routine clinical practice. Instead, this may have to rely on the judgement of

the patient and/or clinician, the approach used, for example, in defining exacerbations in COPD[1]. A similar consideration surrounds the speed of onset, which was not well defined by studies in our review. Drug withdrawal design studies specified washout periods between 2-15 days but this is unlikely to be synonymous with speed of onset. The remaining studies used terms such as 'sudden' and 'quick'. In COPD, for instance, a judgement around 'acute onset' or 'sudden onset' appears to be acceptable for clinical recommendations but we would add that the speed of onset of OA flares ought to be considered also in relation to underlying biologically plausible mechanisms. Indeed presumed aetiology has been argued as a useful feature in defining acute exacerbations in COPD[99]. Minimum duration ranged from 8 hours to 5 days in our review however this was not widely reported. COPD definitions refer to a 'sustained worsening' of symptoms[2] but does not appear to be a feature in other chronic diseases. A minimum duration in OA may help distinguish flares from day-to-day variability. Increase in medication was not found to be a key component in this review despite it being a feature in other chronic diseases; AS[5], SLE[4, 100], Inflammatory Bowel Disease[101], COPD[1]. Interference with function did not emerge strongly from our review as a cardinal feature of OA flare. In other chronic musculoskeletal conditions, such as back pain, interference with function was not shown to be significantly associated with having a flare up[102] and this domain does not feature in the definitions of exacerbations or flares in diseases such as COPD[1, 2], asthma[3], AS[5] or SLE[4].

Our review has several strengths but also some weaknesses that deserve attention. We adopted a broad search strategy, covering a wide range of databases, and featuring bibliography checks, contact with authors, inclusion of conference abstracts, no language restrictions, and a minimal threshold (any description or definition of flare) for inclusion. Five studies that were included in a similar review by Cross et al[103] were not included in this study; four did not contain a clear definition of flare-up, including one which gave a definition of knee OA progression and the final paper by Sands et al[104] was not in our search but the original study was[58]. We did not, however, search the grey literature and we did not include some potential synonyms as search terms ('attack', 'episode', 'fluctuations') although these terms appeared often to relate to comorbidities and other phenomena (e.g. episodes of care) and would therefore have been a less efficient search strategy than relying on snowball references. Data extraction was performed by only a single reviewer. Nevertheless, we argue that our review provides a reasonably comprehensive summary of how 'flares' in OA have been described and defined in the medical literature. In comparison with Cross et al[103] our search strategy appeared comprehensive yet efficient – returning 69 included articles compared with 23. We feel that our review expands on the findings of the Cross et al review and adds strength to this important area. The majority of studies describe experimental 'flare design' trials in which flares are induced by drug withdrawal prior to enrolment and

randomisation. While intentional or unintentional reduction in usual analgesia may indeed be one trigger for flare, experimentally induced flares should not be assumed to represent 'naturally occurring' flares. Flare design trials, for example, are unlikely to capture change in management or healthcare usage that may be a common consequence of OA flares – something that is included in flare definitions in other conditions such as AS[5], SLE[4, 100], inflammatory bowel disease[101], and COPD[1].

A systematic review such as this cannot hope to resolve the need for a common conception and definition of flares in OA. Definitions for exacerbations of disease states are generally reached through a long process of consensus exercises involving key stakeholders, experts and patients in addition to appraisal of relevant literature from studies using multiple methods[6, 8, 105]. However, we believe that a consensus definition that is reliable, valid, and feasible and widely acceptable both clinically and for research purposes should now be sought. The cardinal features described in this review; onset/worsening of symptoms and signs, attainment of a minimum symptom threshold during flare, speed of onset/worsening, and duration of elevated symptoms/signs could help start this discussion. Furthermore, observational studies with repeated measures could give an important insight into the nature of these phenomena.

CONCLUSION

A broad range of ad-hoc definitions currently exist in the medical literature. The majority are from drug-withdrawal or flare-induced trials rather than 'naturally' occurring flares. The cardinal feature is pain intensity with minimum symptom threshold being another important feature. This review has identified the need to gain consensus on a common definition that can be used for research and clinical application.

Acknowledgements

We would like to thank you to Popay et al for allowing us to use their guidance on the conduct of narrative synthesis in systematic reviews. We would also like to thank Jo Jordan and Opeyemi Babatunde for their advice on conducting a systematic review.

Contributions

All authors were involved in conception and design of the study, analysis and interpretation of data, drafting the article, critical revision of the article for important intellectual content, final approval of the article. ELP and MJT extracted and synthesised data. ELP assembled the data.

GMP (g.m.peat@keele.ac.uk) takes responsibility for the integrity of the work as a whole from inception to finished article.

Data sharing statement

No unpublished data is available following this study

Role of the funding source

ELP is funded by a National Institute for Health Research (NIHR) In Practice Fellowship (IPF-2014-08-03). MJT received funding from a NIHR School for Primary Care Research Launching Fellowship and is currently funded by an Integrated Clinical Academic Programme Clinical Lectureship from the NIHR and Health Education England (HEE) (ICA-CL-2016-02-014).

The views expressed in this paper are those of the author(s) and not necessarily those of the NHS, the NIHR, HEE or the Department of Health. This paper presents independent research funded by the Arthritis Research UK Centre in Primary Care grant (Grant Number 18139).

Competing interest statement

GP received consultancy fees from InFirst plc and Good Relations plc.

Figure and Table Legends

Figure 1: PRISMA Flowchart

Table 1: Characteristics of all included studies

Table 2: Summary of number and type of single and multi-item measurement

Table 3: Definition, terminology and measurement instruments used in all included studies

Supplementary data: Database search strategy

REFERENCES

1 Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the diagnosis, management and prevention of COPD: GOLD 2016.

2 National Institute for Health and Care Excellence (NICE). Chronic obstructive pulmonary disease in over 16s: diagnosis and management (CG101). London: NICE 2010.

3 Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention: GINA 2015.

4 Ruperto N, Hanrahan L, Alarcón G, et al. International consensus for a definition of disease flare in lupus, *Lupus* 2011;20:453-62.

5 Stone MA, Pomeroy E, Keat A, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration, *Rheumatology* 2008;47:1213-8.

6 Bingham CO, Alten R, Bartlett SJ, et al. Identifying Preliminary Domains to Detect and Measure Rheumatoid Arthritis Flares: Report of the OMERACT 10 RA Flare Workshop, *The Journal of Rheumatology* 2011;38:1751-8.

7 Bykerk VP, Lie E, Bartlett SJ, et al. Establishing a Core Domain Set to Measure Rheumatoid Arthritis Flares: Report of the OMERACT 11 RA Flare Workshop, *The Journal of Rheumatology* 2014;41:799-809.

8 Bartlett SJ, Hewlett S, Bingham CO, et al. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus, *Annals of the Rheumatic Diseases* 2012;71:1855-60.

9 Taylor WJ, Shewchuk R, Saag KG, et al. Toward a valid definition of gout flare: Results of consensus exercises using delphi methodology and cognitive mapping, *Arthritis Care & Research* 2009;61:535-43.

10 Schmitt J, Spuls PI, Thomas KS, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials, *J Allergy Clin Immunol* 2014;134:800-7.

11 Holla JFM, van dL, Knol DL, et al. The association of body-mass index and depressed mood with knee pain and activity limitations in knee osteoarthritis: results from the Amsterdam osteoarthritis cohort, *BMC Musculoskeletal Disorders* 2013;14:296.

12 Collins JE, Katz JN, Dervan EE, et al. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative, *Osteoarthritis and Cartilage* 2014;22:622-30.

13 Leffondré K, Abrahamowicz M, Regeasse A, et al. Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators, *J Clin Epidemiol* 2004;57:1049-62.

14 Emrani PS, Katz JN, Kessler CL, et al. Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis, *Osteoarthritis and Cartilage* 2008;16:873-82.

15 Bartlett SJ, Ling SM, Mayo NE, et al. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis, *Arthritis Care & Research* 2011;63:1722-8.

16 Hawker GA, Stewart L, French MR, et al. Understanding the pain experience in hip and knee osteoarthritis – an OARSI/OMERACT initiative, *Osteoarthritis and Cartilage* 2008;16:415-22.

17 Arthritis Research UK. Osteoarthritis: Patient Information Booklet. 2012.

18 Bonnet CS, Walsh DA. Osteoarthritis, angiogenesis and inflammation, *Rheumatology* 2005;44:7-16.

19 Smith TO, Zou K, Abdullah N, et al. Does flare trial design affect the effect size of non-steroidal anti-inflammatory drugs in symptomatic osteoarthritis? A systematic review and meta-analysis, *Annals of the Rheumatic Diseases* 2016;75:1971-8.

20 Marty M, Hilliquin P, Rozenberg S, et al. Validation of the KOFUS (Knee Osteoarthritis Flare-Ups Score), *Joint Bone Spine* 2009;76:268-72.

21 Rutjes AS, Jüni P, Da Costa BR, et al. Viscosupplementation for osteoarthritis of the knee: A systematic review and meta-analysis, *Ann Intern Med* 2012;157:180-91.

22 Thomas J, Harden A, Newman M. Synthesis: Combining results systematically and appropriately. In: Gough A, Oliver S, Thomas J, eds. An introduction to systematic reviews. London: Sage publications limited 2013:191-2.

23 Popay J, Roberts H, S, A., et al. Guidance on the conduct of narrative synthesis in systematic reviews: A product of the ESRC methods programme Lancaster: ESRC Method Programme, 2006.

24 Altman R, Hochberg M, Gibofsky A, et al. Efficacy and safety of low-dose SoluMatrix meloxicam in the treatment of osteoarthritis pain: A 12-week, phase 3 study, *Curr Med Res Opin* 2015;31:2331-43.

25 Atukorala I, Pathmeswaran A, Makovey J, et al. Is there a relationship between the intermittent and constant osteoarthritis pain score (ICOAP) and pain flares in knee osteoarthritis? (abstract) [abstract]. *Osteoarthritis and Cartilage* 2016;24:S429-30.

26 Baer PA, Thomas LM, Shainhouse Z. Treatment of osteoarthritis of the knee with a topical diclofenac solution: a randomised controlled, 6-week trial ISRCTN53366886], *BMC Musculoskeletal Disorders* 2005;6:44.

27 Baraf HSB, Gloth FM, Barthel HR, et al. Safety and Efficacy of Topical Diclofenac Sodium Gel for Knee Osteoarthritis in Elderly and Younger Patients, *Drugs Aging* 2011;28:27-40.

28 Battisti WP, Katz NP, Weaver AL, et al. Pain management in osteoarthritis: A focus on onset of efficacy—a comparison of rofecoxib, celecoxib, acetaminophen, and nabumetone across four clinical trials, *The Journal of Pain* 2004;5:511-20.

29 Bingham CO, Sebba AI, Rubin BR, et al. Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies, *Rheumatology* 2007;46:496-507.

30 Birbara C, Ruoff G, Sheldon E, et al. Efficacy and safety of rofecoxib 12.5 mg and celecoxib 200 mg in two similarly designed osteoarthritis studies, *Curr Med Res Opin* 2006;22:199-210.

31 Bocanegra T, Weaver A, Tindall E, et al. Diclofenac/misoprostol compared with diclofenac in the treatment of osteoarthritis of the knee or hip: a randomized, placebo controlled trial. Arthrotec Osteoarthritis Study Group. *Journal of Rheumatology* 1998;25:1602-11.

32 Boswell DJ, Ostergaard K, Philipson RS, et al. Evaluation of GW406381 for Treatment of Osteoarthritis of the Knee: Two Randomized, Controlled Studies, *The Medscape Journal of Medicine* 2008;10:259.

33 Brandt KD, Mazzuca SA, Buckwalter KA. Acetaminophen, like conventional NSAIDs, may reduce synovitis in osteoarthritic knees, *Rheumatology* 2006;45:1389-94.

34 Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: A randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium, *Arch Intern Med* 2003;163:169-78.

35 Ehrich E, Schnitzer T, McIlwain H, et al. Effect of specific COX-2 inhibition in osteoarthritis of the knee: a 6 week double blind, placebo controlled pilot study of rofecoxib. Rofecoxib Osteoarthritis Pilot Study Group. *Journal of Rheumatology* 1999;26:2438-47.

36 Essex M, O'Connell M, Brown PB. Response to Nonsteroidal Anti-Inflammatory Drugs in African Americans with Osteoarthritis of the Knee, *Journal of International Medical Research* 2012;40:2251-66.

37 Gibofsky A, Hochberg MC, Jaros MJ, et al. Efficacy and safety of low-dose submicron diclofenac for the treatment of osteoarthritis pain: A 12 week, phase 3 study, *Curr Med Res Opin* 2014;30:1883-93.

38 Gineyts E, Mo JA, Ko A, et al. Effects of ibuprofen on molecular markers of cartilage and synovium turnover in patients with knee osteoarthritis, *Annals of the Rheumatic Diseases* 2004;63:857-61.

39 Goldberg M, McIlwain H, Poiley J, et al. Controlled-release naproxen in the treatment of osteoarthritis, *Current Therapeutic Research-Clinical and Experimental* 1988;44:51-60.

40 Gottesdiener K, Schnitzer T, Fisher C, et al. Results of a randomized, dose ranging trial of etoricoxib in patients with osteoarthritis, *Rheumatology* 2002;41:1052-61.

41 Hochberg MC, Fort JG, Svensson O, et al. Fixed-dose combination of enteric-coated naproxen and immediate-release esomeprazole has comparable efficacy to celecoxib for knee osteoarthritis: two randomized trials, *Curr Med Res Opin* 2011;27:1243-53.

42 Katz N, Sun S, Johnson F, et al. ALO-01 (Morphine Sulfate and Naltrexone Hydrochloride) Extended-Release Capsules in the Treatment of Chronic Pain of Osteoarthritis of the Hip or Knee: Pharmacokinetics, Efficacy, and Safety, *The Journal of Pain* 2010;11:303-11.

43 Kivitz AJ, Makarowski WS, Fiechtner JJ, et al. A Flexible Daily Dosage Regimen of Oxaprozin Potassium in Patients with Acute Knee Pain Associated with Osteoarthritis, *Clinical Drug Investigation* 2001;21:745-53.

44 Erfani T, Zhang Y, Makovey J, et al. Intermittent analgesic use and risk of pain exacerbation in knee osteoarthritis: A web based case-crossover study (abstract) [abstract]. *Arthritis and Rheumatology* 2014;66.

45 Leung AT, Malmstrom K, Gallacher AE, et al. Efficacy and Tolerability Profile of Etoricoxib in Patients with Osteoarthritis: A Randomized, Double-blind, Placebo and Active-comparator Controlled 12-Week Efficacy Trial, *Curr Med Res Opin* 2002;18:49-58.

46 Luyten FP, Geusens P, Malaise M, et al. A prospective randomised multicentre study comparing continuous and intermittent treatment with celecoxib in patients with osteoarthritis of the knee or hip, *Annals of the Rheumatic Diseases* 2007;66:99-106.

47 Manicourt D, Bevilacqua M, Righini V, et al. Comparative Effect of Nimesulide and Ibuprofen on the Urinary Levels of Collagen Type II C-Telopeptide Degradation Products and on the Serum Levels of Hyaluronan and Matrix Metalloproteinases-3 and -13 in Patients with Flare-Up of Osteoarthritis, *Drugs in R & D* 2005;6:261-71.

48 Mazzuca S, Brandt K, Lane K, et al. Knee pain reduces joint space width in conventional standing anteroposterior radiographs of osteoarthritic knees, *Arthritis Rheum* 2002;46:1223-7.

49 McIlwain H, Silverfield JC, Cheatum DE, et al. Intra-articular orgotein in osteoarthritis of the knee: A placebo-controlled efficacy, safety, and dosage comparison, *Am J Med* 1989;87:295-300.

50 Mendelsohn S. Clinical efficacy and tolerability of naproxen in osteoarthritis patients using twice-daily and once-daily regimens, *Clinical therapeutics* 1991;13:8-15.

51 Moskowitz RW, Sunshine A, Hooper M, et al. An analgesic model for assessment of acute pain response in osteoarthritis of the knee, *Osteoarthritis and Cartilage* 2006;14:1111-8.

52 Pareek A, Chandurkar N, Sharma VD, et al. A randomized, multicentric, comparative evaluation of aceclofenac-paracetamol combination with aceclofenac alone in Indian patients with osteoarthritis flare-up, *Expert Opin Pharmacother* 2009;10:727-35.

53 Pareek A, Chandurkar N, Ambade R, et al. Efficacy and Safety of Etodolac-Paracetamol Fixed Dose Combination in Patients With Knee Osteoarthritis Flare-up: A Randomized, Double-blind Comparative Evaluation, *Clin J Pain* 2010;26:561-6.

54 Ricci JA, Stewart WF, Chee E, et al. Pain Exacerbation as a Major Source of Lost Productive Time in US Workers With Arthritis, *Arthritis & Rheumatism: Arthritis Care & Research* 2005;53:673-81.

55 Schnitzer TJ, Fricke JR, Gitton X, et al. Lumiracoxib in the treatment of osteoarthritis, rheumatoid arthritis and acute postoperative dental pain: results of three dose-response studies, *Curr Med Res Opin* 2005;21:151-61.

56 Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, et al. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis, *Arthritis & Rheumatism* 2001;44:1599-607.

57 Silverfield JC, Kamin M, Wu S, et al. Tramadol/acetaminophen combination tablets for the treatment of osteoarthritis flare pain: a multicenter, outpatient, randomized, double-blind, placebo-controlled, parallel-group, add-on study, *Clin Ther* 2002;24:282-97 doi:http://dx.doi.org/10.1016/S0149-2918(02)85024-X [published Online First: February 2002].

58 Strand V, Simon LS, Dougados M, et al. Treatment of osteoarthritis with continuous versus intermittent celecoxib, *J Rheumatol* 2011;38:2625-34.

59 Wiesenhutter CW, Boice JA, Ko A, et al. Evaluation of the Comparative Efficacy of Etoricoxib and Ibuprofen for Treatment of Patients With Osteoarthritis: A Randomized, Double-Blind, Placebo-Controlled Trial, *Mayo Clin Proc* 2005;80:470-9.

60 Williams GW, Hubbard RC, Yu SS, et al. Comparison of once-daily and twice-daily administration of celecoxib for the treatment of osteoarthritis of the knee, *Clin Ther*

2001;23:213-27 doi:http://dx.doi.org/10.1016/S0149-2918(01)80004-7 [published Online First: February 2001].

61 Wittenberg RH, Schell E, Krehan G, et al. First-dose analgesic effect of the cyclo-oxygenase-2 selective inhibitor lumiracoxib in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled comparison with celecoxib NCT00267215], *Arthritis Research & Therapy* 2006;8:R35.

62 Yeasted R, McPherson J, Schnitzer T. Characterization of osteoarthritis pain variability (abstract) [abstract]. *Osteoarthritis and Cartilage* 2014;22:S390-1.

63 Young C, Parenti D, Hochberg M. Lower-dose diclofenac capsules developed using solumatrix fine particle technology result in clinically meaningful improvements in pain in a phase 3 study of patients with osteoarthritis (abstract) [abstract]. *Osteoarthritis and Cartilage* 2014;22.

64 Zhao SZ, McMillen JI, Markenson JA, et al. Evaluation of the Functional Status Aspects of Health-Related Quality of Life of Patients with Osteoarthritis Treated with Celecoxib, *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 1999;19:1269-78.

65 Zobel I, Erfani T, Bennell K, et al. Relationship of buckling and knee injury to pain exacerbation in knee osteoarthritis: A web-based case-crossover stud, *Interact J Med Res* 2014;66:S560-1.

66 Conrozier T, Mathieu P, Vignon E, et al. Differences in the osteoarthritic synovial fluid composition and rheology between patients with or without flare: a pilot study. *Clinical and experimental rheumatology* 2012;30:729-34.

67 D'Agostino MA, Conaghan P, Le Bars M, et al. EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 1: Prevalence of inflammation in osteoarthritis, *Annals of the Rheumatic Diseases* 2005;64:1703-9.

68 Jawad ASM. Analgesics and osteoarthritis: are treatment guidelines reflected in clinical practice? *Am J Ther* 2005;12:98-104.

69 Murphy SL, Lyden AK, Kratz AL, et al. Characterizing pain flares from the perspective of individuals with symptomatic knee osteoarthritis, *Arthritis Care and Research* 2015;67:1103-11.

70 Wise BL, Niu J, Zhang Y, et al. Psychological factors and their relation to osteoarthritis pain, *Osteoarthritis and Cartilage* 2010;18:883-7.

71 Zhang Y, Zhang B, Wise B, et al. Statistical approaches to evaluating the effect of risk factors on the pain of knee osteoarthritis in longitudinal studies, *Curr Opin Rheumatol* 2009;21:513-9.

72 Zhang Y, Wheaton D, N, J., et al. Recent heavy physical activities trigger knee pain exacerbation in persons with symptomatic knee osteoarthritis (abstract) [abstract]. *Arthritis & Rheumatism* 2011;63(10).

73 Day R, Morrison B, Luza A, et al. A randomized trial of the efficacy and tolerability of the cox-2 inhibitor rofecoxib vs ibuprofen in patients with osteoarthritis, *Arch Intern Med* 2000;160:1781-7.

74 Kivitz AJ, Greenwald MW, Cohen SB, et al. Efficacy and Safety of Rofecoxib 12.5 mg Versus Nabumetone 1,000 mg in Patients with Osteoarthritis of the Knee: A Randomized Controlled Trial, *J Am Geriatr Soc* 2004;52:666-74.

75 Bingham CO, Smugar SS, Wang H, et al. Predictors of Response to Cyclo-Oxygenase-2 Inhibitors in Osteoarthritis: Pooled Results from Two Identical Trials Comparing Etoricoxib, Celecoxib, and Placebo, *Pain Medicine* 2011;12:352-61.

76 Essex MN, Behar R, O'Connell MA, et al. Efficacy and tolerability of celecoxib and naproxen vs placebo in hispanic patients with knee osteoarthritis, *Osteoarthritis and Cartilage* 2013;21.

77 Yocum D, Fleischmann R, Dalgin P, et al. Safety and efficacy of meloxicam in the treatment of osteoarthritis: A 12-week, double-blind, multiple-dose, placebo-controlled trial, *Arch Intern Med* 2000;160:2947-54.

78 Atukorala I, Pathmeswaran A, Chang T, et al. Do Traditional Risk Factors for Knee Osteoarthritis Predict Pain Flares in Knee Osteoarthritis? *Ann Rheum Dis* 2016;75:835.

79 Bartholdy C, Klokker L, Bandak E, et al. A Standardized "Rescue" Exercise Program for Symptomatic Flare-up of Knee Osteoarthritis: Description and Safety Considerations, *J Orthop Sports Phys Ther* 2016;46:942-6.

80 Bassiouni H. Detection of changes in the serum and synovial fluid levels of resistin during flare ups and remissions in primary knee osteoarthritis, *Arthritis and Rheumatology* 2015;67.

81 Erfani T, Makovey J, Bennell K, et al. Psychosocial Factors and Pain Exacerbation in Knee Osteoarthritis: a Web Based Case-Crossover Study, *Intern Med J* 2014;44:16-.

82 Ferreira ML, Zhang Y, Metcalf B, et al. The influence of weather on the risk of pain exacerbation in patients with knee osteoarthritis - a case-crossover study, *Osteoarthritis and cartilage* 2016;24:2042-7.

83 Hunter DJ, Bennell K, Makovey J, et al. Psychosocial Factors and Pain Exacerbation in Knee Osteoarthiritis: a Web Based Case-Crossover Study, *Osteoarthritis and Cartilage* 2014;22:S21-2.

84 Makovey J, Metcalf B, Zhang Y, et al. Web-Based Study of Risk Factors for Pain Exacerbation in Osteoarthritis of the Knee (SPARK-Web): Design and Rationale, *JMIR research protocols* 2015;4.

85 Parry E, Ogollah R, Peat G. Significant pain variability in persons with, or at high risk of, knee osteoarthritis: preliminary investigation based on secondary analysis of cohort data, *BMC musculoskeletal disorders* 2017;18:80.

86 Cibere J, Kopec JA, Thorne A, et al. Randomized, double-blind, placebo-controlled glucosamine discontinuation trial in knee osteoarthritis, *Arthritis Care and Research* 2004;51:738-45.

87 Cibere J, Kopec JA, Esdaile JM, et al. Glucosamine sulfate and cartilage type II collagen degradation in patients with knee osteoarthritis: randomized discontinuation trial results employing biomarkers. In: Anonymous . Journal of rheumatology 2005:896-902.

88 Roth SH, Shainhouse JZ. 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, *Arch Intern Med* 2004;164:2017-23.

89 Simon LS, Grierson LM, Naseer Z, et al. 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* 2009;143:238-45.

90 Weaver A, Rubin B, Caldwell J, et al. Comparison of the efficacy and safety of oxaprozin and nabumetone in the treatment of patients with osteoarthritis of the knee, *Clin Ther* 1995;17:735-45.

91 Rother M, Lavins BJ, Kneer W, et al. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral celecoxib and placebo in osteoarthritis of the knee: multicentre randomised controlled trial, *Annals of the Rheumatic Diseases* 2007;66:1178-83.

92 Zobel I, Erfani T, Bennell KL, et al. Relationship of Buckling and Knee Injury to Pain Exacerbation in Knee Osteoarthritis: A Web-Based Case-Crossover Study, *Interactive journal of medical research* 2016;5:e17.

93 Woolacott NF, Corbett MS, Rice SJC. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials, *Rheumatology* 2012;51:1440-6.

94 National Institute for Health and Care Excellence (NICE). Osteoarthritis: care and management (CG177). London: NICE 2014.

95 Buttgereit F, Burmester G, Bijlsma JWJ. Non-surgical management of knee osteoarthritis: where are we now and where do we need to go? *RMD Open* 2015;1.

96 Porcheret M, Healey E, Dziedzic K, et al. Ostoearthritis: a modern approach to diagnosis and management, *Arthritis Research UK* 2011;Series 6.

97 Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee, *J Rheumatol* 1990;17:364-72.

98 Allen KD, Coffman CJ, Golightly YM, et al. Daily pain variations among patients with hand, hip, and knee osteoarthritis, *Osteoarthritis and Cartilage* 2009;17:1275-82.

99 Makris D, Bouros D. COPD exacerbation: Lost in translation, *BMC Pulmonary Medicine* 2009;9:6.

100 Fitzgerald JD, Grossman JM. Validity and reliability of retrospective assessment of disease activity and flare in observational cohorts of lupus patients, *Lupus* 1999;8:638-44.

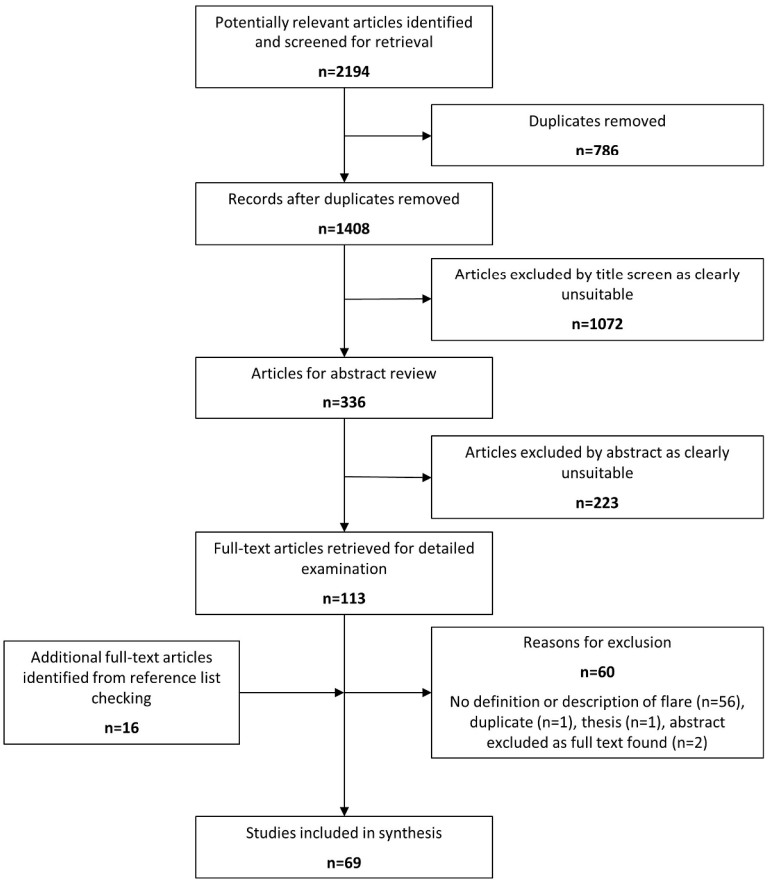
101 Lewis JD, Aberra FN, Lichtenstein GR, et al. Seasonal variation in flares of inflammatory bowel disease, *Gastroenterology*;126:665-73.

102 Suri P, Saunders KW, Von Korff M. Prevalence and Characteristics of Flare-ups of Chronic Nonspecific Back Pain in Primary Care: A Telephone Survey, *Clin J Pain* 2012;28:573-80.

103 Cross M, Dubouis L, Mangin M, et al. Defining Flare in Osteoarthritis of the Hip and Knee: A Systematic Literature Review- OMERACT Virtual Special Interest Group, *J Rheumatol* 2017;44(12):1920-7.

104 Sands GH, Brown PB, Essex MN. The Efficacy of Continuous Versus Intermittent Celecoxib Treatment in Osteoarthritis Patients with Body Mass Index â‰¥30 and *The Open Rheumatology Journal* 2013;7:32-7.

105 Berthelot J, De Bandt M, Morel J, et al. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: The 'FLARE' instrument, *Annals of the Rheumatic Diseases* 2012;71:1110-6.



PRISMA Flowchart

190x275mm (300 x 300 DPI)

Online supplement: Example search strategy

Table 1: Key terms and MeSH headings used for EMBASE database search. The concepts were combined as follows: "KNEE JOINT" AND "ACUTE EVENTS"

Concepts	Search terms
KNEE JOINT	"knee adj3 (pain OR painful)" or "Knee osteoarthritis" or "knee adj3 (arthrosis)" or "knee adj3 (joint OR joints OR degenerative)" or "knee adj3 (osteoarthritis)"
ACUTE EVENTS	"exacerbation" or "flare" or "daily adj3 (pain)" or "pain AND (diary OR diaries)" or "pain adj3 (variab$)"

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6-7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Online supplement
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	n/a
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	9

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n/a
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10 (flowchart)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10, Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A but rationale on 18
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	n/a
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	n/a
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	19-35
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	36-37
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	39
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	40

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

PRISMA 2009 Checklist