Achieving consensus on pain outcome measures in systematic reviews of chronic musculoskeletal conditions : the current state of reporting and identifying key topics for consideration

Maxwell, Lara Jane

UNIVERSITY OF SPLIT

SCHOOL OF MEDICINE

Achieving consensus on pain outcome measures in systematic reviews of chronic musculoskeletal conditions: the current state of reporting and identifying key topics for consideration

Doctoral Thesis

Lara Jane Maxwell

April 21, 2016

Mentor: Professor Peter Tugwell MSc, MD, FRCPC

Acknowledgements

My co-mentors: Professors Peter Tugwell and George Wells:

> Peter, thank you for your unending support, encouragement, and consistently wise advice as I pursued this goal. You have been a tremendous mentor and I very much appreciate how you always made this thesis work a priority for you. I will never forget your "PP" principle!
>
> George, thank you for your valuable guidance, feedback, and always enjoyable and indispensable discussions on both statistical and non-statistical issues. You have taught me so much over these past 12 years, that I simply cannot imagine completing this without your tutelage.

My co-authors: thank you for your constructive and worthwhile discussions and feedback which much improved the outcomes of these projects.

Survey and interview participants: thank you for generously giving your time for this project. Your expertise provided me with valuable insight and understanding into the issues.

My many friends, both old and new: your words of encouragement meant so much to me and kept me inching forward. Special thanks to Andre, Jil, Jennifer and Livia for sharing the journey and going above and beyond for me!

My family: Mum, Dad, Fifi and Peter: thank you for your love, constant support, and for cheering me on every step of the way. I promise that I am finally done with school!

My darling girls, Isla and Keira: one of the hardest things on this journey was the time that it took away from you. I love you both so much and hope that I have shown you that you can achieve anything you put your mind to.

And to my Kevin: without your support and unwavering confidence in me, this would not have been possible. Thank you for inspiring me to have no regrets and to be happy, and for your interminable patience. I love you.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The Comparative Effectiveness Research initiative and regulatory agencies have challenged the Outcome Measures in Rheumatology (OMERACT) initiative, and similar organizations, to establish truthful, discriminative and feasible patient-important outcomes for randomized controlled trials (RCTs) and non-randomized studies (1). As the founders of OMERACT elegantly wrote in 1993, "Clinical trials are only as credible as their endpoints" (2). If we are to believe the results of a clinical trial designed to investigate benefits and harms of a treatment, the trial outcomes measures must be ones that are important and relevant to patients, health care practitioners, policy-makers and other decision makers, and have demonstrated acceptable measurement properties. In addition, the same important outcomes should be measured in all clinical trials of the same conditions and class of interventions to allow for pooling of results across trials, thereby improving the wider evidence base and reducing research waste.

1.1. Cochrane Collaboration

The Cochrane Collaboration is a non-profit, international organization that aims to help people make well-informed healthcare decisions by producing, maintaining, and disseminating systematic reviews on the evidence on benefits and harms of a healthcare intervention. Cochrane systematic reviews are internationally recognized for providing unbiased, methodologically-sound evidence on the benefits and harms of health-care interventions (3). The Cochrane Collaboration is perhaps the largest organization that uses outcomes as the basis to synthesize results of interventional trials to provide high quality, independent evidence to patients, health-care providers and other decision makers. Over 6000 systematic reviews are available electronically in the *Cochrane Database of Systematic Reviews*. Evidence from Cochrane Reviews has been used by patients and health care practitioners to inform decision-making, in healthcare textbooks, and by national and international agencies and clinical guidelines groups, thus demonstrating the wide ranging impact of these reviews (4). A mixed methods assessment of the impact of 1502 Cochrane Reviews produced by Cochrane Review Groups funded by the UK National Institutes for Health Research between 2007 and 2011, found that primary research was influenced by 40

reviews that had identified gaps in the evidence base and that 483 reviews had been used to inform "62 sets of international guidance, 175 sets of national guidance (87 from the UK) and 10 examples of local guidance". As well, the evidence from Cochrane Reviews led to benefits in the UK health care system such as providing patients with more treatment options through the introduction of new effective health care treatments and more appropriate treatment use (5).

1.2.Summary of Findings Tables

Cochrane Reviews are often lengthy documents filled with technical details about the methodology employed during the process of conducting the review. While this detail is necessary to ensure a high quality product, it may be difficult for readers of the review (patients, healthcare practitioners, policymakers) to extract the key information and use it in their decision making. Therefore, various summary formats of the review have been developed, including firstly, a scientific structured abstract and secondly, a plain language summary (6). A third summary format is an integral feature of Cochrane systematic reviews and is entitled a 'Summary of Findings' table. It is a succinct and transparent summary of the key results for each major outcome to enhance interpretability and improve the integration of results into the decision-making process. It provides the reader with an estimate of baseline risk for the population of under study, the corresponding estimate of effect in the treatment group, the number of studies and participants contributing to the estimate, and the quality of evidence of for each major outcome.

There is evidence that Summary of Findings tables improve understanding about the results and quality of the evidence of the main outcomes and help readers find the key information in a systematic review more quickly. In two small randomized controlled trials (n=47 and n=25), researchers found that participants who were given a Summary of Findings table were more likely to respond that they found it easy to find results for important outcomes compared to participants without the Summary of Findings table (68% vs. 40%, (P = 0.021)). As well, the participants with the Summary of Findings table more often provided correct answers about the interpretation of the results (question regarding the control group risk: 93% vs.44% (P = 0.003); question regarding intervention group risk: 87% vs. 11% (P < 0.001)) and spent less time finding information on the main outcomes; an average of 90 seconds compared to 4 minutes (P = 0.002) (7).

More recently, a randomized controlled trial compared several items in a newly designed Summary of Findings table format to the existing format currently used in Cochrane Reviews. The investigators found that the new format improved readers' comprehension of the content, accessibility and ability to quickly find key information, and that readers' preferred and were more satisfied with the new format. The new format included a column entitled 'What happens' in which a brief narrative explanation of the interpretation of the result for each major outcome was provided by describing the direction of the effect, the magnitude of the effect size and the quality of the evidence. For example, in a systematic review on the use of probiotics for the prevention of pediatric antibiotic-associated diarrhea the following was written in the 'What happens' column: "Probably decreases the incidence of diarrhea". As well, the new format allowed for different presentations of risk to help improve interpretability; for example, the number needed to treat (NNT; the number of people that have to be treated in order to achieve the outcome of interest in one person) and the risk/rate difference in percentages (the absolute difference between the risk in the treatment group and the risk in the control group). Trial participants using the new format had a statistically significant higher number of correct answers in questions which were designed to test different parts of their understanding (e.g., ability to determine risk difference, ability to understand quality of evidence) (8).

1.2.1. GRADE

The GRADE (**G**rading of **R**ecommendations **A**ssessment, **D**evelopment and **E**valuation Working Group) criteria are used to assess the quality of the evidence and provide the reader with a "degree of confidence" for the results for each major outcome. The GRADE criteria consist of five individual factors: 1. Limitations in study design or execution (risk of bias), 2. Inconsistency, 3. Imprecision, 4. Indirectness, and 5. Publication bias (9-11). 'Limitations in study design or execution (risk of bias)' refers to issues with the internal validity of the trials contributing information to an outcome and the Cochrane Risk of Bias tool (12) is used to assess this GRADE factor. The Risk of Bias tool for use with randomized controlled trials evaluates the following aspects that may affect the internal validity of the trial and result in concerns about biased results: generation of the randomization sequence, allocation concealment, blinding of patients and personnel, blinding of outcome assessor, selective outcome reporting, and incomplete outcome data. 'Inconsistency' assesses the amount of

heterogeneity in a meta-analysis and whether there is large, unexplained heterogeneity so that it is difficult to interpret the meaning of the overall estimate. 'Imprecision' addresses whether the 95% confidence interval of the estimate of effect is narrow or wide. An estimate may be considered imprecise if the ends of the confidence interval encompass both appreciable harm or benefit and the null effect, resulting in uncertainty about the true estimate of effect. 'Indirectness' assesses the whether the evidence fully addresses the 'PICO' [Population, Intervention, Comparator, Outcome] elements of the systematic review question. Within the aspect of outcomes, a concern may arise when a 'surrogate' outcome is used; that is, an outcome that provides indirect evidence in that it approximates a patient-important outcome. A common example from rheumatology is bone mineral density, which is a frequent outcome in trials on osteoporosis, and is used as a surrogate marker for fracture data. Fractures are, of course, of primary interest and are more meaningful to a decision-maker, but are fairly rare events. Thus, trials with large sample sizes and of long duration are necessary in order to have enough fracture events so that the trial is not underpowered to find an effect. But because it is expensive to conduct large trials of long duration, trialists use a surrogate outcome like bone mineral density in their trials. However, the relationship between bone mineral density and fractures has been investigated using logistic regression in clinical trials of antiresorptive therapies for osteoporosis and found that increases in bone mineral density does not fully explain resulting decreases in fractures (13-15). Thus, there may be concerns about the applicability of evidence from clinical trials which use surrogate outcomes to answer the systematic review question. The final GRADE factor is 'publication bias' which is a form of reporting bias in which the publication of journal articles describing results of clinical trials is more likely if the intervention demonstrates a large and statistically significant effect.

Using the GRADE criteria, the quality of evidence can be graded as high, moderate, low, or very low. The highest quality rating is for a body of evidence based on data from randomized trials without important limitations. The evidence rating may then be downgraded one level for limitations in one of the five factors described above; it may be downgraded two levels in the case of serious limitations. If the source of the evidence is observational studies, the quality of the evidence is rated as low and it can be upgraded due to the following factors: a large, consistent, and precise magnitude of effect (for example, as seen with total hip joint replacement); all plausible confounders would be expected to reduce the observed effect (thereby increasing confidence that there is indeed an effect) or in the case where

observational studies did not observe an effect, the confounders would be expected to increase an effect; and evidence of a dose-response gradient (9).

The meaning of each rating is as follows: high quality: we are very confident that the true effect lies close to that of the estimate of the effect; moderate quality: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different; low quality: confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect ; very low quality: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect (16).

1.3. Patient-reported outcomes

A patient-reported outcome is defined as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" (17). Patient-reported outcomes are important as they serve to characterize the effect of a health condition from the patient perspective. Over the past decades, there have been increased efforts to involve patients more actively in their healthcare decision-making. For example, in 2010, the US Congress authorized the establishment of the Patient-Centered Outcomes Research Institute (PCORI), an independent nonprofit, non-governmental organization whose aim is to provide high-quality research evidence that is guided by the "evaluation of questions and outcomes meaningful and important to patients and caregivers" (18).

Patient-reported outcomes can be obtained by directly asking a patient or having a patient fill in a self-completed checklist, diary, or other type of form. In the context of a study designed to assess the effectiveness of an intervention, patient-reported outcomes can inform whether a treatment has improved those qualities about their health that patients consider most important. These outcomes are in addition to more objective outcomes such as mortality, major morbid events, and laboratory and physiological measures (e.g. biomarkers, outcomes based on physical examination) that are often measured in clinical trials (19).

Different patient-reported outcome measures can be used to measure different 'domains', also known as 'concepts' or 'constructs', and which can be defined as the "what to measure" (20) in a clinical trial. These can include signs or symptoms, or group of symptoms; the effects on a particular functional ability or group of abilities; a measure of feeling about overall health

and well-being; and opinions on satisfaction or acceptability of treatment. Outcome measures covering these aspects can be focused on disease-specific, condition-specific or on 'generic' features; that is, those applicable to populations regardless of the disease or condition (21).

The Cochrane Patient-Reported Outcomes Methods Group has a chapter in the Cochrane Handbook which discusses key issues that systematic review authors should consider when reporting on patient-reported outcomes (21). A checklist for authors on important considerations when describing and assessing patient-reported outcomes in systematic reviews has been developed. The checklist includes considerations such as rationale for the construct, evidence for reliability, validity, and responsiveness of the measurement instruments, and interpretability of the result.

The OMERACT initiative has pioneered methodological work to develop a framework for evaluating outcome measures for use in clinical trials. OMERACT's focus is on interventions for rheumatologic conditions and one of the key features of OMERACT is that they have intensively involved patients in their methods work to ensure the patient perspective is captured when identifying important outcome domains. As one example, this integration of 'patient researcher partners' led to the identification of fatigue being an important outcome to patients with rheumatoid arthritis and a recommendation that fatigue should be measured in clinical trials on interventions for rheumatoid arthritis, in addition to the outcomes previously recommended by OMERACT (22, 23).

1.3.1. Measuring pain outcomes

The International Association for the Study of Pain (IASP) defines pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (24). The global burden of chronic pain has been highlighted by IASP for their "Global Year Against Pain" (2004-2005) campaign. The campaign called for the recognition of the 'right to pain relief' as a human right. Evidence presented in support of this proposal was based on large epidemiological studies conducted across various world regions which estimated that chronic pain affects one in five adults and household surveys in Europe and the United States showed that over one-third have a chronic pain sufferer (25). More recent evidence demonstrated that in a cohort of 6940 people from Scotland, ten years later, those with severe chronic pain had an increased risk of mortality (hazard ratio 1.49, 99% CI 1.21 to 1.84), after adjusting for socio-demographic factors (26).

Pain is a subjective and complex experience which complicates efforts to quantify it. It is often the reason that patients seek professional care. Yet in order to effectively manage it, and identify interventions which attempt to reduce it, an accurate assessment is necessary. In clinical trials of interventions to alleviate suffering from chronic painful conditions, measurement of mortality or major morbidity (which can be easier to measure) is necessary, but evidence of reduction in key patient-reported outcomes like pain and disability is of equal or greater importance to patients (27).

The effect of treatment on the outcome of pain does not always correlate well with other patient-important outcome domains so it is necessary to provide a measure of pain itself. For example, in a systematic review of opioids versus placebo for chronic non-cancer pain, the effect size was twice as large for pain relief (standardized mean difference (SMD) = –0.60; 95%
confidence interval (CI) = –0.69 to –0.50) versus improvement in function (0.31; 95% CI = 0.41 to 0.22) (28).

Pain is a decidedly personal, multidimensional experience and various factors including beliefs, cultural background, cognitive ability, past and current experiences, personality, emotional and social support and networks can influence a patient's report of pain (27, 29). Given its highly subjective nature, researchers have investigated whether more 'objective' outcomes can substitute for a patient-reported outcome of pain. Investigation of the correlation of patient-reported pain with physical measurement outcomes such as range of motion and muscle strength, and imaging and laboratory outcomes has not revealed strong associations between the two (29). For example, in knee osteoarthritis, evidence of radiographic changes was not highly associated with self-reported pain (30). Therefore, a self-report outcome of pain is necessary when assessing the benefits and harms of an intervention.

As described by Jensen and Karoly (31), pain intensity, affect, quality and location are four dimensions that can be assessed in most pain patients. Pain intensity is defined as the severity of pain, or "how much it hurts". Patients are usually asked to relate their intensity in a quantitative measure, for example, on a 0 to 10 scale. Pain affect is the emotional stimulation caused by the sensation of pain. Chronic pain suffers may feel anxiety or fear due to their condition and this can impact their usual activities; this is a more complex and

difficult outcome to measure than pain intensity. Pain quality refers to the description of the sensations of pain (e.g. burning, throbbing, stabbing) and can be described using a variety of different adjectives organized into categories. Pain location refers to the position of the pain sensation on the body and can be assessed by pain drawings.

Pain intensity is often measured using a simple, unidimensional scale. Although intensity captures only one aspect of the complex pain experience, it is recommended as a core outcome domain in clinical trials and is considered a sensitive outcome measure (32). Common single-item pain intensity instruments include the visual analogue scale, numerical rating scale, and verbal rating. The visual analogue scale (VAS) is a 0 to 10cm (equivalent to a 0 to 100mm) line on which the respondent is asked to put an 'x' on the line to rate their pain intensity. The numerical rating scale (NRS) can be administered either verbally or graphically and common response options are 0 to 10, 0 to 20, and 0 to 100. In the graphic version, the numbers are often put in boxes on a straight line. The scale anchors of the VAS and NRS can vary in the wording but are usually similar to '0 = no pain; 10 = worst pain imaginable'. The verbal rating scale (VRS) provides patients with a descriptive list, increasing in intensity; for example, no pain, mild pain, moderate pain, severe pain. The number of VRS response options can vary, but is usually between 4 to 7 descriptors (33).

For all these outcome instruments, the question asked of the respondent around the timing of the assessment (e.g. current, last week, last month), the type of the pain (e.g. least, worst, average), and any specified location of pain (e.g. overall, knee pain) should be explicitly reported so as to facilitate appropriate comparisons between clinical trials. A recent systematic review (conducted after the work in this thesis) by the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION) initiative investigated the reporting of these features in 262 recent journal articles of pain studies. They found that the description of the pain intensity outcome was not explicitly reported in 24% of the studies and over a third of all the individual elements were not reported, including 43% not clearly reporting the type of pain assessed. This ACTTION article provided recommendations for trial authors on reporting pain intensity assessments (34).

Various issues have been raised around the use of a unidimensional scale to capture the domain of pain intensity. The ability of patients to quantify and reduce a qualitative

experience which can be influenced by many variables to a simple number has been questioned. Another issue often raised is whether a patient is able to validly report a previous level or a change in the level of pain; this is known as 'recall bias'. There has been research demonstrating that a retrospective report of pain intensity may be over- or under-estimated and influenced by present levels of pain (29). Other evidence demonstrated that a 3-month recall period of average pain intensity, interference, and days with pain was adequately valid (27). The evidence for the validity of VAS, NRS, and VRS scales in different chronic pain conditions was thoroughly explored in the chapter by Jensen in the 'Handbook of Pain Assessment', edited by Turk and Melzack (31). The authors recommended the use of the NRS as patients make fewer errors while using it compared to the VAS and it is more sensitive than the VRS. However, if researchers want to undertake calculations that require the use of a scale with ratio properties, then the VAS was recommended (35).

1.4.Challenges for systematic review authors

Up to seven outcomes may be included in the Summary of Findings table; they should be those that are deemed most important from a patient perspective and represent both benefit and harm. While the guidance provided in the Cochrane Handbook states that outcomes important to both patients and other decision makers should be included, there is no explicit process for deciding which outcomes, which outcome instruments, and which effect size metric should be reported in the Summary of Findings table (36).

To answer a question about the benefit and harm of a treatment, systematic review authors must rely on the outcomes measured in the clinical trials that meet a review's inclusion criteria. The outcome instruments should be valid, reliable, and responsive, but the evidence for the adequacy of the measurement properties of the outcome instruments may not be readily available to systematic review authors.

A major problem faced by systematic review authors is the lack of standardization of outcomes measured in clinical trials of the same intervention (37). A key objective of undertaking a meta-analysis is to increase statistical power by combining trials and thereby increasing the sample size. However, the use of different outcomes makes it difficult for systematic review authors to pool together trial results in a meta-analysis. This can be an issue at the level of both the outcome domain (concept) and outcome instrument; i.e. when

trials measure different domains (e.g. pain intensity, pain interference) or when they measure the same domain, such as pain intensity, but do so using different outcome instruments. For example, pain intensity may be measured using a single-item instrument (e.g. visual analogue scale, numerical rating scale) or as part of a multidimensional pain scale (e.g. Western-Ontario McMaster (WOMAC) pain subscale). In a systematic review of interventions for fibromyalgia, review authors found that pain was measured using 75 different outcomes measures in 241 included studies (38). Pain may also be measured as part of an instrument that has been designed to capture multiple different domains, e.g., pain intensity as well as pain interference. Systematic review authors face the difficult task of deciding when it is appropriate to pool different outcomes together in a single meta-analysis.

1.4.1. Measurement properties of health outcome instruments

For each outcome domain that is identified as important to measure in a clinical trial, an outcome instrument capable of appropriately measuring the outcome is needed. The use of an outcome instrument lacking in ability to accurately capture the concept of interest or deficient in its ability to measure a change in the status of an outcome will result in an inaccurate assessment of the effectiveness of an intervention. The measurement properties (or 'psychometric properties') of an instrument can be demonstrated by studies designed to assess those properties.

OMERACT developed a framework, entitled the 'OMERACT Filter', to evaluate the measurement properties of outcome instruments. The Filter consists of three elements, 'Truth, Discrimination, and Feasibility'. The 'truth' element questions whether the outcome instrument measures what it intended to measure and addresses issues of face, content, and construct validity. 'Discrimination' refers to the ability of the instrument to differentiate between situations that change over time (sensitivity to change) and stable situations, (the reliability of the instrument). The 'feasibility' aspect of the filter examines whether issues such as cost, time to complete, and ability to access/obtain or interpret the instrument are limitations to its use (39). OMERACT has recently undertaken a project, 'OMERACT Filter 2.0', to further clarify and elaborate on this original filter, and provide more explicit details about the process of developing a core outcome set with valid outcome measurement instruments (20). Further information about OMERACT is provided in section 1.5.

The terminology and definitions used when discussing measurement properties is not consistent across the body of literature on this topic and this can result in confusion and misunderstanding. In 2010, the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative undertook an international Delphi exercise to obtain consensus on the taxonomy and definitions of different measurement properties for health-related patient-reported outcomes (40). This project resulted in consensus on three quality domains - validity, reliability, and responsiveness – with specific measurement properties included in each as described in Figure 1 below.

Figure 1: COSMIN taxonomy of relationships of measurement properties

Figure 1 reprinted from Journal of Clinical Epidemiology, Vol.63, Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW, The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes, pg. 737-45, Copyright 2010, with permission from Elsevier.

Often the evidence base for the adequacy of the measurement properties of outcome instruments used in clinical trials is not readily evident and it can be a challenge for systematic review authors to determine whether an instrument used in clinical trial has adequate measurement properties. The COSMIN initiative has developed guidance for

conducting systematic reviews of measurement properties of measurement instruments, including a checklist (the COSMIN checklist) for assessing the methodological quality of the individual studies investigating the measurement properties of an outcome instrument (41). The COSMIN website (www.cosmin.nl) hosts a repository of systematic reviews of measurement properties which could assist systematic review authors in obtaining evidence on the adequacy of the instruments included in their review.

1.4.2. Selective outcome reporting

There is direct empirical evidence from meta-epidemiological studies that full reports of randomized controlled trials are more likely to be published if they report positive or statistically significant results (42, 43). As well, those individual outcomes measured in randomized controlled trials that show a statistically significant effect are more likely to be published, though the direction of the effect may vary over time (44, 45). Clinical trials may measure the same or similar outcome domain using different outcome instruments and then selectively report those that show a statistically significant and/or larger effect size.

Both study publication bias and selective outcome reporting bias may impact on the validity of study results, likely overestimating the treatment effect, leading to a decrease in the credibility of the results and making them untrustworthy for decision making. This can be compounded if biased studies are pooled together in a meta-analysis. In a systematic review of empirical cohort studies of randomized controlled trials that investigated evidence for publication bias or selective outcome reporting, the odds of a statistically significant outcome being fully reported compared to non-significant outcomes ranged from an odds ratio (OR) of 2.2 to 4.7 (based on three studies) (44). Another process that contributes to selective outcome reporting is when the primary outcome(s) of a clinical trial is changed after the trial protocol is complete. The frequency of the occurrence of changes in primary outcome(s) can be assessed by comparing trial protocols to final publications. In the Dwan et al systematic review, researchers found that, "40–62% of studies had at least one primary outcome that was changed, introduced, or omitted" (44).

The prevalence and potential impact of selective outcome reporting was assessed in the included trials in a group of Cochrane Reviews (n=283) (45). Over half (55%) of the Cochrane Reviews did not provide data from all eligible trials for the primary outcome of the review and the median amount of missing review outcome data was 10%, though in 70 reviews more than 50% of the data were missing. There was strong evidence that in 6%

(n=155) of the included trials, the trial investigators had measured the outcome but it was not reported in full in the published trial report and 96 Cochrane Reviews included a trial with a high risk of selective outcome reporting bias. In terms of the potential impact, a sensitivity analysis showed that the effect of treatment was decreased by 20% or more in 19/81 Cochrane Reviews when adjusted for outcome reporting bias and the primary outcome in 8/42 reviews changed from statistically significant to non-statistically significant. The authors concluded, "Outcome reporting bias is an under-recognized problem that affects the conclusions in a substantial proportion of Cochrane Reviews."

Concerns about selective outcome reporting bias are not specific to clinical trials; this bias can also occur at the level of the systematic review. Cochrane systematic review authors must pre-specify the primary and secondary outcomes at the protocol stage of conducting a systematic review and all protocols are published in the Cochrane Library. The rationale for this pre-specification of outcomes is to avoid authors omitting or changing outcomes after they are aware of the review's results, and in doing so, especially if changed based on the magnitude of effect size, could introducing a bias into the systematic review process. Kirkham et al investigated changes made to specified primary outcomes during the systematic review process in 288 Cochrane Reviews (46). While the majority of reviews did show evidence of agreement in the primary outcomes reported in the protocol with those in the completed review, they found that out of 28 reviews which reported different primary outcomes from those stated in the protocol, eight (29%) did not explicitly explain this change in outcomes and the authors judged these reviews might have been at a risk of selective outcome reporting bias. As well, in those cases where a secondary outcome in the protocol was moved up to a primary outcome or if a new primary outcome was added at the full review stage, there was an increased risk of obtaining a statistically significant result in the meta-analysis (relative risk 1.66 95% CI (1.10 to 2.49) (46).

One strategy to reduce publication bias and selective outcome reporting bias is to register all clinical trials in an easily accessible database (e.g. ClinicalTrials.gov from the U.S. National Institutes of Health; WHO International Clinical Trials Registry Platform) prior to the start of the trial and to a priori list all the outcomes that will be collected during the trial (47). This will enable systematic review authors to more easily check if full trial results have not been published and if trialists have selectively chosen to report certain outcomes in published reports.

The establishment of a set of outcomes recommended to be measured in all clinical trials for a specified condition, known as a "core outcome set", been proposed as a tool to reduce both the variability of outcome measures and selective reporting bias (37, 48). With the sharp increase in the number of systematic reviews produced over the last fifteen years, the lack of standardized outcome measures in clinical trials hampers efforts to synthesize and interpret the evidence base. Some clinical areas show great variability in the outcomes measured in intervention trials; for example, a systematic review on psychosocial interventions for treating pre-menstrual syndrome found 25 different patient-reported outcomes in nine included studies (21). Williamson et al note that, "…the most accessed and the top cited Cochrane Reviews in 2009 all describe problems due to inconsistencies in the outcomes reported in trials" (37).

In a survey of co-ordinating editors of Cochrane Review Groups, 73% (33/45) of them thought that using a core outcome set established for clinical trials to inform the outcomes for a Summary of Findings table is useful and will probably improve the ability to conduct meta-analysis. As well, the outcome is likely to be an appropriate one that is important for decision makers and may reduce outcome reporting bias. The editors identified some challenges as to the consistent use of a core outcome set: core outcome sets are resource-intensive to create, it may be difficult to convince trialists to employ them, and the most appropriate outcomes may depend on the expected mechanism of action of the intervention (e.g. pharmacological compared to non-pharmacological treatments) (49).

OMERACT has developed a framework, 'OMERACT Filter 2.0', for developing core outcome sets. The first step in the development is to get agreement on what core domains (e.g. pain, function) should be measured in all clinical trials for a certain health condition. This framework shown in Figure 2 recommends that core outcome domain sets encompass concepts of both the impact of the health condition and pathophysiological manifestations when assessing the effect of an intervention.

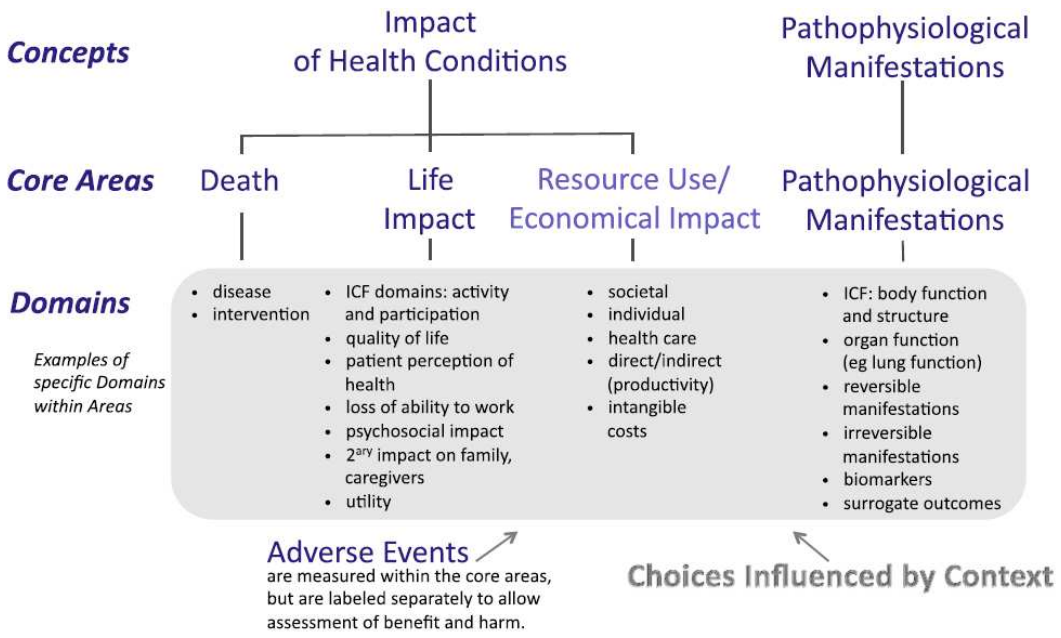


Fig. 1. Conceptual framework of Core Areas for outcome measurement in the setting of health intervention studies. Resource Use has a lighter shade to indicate it is currently strongly recommended, but not mandatory for inclusion. The choice of specific Domains within an Area depends on the context for which the core set is being developed in all areas, domains can be generic or made more specific, for example disease-specific, time-specific (eg, short or long-term), specific for patient preference, and so forth. ICF, International Classification of Functioning, Disability and Health.

Figure 2: OMERACT Filter 2.0 conceptual framework of Core Areas for outcome measurement

The second step is to evaluate potential measurement instruments for each core domain to identify at least one measurement instrument that demonstrates it is a truthful, discriminative, and feasible instrument (Figure 3).

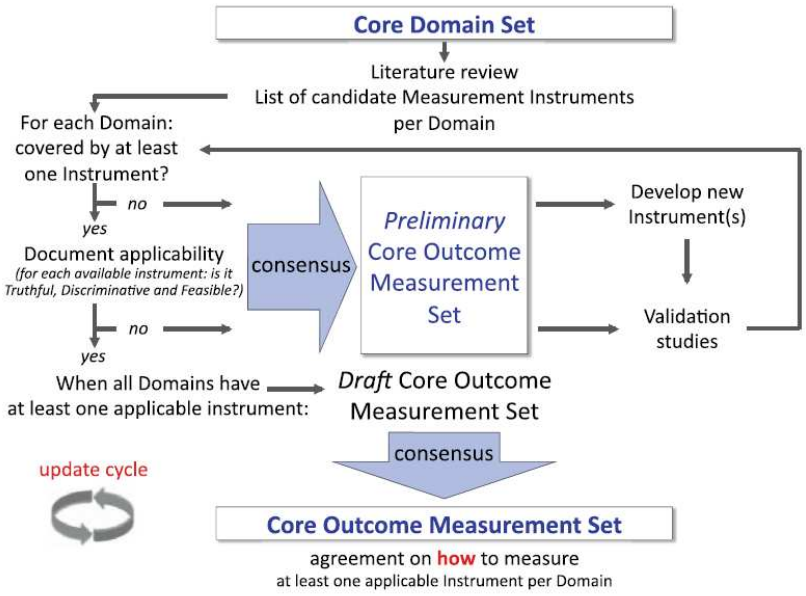


Fig. 3. Development of a Core Outcome Measurement Set from a Core Domain Set. The Core Outcome Measurement Set is defined as: the minimum set of outcome measurement instruments that must be administered in each intervention study of a certain health condition within a specified setting to adequately cover a corresponding Core Domain Set. As depicted, the development process allows core set developers to declare a Preliminary Core Outcome Measurement Set when not all Domains are covered by at least one applicable measurement instrument.

Figure 3: OMERACT Filter 2.0 conceptualization of the development of a core outcome measurement set

Figures 2 and 3 reprinted from Journal of Clinical Epidemiology, Vol.67 (7), Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA et al., Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0, pg. 745-53., Copyright 2015, with permission from Elsevier.

1.4.3. Presenting results of continuous outcomes

Presenting the results of a continuous outcome in an easily interpretable way can be challenging.

The GRADE working group has developed guidance on different approaches to pooling and presenting results for continuous outcome measures (50, 51). If the same outcome instrument has been used in different studies, then a weighted mean difference between the intervention and control group can be presented. To interpret this mean difference, information on the minimum and maximum scores of the instrument and the direction of the scale (e.g. 0 to 10 where 0 means 'no pain' and 10 means 'worse possible pain') is needed. When a common or familiar instrument is used, the interpretation can be quite straightforward. However, if a less familiar outcome instrument is used, presenting the minimal important difference (MID; the smallest change on the scale that patients consider important) can further assist with interpretation. Yet comparing the MID with the mean difference should not be used alone; it

is incorrect to infer that there is a lack of benefit of the treatment if the mean difference is less than the MID because the proportion benefiting is based on a distribution between individuals. The proportion of patients benefiting may still be substantive so it is important to show this alternative presentation as well. There are different methods for determining a MID, and these can change across populations, and an additional issue is that there are many situations where no MID has been established. Another option is to present dichotomized results based on a threshold (ideally pre-specified before seeing the data), though this can lead to loss in statistical power. The threshold can either be change (e.g., a 50% improvement in pain from baseline) or absolute (achieving a state that the patient finds is acceptable, e.g. no worse than mild pain (52) , though presenting both is preferable.

Pooling results from continuous outcomes becomes more complex when clinical trials use different outcome instruments to measure similar domains. The most common approach to dealing with this situation is to present the standardized mean difference (SMD) which is the mean difference divided by the pooled standard deviation. The SMD presents the results in standard deviation units and it is difficult to interpret the meaning of an SMD, though there are guidelines on what may constitute a small, moderate, or large effect size (53). As well, this approach is vulnerable to the heterogeneity of the patients enrolled in the trial. Another option is to rescale the different instruments to a common outcome instrument well-known by clinicians (if there is one available in the outcome of interest), and then undertake meta-analysis using the rescaled scores. If an SMD is calculated in a meta-analysis, both the GRADE paper (50) and the Cochrane Handbook (54) recommend re-expressing the SMD in another format to help readers interpret the result more easily. One method is to multiply the SMD by the standard deviation associated with the most familiar instrument to convert the results back into the natural units of that instrument. The issues outlined in the previous paragraph would still apply to these options.

As described in the GRADE paper (50), another approach is to dichotomize the continuous outcome and present the relative and absolute effects at a threshold. This can be done using a variety of statistical techniques but they rely on the SMD and some statistical assumptions and the meaning of the threshold may not be clear. A less frequently used method is to calculate a ratio of the mean of the intervention group compared to the mean of the control group; however, to use this method both intervention and control groups must show change in the same direction on the instrument and if the control group change is very small, the ratio

result will be deceptively large. The last approach described is similar to the SMD approach but instead of dividing by the standard deviation, the mean difference is divided by the established MID of the outcome. The magnitude of the effect is then described in MID units. This approach depends on the availability of a well-defined MID.

As there are pros and cons to each of the methods described above, it is recommended to provide the magnitude of the effect using more than one presentation method.

Although various methods (as described in the previous section) have been developed to present the results from meta-analyses of continuous outcomes, there is evidence that readers have difficulty with interpreting the magnitude of the effect size between intervention groups. Johnston et al (55) recently conducted a randomized survey (n=531) of clinicians trained in internal and family medicine across eight countries. Participants were presented with results of a hypothetical intervention study to improve chronic pain using different statistical formats (standardized mean difference, mean difference in units of the scale, risk difference, minimal important difference units, relative risk, and ratio of means) and questioned on their understanding of the magnitude of the effect size and how useful they found each format. The participants understanding was greatest when the risk difference was used, followed by relative risk and ratio of means; standardized mean differences were the least understood and reported to be the least useful. There was not a single format which was clearly well understood and perceived to be the most useful by a majority of the respondents. The lack of understanding around SMDs is an important point to highlight; SMDs are recommended in the Cochrane Handbook to conduct a meta-analysis when different clinical trials use different outcome instruments to measure similar domains and the SMD should be re-expressed in another format to enhance understanding (56). Further research is needed into the consistency of results when using these different approaches and which readers find to be the most understandable and easy to interpret.

1.4.4. Cochrane Review Group guidance

Within the Cochrane Collaboration, some clinical conditions fall under the scope of more than one Cochrane Review Group. In the case of fibromyalgia, both the Musculoskeletal Group and the Pain, Palliative, and Supportive Care Group conduct and publish reviews on fibromyalgia. The guidance that the two review groups offers to their systematic review authors differs with respect to recommended outcomes, choices of cut-offs for minimal

important differences, and methods of presentation. The Musculoskeletal Group recommends showing the between group difference in pain intensity as a continuous measure; that is the average change between groups. The Pain, Palliative, and Supportive Care Group advises their authors to present pain in terms of a responder analysis using the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations of a 50% improvement from baseline as being evidence of a substantive improvement and a 30% improvement from baseline as evidence of an important improvement (57, 58). With respect to outcome domains, the Musculoskeletal Group recommends that in addition to pain, sleep and fatigue outcomes are reported in the Summary of Findings table as patients have identified those as important and they form part of the OMERACT core set for fibromyalgia domains (59).

It is not surprising that there are some differences in the guidance provided by different review groups within the Cochrane Collaboration. The guidance has been developed independently with consultation usually focused within specialty clinical areas. In addition to the advice offered by the Cochrane Handbook, methods guidance more specific to a clinical area in terms of searching, inclusion criteria, outcome domains and instruments, and methods of synthesis has been published by some review groups including the Back, Musculoskeletal, and Pain, Palliative and Supportive Care Groups. These have been published in specialty clinical journals: Back Group is published in 'Spine' (60), the Musculoskeletal Group in 'The Journal of Rheumatology' (13, 61), and the Pain, Palliative and Supportive Care Group in 'Pain' (62).

For both systematic review authors and readers of Cochrane Reviews, it is most useful to have standardized guidance about how to report pain outcomes in Summary of Findings tables. For a patient or health care professional faced with various treatment options, having the same outcomes reported using the same methods across systematic reviews of the different treatment options will facilitate easier comparisons for decision making. Similarly, policymakers may be interested in the evidence base for the same intervention used for different health conditions (e.g. biologics) and consistent outcomes and presentation of results may simplify their assessments.

1.5.Outcome measures initiatives

Various international initiatives involved in outcomes research have developed much expertise over the last couple of decades.

The Outcome Measures in Rheumatology (OMERACT) initiative (2) originated over 25 years ago when researchers realized that clinical trials of interventions for rheumatoid arthritis conducted in Europe were using different outcome domains and measures compared to trials of the same interventions being conducted in North America. This variability and lack of standardization of outcomes made it impossible to synthesize and pool the results of trials investigating the same intervention in order to improve the precision of the estimates of benefit and harms of the intervention. The aim of the OMERACT initiative was to reach consensus on which outcomes should be measured in clinical trials on interventions in different conditions in the field rheumatology. The strategy of this initiative was to involve an internationally-based group of health care practitioners, methodologists, and patients and use a data-driven approach to provide the evidence base for discussions until consensus on which outcomes should be measured in clinical trials was reached or if evidence was insufficient, a proposed research agenda. The evidence base consists of reviews of the existing literature, and, if necessary, conducting new studies when there a gap in the evidence base. The aim of the evidence is to demonstrate that an outcome measure has met the 'OMERACT Filter' of 'Truth, Discrimination, and Feasibility' (39) as described above in section 1.4.1. OMERACT Filter 2.0 has been recently developed to provide a more explicit framework for the process of developing core outcome sets (20) (see section 1.4.2).

OMERACT holds a conference every two years to discuss the evidence base and vote on whether there is enough evidence to recommend core sets of outcome measures on various clinical conditions (63). OMERACT has achieved consensus and published recommendations on the key outcomes that should be measured in all clinical trials for various conditions, including: rheumatoid arthritis (64), osteoarthritis (65), osteoporosis (66), ankylosing spondylitis (67), psoriatic arthritis (68), and fibromyalgia (59). The first core set was developed on rheumatoid arthritis and the following outcomes were recommended for inclusion in all clinical trials of interventions for rheumatoid arthritis: tender and swollen joint counts, pain, function, patient and physical global assessment, and imaging (for clinical trials lasting longer than one year). Other projects undertaken by OMERACT include facilitating efforts to define and achieve consensus on minimal clinically important

differences, biomarkers and surrogate endpoints, magnetic resonance imaging (MRI), drug safety, and minimum criteria for economic evaluations.

The work of OMERACT has greatly influenced the clinical trial landscape in rheumatology over the last twenty years. Kirkham et al (69) evaluated the use of the OMERACT core outcome set for rheumatoid arthritis in a cohort of 350 randomized controlled trials on interventions for rheumatoid arthritis to assess the uptake of the core outcome set by trialists since it was published. They found that there was an increase in the number of trials measuring the full rheumatoid arthritis core outcome set in trials on pharmacologic interventions, however the proportion varied depending on the intervention; 85% and 93% of trials assessing the effects of disease-modifying anti-rheumatic drugs (DMARDs) and biologics respectively measured the full core outcome set but only 29% of trials assessing symptom-modifying anti-rheumatic drugs (SMARDs) reported on the full core outcome set. Within non-pharmacological treatments, alternative therapies had a reasonable uptake (88%), but other interventions such as diet, exercise, and rehabilitation had much less. The researchers concluded that, "The adoption of a COS has the potential to increase the consistency in outcomes measured across trials, reduce selective reporting and ensure that trials are more likely to measure appropriate outcomes."

The Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) initiative aims to develop consensus on recommendations for the conduct and interpretation of clinical trials focused on treatments for acute and chronic pain. Similar to OMERACT, IMMPACT holds meetings at which a broad range of stakeholders are represented (different clinical disciplines, regulatory agencies, industry, and consumer groups) and consensus recommendations are reached after extensive discussions. Eighteen meetings have been held since 2002, resulting in a wide range of publications, including recommendations for core outcome domains and outcome measures for chronic pain clinical trials (32, 70), interpreting the clinical importance of outcomes (58), and between-group differences in chronic pain clinical trials (57). Since 2011, IMMPACT has been part of the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION) initiative, a public-private partnership with the United States Food and Drug Administration (FDA) which aims to "to streamline the discovery and development process for new analgesic, anesthetic, addiction, and peripheral neuropathy medications and to more generally accelerate the development of treatments with improved efficacy and safety" (71).

In 2010, the 'COMET' (Core Outcome Measures in Effectiveness Trials) initiative was established with the aim to stimulate research in the area of the standardization of core outcome measures and to develop a repository of resources, including published core outcome sets from all clinical areas (37). It is an international collaboration with the following funders: the UK Medical Research Council, the European Commission, PCORI, and the UK National Institutes for Health Research. COMET holds meetings and workshops to encourage the evidence-based development of core outcome sets and their uptake by trialists. They undertake research on methodological techniques to develop core outcome sets (72) and plan to develop guidelines on producing core outcome sets. A key focus of their activity is to promote patient involvement in core outcome set development and a working group has been established to undertake research on the most effective way to involve patients and the public in this initiative. COMET collaborates with a range of players involved in clinical research including trialists and trial funders, regulators, industry, systematic reviewers, and journal editors. A searchable online database of established core outcome sets is available on the COMET website (www.comet-initiative.org), which helps to promote uptake and reduce duplication of research efforts.

The COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) international initiative established in 2005 complements the work of COMET; within each core outcome domain, a measurement instrument capable of accurately measuring the domain is needed. The purpose of COSMIN is to ensure that the outcome measure instruments used in clinical trials have appropriate measurement properties. As described briefly in section 1.4.1, the COSMIN group has developed a checklist of standards to appraise the methodological quality of studies focused on determining the measurement properties of health-related measurement instruments (41, 73, 74). Similar to systematic reviews of intervention studies, studies on the measurement properties of instruments can be systematically reviewed in order to obtain a complete, unbiased assessment of the evidence base on the validity, reliability, and responsiveness of an instrument. This COSMIN checklist is often used in systematic reviews of the measurement properties of an instrument to determine whether the individual studies included in the review are methodologically sound and whether the appropriate study designs and statistics have been used. As well, it helps guide those who want to undertake a study on assessing the measurement properties of outcome instruments. A searchable database of published systematic reviews of outcome measurement instruments is available on the COSMIN website (www.cosmin.nl).

We are at a point where we need to bring together key players in this field to try to achieve consensus on the best methods for presenting results of chronic pain outcomes in a harmonious manner.

1.6.Background for Cochrane review: Abatacept for rheumatoid arthritis

This section describes the background material for the Cochrane Review which is submitted as part of the requirements for this thesis dissertation. In undertaking this review, we identified the need for the two other projects that are part of this thesis.

1.6.1. Description of the condition

Rheumatoid arthritis is a chronic auto-immune disease which affects the synovial lining of many joints and tendon sheaths resulting in persistent inflammation (75). It is associated with significant morbidity, disability, and impaired quality of life (76). Rheumatoid arthritis incidence is estimated to be 13 to 36 per 100,000 for females and less for males, with a prevalence in the UK as high as 0.8% (77). In terms of costs to society, the loss due to sick leave and disability due to rheumatoid arthritis represents an annual loss in productivity of 1.8 billion GBP with a cost to the UK National Health Service estimated at 560 million GBP (78). In the US, the prevalence of rheumatoid arthritis in white adults over 18 years old was estimated to be 0.6% (79). An estimate of the cost burden of rheumatoid arthritis in the US based on administrative claims databases was that patients' annual health care costs were $8.4 billion while the effects of rheumatoid arthritis cost $10.9 billion. When considering the total of direct, indirect, and intangible (i.e. reduced quality of life and premature mortality) costs, the estimate was $39.2 billion (80).

1.6.2. Description of the intervention

Disease-modifying anti-rheumatic drugs (DMARDs), such as methotrexate (81), leflunomide (82), hydroxychloroquine (83), sulfasalazine (84), and glucocorticoids (85) have been shown to reduce disease activity, to slow disease progression (i.e. reduce the rate of new joint erosions) and/or to improve patients' quality of life. However, a significant proportion of rheumatoid arthritis patients are unable to tolerate these agents for long periods of time or

only experience a partial benefit from these traditional DMARDs, or both. Another class of drugs called 'biologics' has been developed over the past ten years. These drugs mimic substances that occur in the immune system during an inflammatory reaction and are able to specifically target parts of the immune system to reduce inflammation, which in turn reduces the symptoms of rheumatoid arthritis.

Tumour necrosis factor (TNF)-alpha is a protein that the body produces during the inflammatory response. The following biologic agents that target TNF-alpha are currently available: infliximab (Remicade) is a chimeric (mouse/human) monoclonal antibody, golimumab (Simponi) is a fully human monoclonal antibody, etanercept (Enbrel) is a receptor fusion protein that binds to TNF-alpha, adalimumab (Humira) is a recombinant human IgG1 monoclonal antibody specific for human TNF-alpha, and and certolizumab pegol (Cimzia) is a recombinant, humanized antibody Fab' fragment specific for human TNF-alpha. Infliximab, etanercept, and adalimumab have been shown to substantially and rapidly improve rheumatoid arthritis symptoms and to slow radiographic progression (86-88). Golimumab and certolizumab pegol have also received licensing approval and Cochrane Reviews have been conducted on them (89, 90).

Despite their effectiveness, not all patients respond to TNF-alpha blockade and therefore other therapeutic options are needed. Abatacept (brand name Orencia) was approved by the US Food and Drug Administration (FDA) in December 2005 for use in adult patients with moderate to severe rheumatoid arthritis who have not responded adequately either to oral DMARDS (such as methotrexate) or to the TNF-alpha antagonists. Other non-TNF biologic therapies have been developed: rituximab (91) and tocilizumab (92) and are the focus of separate Cochrane Reviews.

1.6.3. How the intervention might work

Abatacept is a selective costimulation modulator, inhibiting T-cell (T lymphocyte) activation by binding to CD80 and CD86 (the costimulatory antigens), thereby blocking interaction with CD28 (the costimulatory receptor). It is the first biologic to work by disrupting T-cell activation. Activated T-cells occur early in the inflammatory reaction so by preventing their activation, the chain of events that leads to joint inflammation, pain, and damage is prevented. Abatacept is administered intravenously over approximately 30 minutes and after

the first dose additional doses are given at two and four weeks and then every four weeks (93).

1.6.4. Why is it important to do this systematic review?

The use of biologics is limited by their high cost and uncertainty about adverse events. Although estimates vary by country, the annual cost of etanercept treatment is estimated at $17,160 CAD and $21,385 CAD for infliximab and abatacept is similar at $21,384 CAD (94). While expensive, if supported by the overall body of evidence, the claims of their benefit upon both symptoms and radiographic progression, and their low rate of short term side effects make them of great interest to patients with rheumatoid arthritis. At this time it is appropriate to conduct a systematic review of randomized controlled trials of abatacept to quantify the benefits and potential harms of its use.

This chapter contains text from the two published qualifying manuscripts submitted as per the TRIBE PhD requirements: (1) Maxwell, L. J. et al: Current State of Reporting Pain Outcomes in Cochrane Reviews of Chronic Musculoskeletal Pain Conditions and Considerations for an OMERACT Research Agenda. J Rheumatol 2015;42 (10);1934-1942. All rights reserved; (2) Maxwell L, Singh J. Abatacept for rheumatoid arthritis. The Cochrane Database of Systematic Reviews 2009, Issue 4. Art. No.: CD007277. DOI: 10.1002/14651858.CD007277.pub2

2. AIMS OF RESEARCH AND HYPOTHESES

Different Cochrane Review Groups and other organizations have developed guidance on reporting pain outcomes from intervention trials. Since this guidance was developed mostly independently, it is not surprising that there are differences; these include different outcome domains, instruments, methods of analysis and transformations, cut points and thresholds for minimally important and clinically important differences, interpretation of clinical or policy relevance, and methods of presentation.

I have published a Cochrane systematic review of Abatacept for rheumatoid arthritis and will use this as my second qualifying paper (95). The objective of this systematic review was to assess the benefit and harm of abatacept in reducing disease activity and pain, and improving function in people with rheumatoid arthritis. It was undertaking this systematic review, along with other systematic reviews that I have co-authored, that provided me with the idea and initiative to conduct the two studies which I describe below. The different pain outcome measures reported in both individual trials and other systematic reviews provided a challenge in both synthesizing the evidence in meta-analyses within systematic reviews and ensuring comparability of results across different systematic reviews.

Given we suspect that the results of pain outcomes are not reported in a consistent manner in Cochrane Summary of Findings tables, there is a need to ascertain the extent to which this is true. Variability and lack of consensus hampers efforts to synthesize evidence and results in an inconsistent and possibly confusing message for readers (including patients, caregivers, practitioners, and policymakers) of Cochrane Reviews. There is a need to develop a research program in partnership with different organizations working on the development and assessment of patient-important outcomes to address this problem.

Hypotheses

Pain outcomes reported in Cochrane Reviews of chronic musculoskeletal conditions are not reported in a consistent manner. Eliciting the opinions of stakeholders, including patients, health care providers, and methodologists to identify key research topics will improve commitment to this project and contribute to achieving consensus on best practices for reporting in Cochrane Summary of Findings tables on chronic musculoskeletal conditions.

3. METHODS

3.1.Methods for Cochrane Review: Abatacept for rheumatoid arthritis

Criteria for considering studies for this review

Types of studies

Randomized controlled trials (RCTs) were included. To be eligible for inclusion, the generation of the allocation sequence had to be truly random; for example, generation of the sequence by a computer or random numbers table. Trials had to be a minimum of three months duration. Trials of less than six months duration were used to investigate short-term benefit and harm while studies longer than six months addressed longer-term benefit and harm. We considered data from published and unpublished RCTs for inclusion. We checked websites of regulatory agencies for reported adverse effects.

Types of participants

Patients at least 16 years of age meeting the ACR 1987 revised criteria for rheumatoid arthritis (96).

Types of interventions

RCTs comparing abatacept alone or in combination with DMARDs or biologics to placebo or other DMARDs or biologics. There were no restrictions with regard to dosage or duration of intervention.

Types of outcome measures

Major outcomes:

Benefit

The primary outcome is the ACR 50 response rate to treatment with abatacept as defined by the American College of Rheumatology (ACR) (97). The variables included in this definition are:

- tender joint count;
- swollen joint count;
- patient's assessment of pain (visual analogue scale (VAS) or Likert scale);

- patient and physician assessment of disease activity (VAS or Likert scale);
- patient assessment of functional ability (Health Assessment Questionnaire (HAQ), Arthritis Impact Measurement Scales (AIMS), McMaster Toronto Arthritis (MACTAR)); and
- laboratory parameters (i.e. acute phase reactants, such as erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP).

An ACR 20/50/70 response is defined as a 20%/50%/70% improvement in tender and swollen joint counts and the same level of improvement in three of the five following variables: patient and physician global assessments, pain, HAQ, and acute phase reactants.

Adverse events

Since RCTs are usually of limited duration, mainly short-term adverse events were assessed. However, regulatory agency websites and long-term extensions of included RCTs were also reviewed for potential longer-term adverse events.

Specific adverse event outcomes of interest were:

- adverse events, including allergic reactions, and infections;
- serious adverse events, including serious infections, and lymphoma; and
- withdrawals due to lack of efficacy, and adverse events.

Secondary outcomes:

- Individual ACR criteria and ACR 20 and 70 response criteria as outlined above.
- Radiographic progression, as measured by the Sharp, modified Sharp or Larsen methods (also considered a primary outcome for studies longer than one year in duration).
- European League Against Rheumatism (EULAR) criteria (98) which define response (good, moderate and none) according to certain cut-offs for both the absolute values and relative changes in the Disease Activity Score (DAS) (99). The DAS is a composite index that includes the combination of the values of tender and swollen joint counts, patient's global assessment of disease activity, and erythrocyte sedimentation rate (ESR) value. When a 28-joint count is used the index is reported as DAS 28. The DAS28 is scored on a scale from 0 to 10 to indicate the current activity

of rheumatoid arthritis; a higher number indicates higher disease activity. According to the DAS-Score website, “A DAS28 above 5.1 means high disease activity whereas a DAS28 below 3.2 indicates low disease activity. Remission is achieved by a DAS28 lower than 2.6.”(100). A 'good' EULAR response is defined as a decrease in the DAS or DAS 28 of more than 1.2 points from baseline with a final DAS less than 2.4 (or DAS 28 less than 3.2). A EULAR response of 'None' is defined as a decrease in DAS or DAS 28 less than 0.6 or a decrease greater than 0.6 and less than 1.2 with a final DAS greater than 3.7 (or DAS 28 greater than 5.1). Any other scores are regarded as 'moderate' response.

- Health-related quality of life (HRQOL) as measured by the SF-36 or other instruments.

Search methods for identification of studies

Electronic searches:

The original search strategy developed for MEDLINE in the protocol was further refined and is reported in Appendix of the Cochrane Review. We screened 492 records from the original search in MEDLINE and compared these to the results of the new search strategy. All the records of interest retrieved using the original strategy were contained in the new search results. The MeSH headings of 'Immunosuppressive Agents' and 'Antirheumatic Agents' in the original search strategy were removed. In addition, since abatacept has a different mechanism of action from the tumour necrosis factor (TNF) biologics, references to TNF were removed from the original strategy. The MeSH headings of 'Immunoconjugates' and 'Antigens, Differentiation' were retained in the revised strategy. The new search strategy for MEDLINE was adapted for the other electronic databases as shown in the appendices.

We searched the following electronic databases initially up to March 2007: the Cochrane Central Register of Controlled Trials (CENTRAL) (*The Cochrane Library,* 2007 Issue 1) MEDLINE, EMBASE, ACP Journal Club, and ISI Web of Science (Biosis Previews). We searched the FDA website for references to trials of abatacept. We searched abstracts from ACR and EULAR conferences using Biosis Previews.

The search was not limited by language, year of publication or type of publication.

We ran an updated search in January 2009 to capture publications between 1 January 2007 and 31 December 2008.

Searching other resources

We also searched reference lists from comprehensive reviews and identified clinical trials. We contacted content experts and the pharmaceutical company that manufactures abatacept to obtain clarification and any relevant additional unpublished data.

We searched websites of the following regulatory agencies for reported adverse events using the terms 'rheumatoid arthritis', 'abatacept' and 'orencia' on 1 April 2009.

- 'Current Problems in Pharmacovigilance' (http://www.mhra.gov.uk/Publications/Safetyguidance/CurrentProblemsinPharmacovigilance/index.htm) (this was superseded by 'Drug Safety Update' in July 2007. Both databases were searched under 'drug alerts').
- Australian Adverse Drug Reactions Bulletin (http://www.tga.gov.au/adr/aadrb.htm).
- Food and Drug Administration FDA Medwatch (US) - Adverse Event Reporting System (AERS) FDA Medwatch (http://www.fda.gov/medwatch/safety.htm).
- European Public Assessment Reports from the European Medicines Evaluation Agency (http://www.emea.europa.eu/).

Data collection and analysis

Selection of studies

We used Reference Manager 11 software to manage the records retrieved from the searches of the electronic databases. We tracked results from handsearching on paper. We created the data extraction form in Word and captured all article information except outcome results in this form. We tracked outcome results in an Excel spreadsheet for easier entry into RevMan (102).

Two authors (LM, JS) independently reviewed the results of the various searches. We reviewed titles and abstracts and when more information was required to determine whether the trial met the inclusion criteria, we obtained the full text. We kept a record of reasons for excluding studies. We resolved disagreement by consensus and there was no need to contact a third party for a decision. Two German language articles were summaries of included studies so no further translation was required.

Data extraction and management

Two authors (LM, JS) independently extracted data from the included trials and entered these into RevMan 5. Variance measures were missing for many continuous outcomes (only P values were reported in the published articles) so we obtained additional data from Bristol-Myers Squibb.

We decided a priori that the following data from each trial would be extracted.

- General study information, such as title, authors, contact address, publication source, publication year, country, study sponsor.
- Characteristics of the study: design, study setting, inclusion/exclusion criteria, quality criteria (e.g. randomization method, allocation procedure, blinding of patients, caregivers and outcome assessors, withdrawals and drop-outs, intention-to-treat (ITT) analysis).
- Characteristics of the study population and baseline characteristics of the intervention and control groups (age, sex, duration of disease, treatment history, presence of co-morbidity and peripheral disease, concurrent treatments) and numbers in each group.
- Characteristics of the intervention, such as treatment comparators, dose, method of administration, frequency of administration, and duration of treatment.
- Outcomes measures as noted above (changes in disease outcome, adverse events, withdrawal from treatment).
- Results for the intention-to-treat population (if reported), outcome measures at the end of the placebo phase, and any summary measures with standard deviations, confidence intervals, and P values where given, drop-out rate, and reasons for withdrawal.

Assessment of risk of bias in included studies

Two independent authors (LM, JS) assessed the risk of bias of the included studies. As recommended by the *Cochrane Handbook for Systematic Reviews of Interventions* (101), the following methodological domains were assessed.

- Sequence generation - was the method used to generate the allocation sequence appropriate to produce comparable groups?
- Allocation sequence concealment - was the method used to conceal the allocation sequence appropriate to prevent the allocation being known in advance of, or during, enrolment?
- Blinding of participants, personnel and outcome assessors - were measures used to blind study participants, personnel, and outcome assessors from knowledge of which intervention a participant received?
- Incomplete outcome data - how complete were the outcome data for the primary outcomes? Were drop-out rates and reasons for withdrawal reported? Were missing data imputed appropriately? We considered an overall completion rate of 80% or higher as a low risk of bias. If completion rates were only provided by group, a less than 80% completion rate in the treatment group was considered a high risk of bias.
- Selective outcome reporting - were appropriate outcomes reported and were any key outcomes missing?
- Other potential threats to validity (considering external validity, e.g. relevant use of co-interventions) - what was the funding source of each of the studies?

We explicitly judged each of these criteria using: low risk of bias; high risk of bias; and unclear risk of bias due to either lack of information or uncertainty over the potential for bias.

Measures of treatment effect

We analyzed the results of the studies using RevMan 5.0 (102). We summarized data in a meta-analysis if they were sufficiently homogeneous, both clinically and statistically. We expressed continuous data as mean difference (MD) or standardized mean difference (SMD), depending on the similarity of scales measuring an outcome. We expressed dichotomous data as relative risk (RR) or in the case of rare events (< 10%), such as death, we used the Peto odds ratio (Peto OR).

Some transformations were necessary to enter continuous data into RevMan. For Kremer 2006 (103), standard error (SE) was converted to standard deviation (SD) using the formula, SD= SE x sqrt(N). The mean percent improvement from baseline and standard error were provided for Kremer 2003 (104). Mean percent improvement was used to calculate the end of

study score using the formula e-b/b x 100=% improvement from baseline and the standard deviation at baseline was assumed for the standard deviation at end of study.

Assessment of heterogeneity

In addition to reviewing forest plots, we formally tested heterogeneity of the data using the Chi^2 with a P value < 0.10 indicating significant heterogeneity. We also assessed the I^2 statistic (105). A value greater than 50% may indicate substantial heterogeneity. In the case of substantial heterogeneity, we explored the data further, including subgroup analyses, in an attempt explain the heterogeneity.

Assessment of reporting biases

A funnel plot was performed to assess the possibility of publication bias.

Data synthesis

It was expected that the trials would be performed in similar populations and that there would be little 'between-study' variation. Thus, we specified a fixed-effect model a priori. However, if significant heterogeneity was found and could not be explained, we decided that a random-effects model would be used to assess the results.

Subgroup analysis and investigation of heterogeneity

We planned the following subgroup analyses a priori in order to explore possible effect size differences.

1. Intervention - different dosage, duration of treatment.
2. Characteristics of participants - severity of baseline disease; age; disease duration; sex; disease with or without peripheral joint involvement.

For this review, we assessed results separately at three, six and 12 months, by two dosages (2 mg/kg and 10 mg/kg), by duration of disease (average of less than eight years and greater than eight years), and by study eligibility criteria (anti-TNF failures or DMARD failures).

Sensitivity analysis

We planned the following sensitivity analyses a priori in order to explore effect size differences and the robustness of conclusions.

1. Effect of study quality - defined as adequate allocation concealment and outcome assessor blinding.

2. Effect of imputation of missing data or statistical transformations.

Summary of findings table

We completed 'Summary of findings' tables included in RevMan 5 in order to communicate the key outcomes of the review. The outcomes for inclusion were: ACR50% improvement, Pain, Function, Achievement of low disease activity state, Total serious adverse events, Change in radiographic progression, Long-term serious adverse events.

We determined the absolute risk difference and relative percent change and entered these into the comments column of the 'Summary of findings' table. For dichotomous data, the absolute risk difference is calculated by using RevMan to generate the Risk Difference analysis and then reporting the result as a percentage. The relative percent change is calculated by finding the relative risk (RR) from RevMan and then applying the formula RR-1 equals the relative percent change. The number needed to treat (NNT) was calculated from the control group event rate (unless the population event rate was known) and the relative risk using the Visual Rx NNT calculator (106).

For continuous outcomes, the absolute risk difference is the mean difference expressed as a percentage. The relative percent change is the absolute change divided by the baseline mean of the control group. The NNT was calculated using the Wells calculator software available at the Cochrane Musculoskeletal Group editorial office. The minimal clinically important difference (MCID) for pain was 20%, based on Tubach et al (107) for input into the calculator. We also carried out a sensitivity analysis for 30%, based on Farrar et al (108).

We used GRADEPro software to create the Summary of findings table and the GRADE criteria to provide an overall grading of the quality of the evidence.

Additional data

We contacted trial authors and Bristol-Myers Squibb, the manufacturer of abatacept, additional information about risk of bias aspects of the trials (e.g. allocation concealment and

blinding) and variance and other outcomes not reported in the published reports. Some of this data was provided.

3.2. Methods for Study 1: Assessment of reporting of pain outcomes in Cochrane Reviews

On July 17, 2013, we conducted a search of titles/abstracts/keywords of the Cochrane Library using the key word "pain" to identify all intervention reviews (excluding overviews) in defined chronic musculoskeletal painful conditions from the Back, Musculoskeletal, and Pain, Palliative, and Supportive Care Cochrane Review Groups that contained a SoF Table. These three Cochrane Review Groups are responsible for the conditions that fall within the World Health Organization definition of a chronic musculoskeletal condition: "inflammatory rheumatic diseases such as rheumatoid arthritis, osteoporosis and other bone diseases, osteoarthritis and related conditions, soft-tissue periarticular disorders, back pain" (109). We independently extracted data in duplicate on the pain domains and instruments in Summary of Findings tables. The pain domain and instrument data from the abstracts and included studies data was extracted by one person for feasibility reasons. We used Excel2010 for data management and analysis.

We extracted composite measures if they contained a pain component. If a SoF table reported more than one pain outcome and/or instrument, all were extracted. When a SoF table reported only a standardized mean difference (SMD), this was recorded as 'no instrument reported'. If a SMD was reported along with the transformation of the SMD to a specific instrument, then the instrument used for the back-transformation was extracted.

We also extracted data on pain outcomes reported in the abstract and included studies of the Cochrane reviews. We used the 'Characteristics of Included Studies' table in each review to obtain information on the pain domain and instrument reported in the included studies. This information was not presented in the published journal article due to space limitations but will be presented in the Results section of the thesis dissertation.

3.3.Methods for Study 2: Survey and interviews of key stakeholders regarding presentation of pain outcomes in Cochrane SoF tables

The aim of our survey and interviews was to obtain information from key stakeholders, including patients, clinicians, and methodologists, on the most important aspects to consider when expressing the pain response of trial participants in chronic musculoskeletal pain intervention studies with respect to Cochrane systematic reviews.

Participants were asked: (1) which of the following domains of pain are important for reporting in a SoF table: pain intensity, pain frequency, pain interference with function, or other domains; (2) what is the best way to present measures of change; and (3) what are the important thresholds/cut-off for identifying responders in (i) change scales and (ii) achieving predefined absolute 'states'.

We used a purposive, expert sampling technique to select survey and interview participants to obtain representation from Cochrane Review Groups, international initiatives involved in outcome measures methodology, patients with painful musculoskeletal conditions, health care practitioners, and methodologists. Prospective participants were sent a link to the survey via email and were asked to participate in an interview to provide more detailed comments. Two members of the project team drafted the survey (a rheumatologist/journal editor/systematic reviewer and senior outcomes researcher [PT], and a managing editor and systematic reviewer [LJM]) that consisted of open-ended text responses.

We piloted our survey with three invitees, and revised in response to comments. We administered our survey using SurveyMonkey™ and invitees were sent two reminders to complete the survey. The interviews were conducted via telephone by a researcher trained in conducting semi-structured interviews.

The interview guide followed the sequence of the survey (Appendix A). Interviews lasted between 30 and 75 minutes. The majority of interviews (21/24, 88%) were recorded, transcribed verbatim and then coded. Three were coded from notes taken during the interviews as the audio recording equipment malfunctioned.

The text data were analyzed by one researcher (LJM) and checked by a second (PT) for themes using a directed approach of qualitative content analysis (110). This directed

approach was used as the survey was developed around existing ideas for themes as identified by OMERACT Executive Committee members.

Interpretation of the results started from the responses to these ideas and then ascertained and counted new topics identified by the participants.

Ethics approval for the survey and interviews was obtained from the University of Split, School of Medicine Ethical Committee, Croatia.

This chapter contains text from the two published qualifying manuscripts submitted as per the TRIBE PhD requirements: (1) Maxwell, L. J. et al: Current State of Reporting Pain Outcomes in Cochrane Reviews of Chronic Musculoskeletal Pain Conditions and Considerations for an OMERACT Research Agenda. J Rheumatol 2015;42 (10);1934-1942. All rights reserved; (2) Maxwell L, Singh J. Abatacept for rheumatoid arthritis. The Cochrane Database of Systematic Reviews 2009, Issue 4. Art. No.: CD007277. DOI: 10.1002/14651858.CD007277.pub2

4 RESULTS

4.1 Results for Cochrane Review: Abatacept for rheumatoid arthritis

The entire publication version of this Cochrane Review is provided as a separate submission to this thesis. Key results related to the reporting of pain outcomes will be highlighted in this section.

Seven clinical trials reported in eleven journal articles were included in this review. The seven trials are: Genovese 2005 (111); Kremer 2003 (104); Kremer 2006 (103); Moreland 2002 (112); Weinblatt 2006 (113); Weinblatt 2007 (114); Schiff 2008 (115).

There were multiple publications of some trials which reported additional outcomes different from the main trial publication. Emery 2006 (116), Kremer 2003, and Kremer 2005 (117) all referred to one trial; for the purpose of this review, Kremer 2003 is considered the primary publication. Kremer 2006 and Russell 2007 (118) referred to one trial and for the purpose of this review, Kremer 2006 is considered the primary publication. Genovese 2005 and Westhovens 2006 (119) referred to a single trial and Genovese 2005 is considered the primary publication. Cole 2008 (120) is a publication of health related quality of life data from the Genovese 2005 and Kremer 2006 trials

Trials reporting results for patient-reported pain did not provide any measures of variance in Kremer 2003, Moreland 2002 and Weinblatt 2006. Weinblatt 2007 did report variance in the published article and Kremer 2006 provided the mean and standard error in an appendix (which we converted to standard deviation as required for the meta-analysis). Moreland 2002 reported pain on a 1 to 5 scale. Kremer 2003 used a 0 to 100 VAS for pain. Weinblatt 2006 reported pain on a VAS and we assumed 0 to 100 given the numerical results. Weinblatt 2007 did not report the pain scale. The Genovese 2005 and Schiff 2008 trials did not report separate results for pain in the published articles.

Table 1 provides the results for pain outcomes that were reported in the trials for the comparison of abatacept (10 mg/kg and 2 mg/kg combined) + DMARDs or biologics versus placebo + DMARDs or biologics. We contacted trial authors and Bristol-Myers Squibb to obtain the missing information. Variance measures were obtained for Kremer 2003 and Weinblatt 2006. In Kremer 2003 there was a statistically and clinically significant reduction in pain on a 100 mm visual analogue scale (VAS) (lower score means less pain) in

the abatacept + MTX group compared to placebo + MTX at both six and 12 months (MD -22.49, 95% CI -28.00 to -16.98 at 12 months). In Kremer 2006, the abatacept group had statistically significantly less pain at 12 months compared to placebo (MD -12.60, 95% CI -16.82 to -8.38), though this is a small clinical difference. Based on moderate quality evidence from Weinblatt 2006, there was a statistically significant and small clinical reduction in patient-reported pain between groups at 12 months (mean difference (MD) -10.71, 95% CI -12.97 to -8.45). When pooling these three studies, the overall MD was -12.45, 95% CI -14.33 to -10.57, but there was high heterogeneity: I^2=87% (Figure 4). Results from Weinblatt 2006 were chosen to be presented in the Summary of findings table for pain because of the high heterogeneity when the results were pooled and because this was a large study with a wide variety of participants which is more generalizable to the general population. The relative percent change from baseline was -18% (95% CI -22% to -14%). The NNT was 5 (95% CI 4 to 6) when a minimal clinically important difference (MCID) of 20% was assumed and 8 (95% CI 6 to 10) when an MCID of 30% was assumed. The mean difference in pain scores between abatacept + etanercept and placebo + etanercept groups was not statistically or clinically significant in Weinblatt 2007.

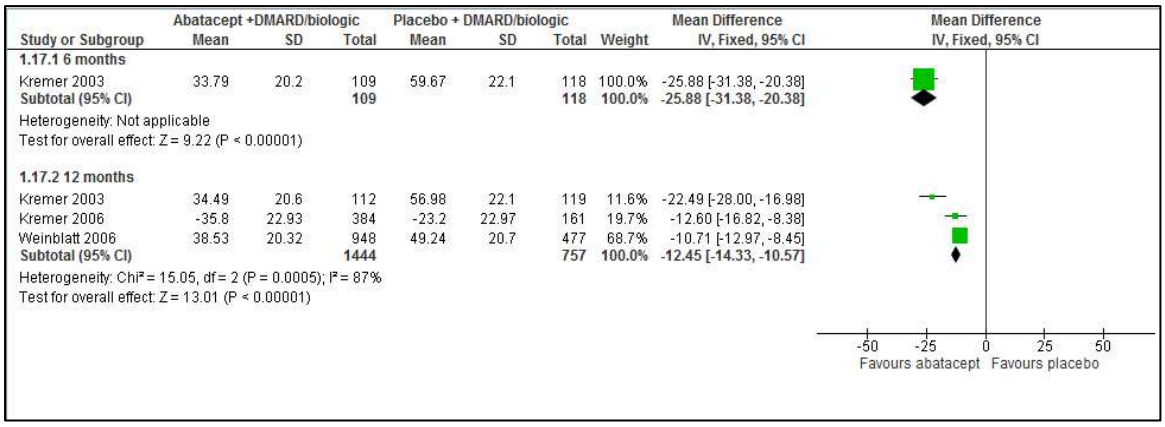


Figure 4: Forest plot of abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic versus placebo + DMARDs/biologic, Outcome: Patient reported pain (100 mm VAS)

Table 1: Patient-reported pain outcomes in Abatacept for rheumatoid arthritis systematic review

Study ID	Pain scale	Baseline mean	End of study mean	% improvement, mean	Mean change from baseline	Variance (P value or CI)
Moreland 2002 - abatacept; 85 days	1-5	3.47	2.43	28.1	-	NR
Moreland 2002 - placebo; 85 days	1-5	3.55	3.24	4.6	-	NR
Kremer 2003 - abatacept; 6 months	100mm VAS	NR	NR	-	-46.4	P < 0.05 (between groups)
Kremer 2003 - placebo; 6 months	100mm VAS	NR	NR	-	-8.4	-
Weinblatt 2006 - abatacept; 1 yr	VAS	NR	NR	-	-26.3	P < 0.001 (within group)
Weinblatt 2006 - placebo; 1 yr	VAS	NR	NR	-	-16.4	P < 0.001 (within group)
Weinblatt 2007 - abatacept; 1 yr	NR	65.5	43.6	33.4	-22.0	P < 0.001 (within group)
Weinblatt 2007 - placebo; 1 yr	NR	53.2	47.4	10.9	-7.1	P < 0.001 (within group)
Kremer 2006 - abatacept; 1 yr	100mm VAS	NR	NR	-	-35.8	-12.6 95% CI (-16.9 to -8.39)
Kremer 2006 - placebo; 1 yr	100mm VAS	NR	NR	-	-23.2	

NR = not reported; VAS = visual analogue scale; *calculated as the average of the changes in the individual patient data

Table 2: Summary of findings: Abatacept (2 and 10 mg/kg) +DMARDs/biologic versus placebo + DMARDs/biologic for RA

Patient or Population: patients with rheumatoid arthritis **Settings: International; clinic/hospital** **Intervention: Abatacept (2 and 10 mg/kg) + DMARDs/biologic** **Comparison: Placebo +DMARDs/biologic**						
Outcomes	**Placebo +DMARDs/ biologic**	**Abatacept (2 and 10 mg/kg) +DMARDs/ biologic**	**Relative Effect (95% CI)**	**No of Participants (Studies)**	**Quality of Evidence (GRADE)**	**Comments (95% CI)**
ACR 50% improvement **Follow-up: 12 months**	168 per 1000	371 per 1000 (291 to 474)	RR 2.21 (1.73 to 2.82)	993 (3 studies)	⊕⊕⊕⊖ Moderate [1, 2, 3]	Absolute difference= 21% (16% to 27%). NNT=5 (4 to 7) Relative percent change=121% (73% to 182%).
Pain **measured at end of study on a 100 mm visual analog scale. Scale from 0 (better) to 100 (worse).** **Follow-up: 12 months.**	The mean pain in the control group was 49.24 mm	The mean pain in the intervention group was 10.71 lower (12.97 to 8.45)		1425 (1 study)	⊕⊕⊕⊖ Moderate[2]	Absolute difference= -11% (-13% to -8.5%). NNT=5 (4 to 6) Relative percent change=-18% (-22% to -14%).
Improvement in physical function (HAQ: greater than 0.3 increase from baseline, 0-3 scale) **Follow-up: 12 months**	393 per 1000	637 per 1000 (531 to 766)	RR 1.62 (1.35 to 1.95)	638 (1 study)	⊕⊕⊕⊖ Moderate[1]	Absolute difference= 24% (16% to 32%). NNT=5 (4 to 7) Relative percent change=62% (35% to 195%).
Achievement of low disease activity state (DAS 28 less than 3.2, scale 0-10) **Follow-up: 12 months**	98 per 1000	424 per 1000 (278 to 646)	RR 4.33 (2.84 to 6.59)	683 (1 study)	⊕⊕⊕⊖ Moderate[1]	Absolute difference= 33% (26% to 39%). NNT=4 (3 to 5) Relative percent change=333% (184% to 559%).
Total serious adverse events **Follow-up: 6 to 12 months**	121 per 1000	127 per 1000 (105 to 155)	RR 1.05 (0.87 to 1.28)	3151 (6 studies)	⊕⊕⊕⊖ Moderate[1, 2, 3, 7]	Absolute difference= 1% (-2% to 3%). NNT=n/a[4] Relative percent change=5% (-14% to 29%).

Change in radiographic progression measured by Genant-modifed Sharp erosion score (increase in score means more joint damage). Scale from: 0 to 145. Follow-up: 12 months	The median change in radiographic progression in the control group was 0.27 units	The median change in radiographic progression in the intervention group was 0 units		586 (1 study[6])	⊕⊕⊕⊖ moderate[1,8]	Note there was no change in the abatacept group. MD -0.27 (-0.42, -0.12).Absolute RD=-0.2% (-0.3% to -0.08%). Relative percent change=-1.2% (-1.9% to -0.6%).[9]
Long-term serious adverse events Follow-up: 2 years	See comment	See comment	Not estimable	950 (2 studies [11])	⊕⊕⊖⊖ low[10]	Number of patients with SAE: Genovese 2005: 103/357; 23.4 SAE/100 patient-years; 70% completed the LTE. Kremer 2006: 149/593; 16.3 SAE/100 patient-years; 90.5% completed the LTE

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio;

GRADE Working Group grades of evidence:

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

1 Kremer 2006: Intention to treat analysis not performed. 9 patients in abatacept group and 5 in placebo group excluded from analysis.

2 Weinblatt 2006: 15 people randomized were not treated and not included in analysis

3 Kremer 2003: Risk of attrition bias - less than 80% completion rate in treatment group at 12 months

4 NOTE: Number needed to treat (NNT)=n/a when result is not statistically significant. NNT for dichotomous outcomes calculated using Cates NNT calculator (http://www.nntonline.net/visualrx/). NNT for continuous outcomes calculated using Wells Calculator (CMSG editorial office).

5 Outcome based on Weinblatt 2006

6 Outcome based on Kremer 2006

7 Weinblatt 2007: Risk of attrition bias - less than 80% completion rate in the treatment group at 12 months

8 Radiographic data obtained for 90% of study participants

9 RD=risk difference

10 Long-term serious adverse events based on observational data. Two RCTs had a long-term extension (LTE) phase in which people in the placebo group during the RCT switched to abatacept for the LTE.

11 Based on 2 long-term extension studies (LTE) of RCTs. Participants on placebo in the RCT switched to abatacept treatment

4.2 Results for Study 1: Assessment of reporting of pain outcomes in Cochrane Reviews

Our search of the Cochrane Library website identified 57 reviews containing a Summary of Findings table that assessed interventions for chronic painful musculoskeletal conditions (see study flow chart in Appendix B and list of included conditions in Appendix C). All but one Summary of Findings table (98%, 56 of 57) reported an outcome of pain (Table 3). Over two-thirds of the Summary of Findings tables (41, 72%) reported only the word "pain" in the outcome column in the table and we assumed the domain of interest was pain intensity based on the scales that were reported.

All 56 Summary of Findings tables that reported pain presented a measure of pain intensity, in either single or multiple domain instruments; twenty different instruments were reported, with the visual analogue scale (VAS) being the most frequent (45%). Pain intensity was measured using a continuous scale for all but five outcomes: four responder analyses and the outcome 'number of people with resting pain' (Table 3). Pain interference was reported in eight Summary of Findings tables (5 different instruments) and pain frequency was reported in one multiple domain instrument in a single Summary of Findings table.

No Summary of Findings tables reported other aspects of pain, such as pain quality or pain affect (14). In ten (16%) Summary of Findings tables, the instrument for measuring pain intensity was not reported, and of these, six reported a standardized mean difference (SMD) with no re-expression using a familiar instrument.

Table 3: Pain outcome domains and instruments reported in included Cochrane Summary of Findings (SoF) tables

SoF tables N=57 reviews*			
Outcome Domain/Subdomain	*N SoF tables*	*Outcome Instrument*	*N*
Pain intensity	48**	***Unidimensional pain intensity scales:***	
		VAS (0-10 cm or 0-100 mm)	28
		VAS (1-9)	11
		Verbal rating score (0-10)	1
		10-point Likert scale	1
		Multidimensional pain intensity scales:	
		WOMAC pain subscale score	
		Hospital for Special Surgery pain subscale score	1
		Dichotomous outcomes – instrument not reported	1
		At least 50% improvement from baseline	
		Patient Global Impression of Change [in pain] much or very much improved	2
		IMMPACT definition - any substantial pain benefit	2
		IMMPACT definition - at least moderate pain benefit	1
		Number of participants with resting pain	1
		Instrument not reported (only SMD reported)	1
		Pain reported in SoF as an outcome, but not measured in included studies	10 (6)
			6
Pain intensity/tender joints	2	Number of tender joints	1
		Number of tender points	1

Multi-domain outcomes including pain intensity	14	ACR50 response criteria ASAS40 response criteria[#] ASAS Partial remission response criteria[#] ASES Shoulder Score Disease Activity Score (DAS28) Hospital for Special Surgery knee score Lequesne Index QUALEFFO[#]	9 1 1 1 6 1 1 1
Multi-domain/dimension outcomes including pain intensity and pain interference	6	Neck Disability Index DASH Fibromyalgia Impact Questionnaire	2 3 1
Multi-domain/dimension outcomes including pain interference	1	SF-12	1
Multi-domain/dimension outcomes including pain intensity, pain frequency and pain interference	1	Osteoporosis quality of life	1
Pain not reported in SoF table	1	Not applicable	

*Note: More than one pain outcome can be reported per SoF table; 1 review had no included studies but had an SoF table for the major outcomes **41 reported only 'pain' but we assumed the domain was pain intensity from the scale; [#]an outcome of interest for the SoF table but no study reported it; ACR 50=American College of Rheumatology 50% response criteria; ASES=American Shoulder and Elbow Surgeons Shoulder score; DASH=Disabilities of the Arm, Shoulder, and Hand ;VAS=visual analogue scale; WOMAC=Western Ontario and McMaster Universities Arthritis Index. QUALEFFO-41 Questionnaire is a quality of life questionnaire in patients with vertebral fractures; SMD=standardized mean difference

In addition to the assessment of the reporting of pain outcomes in the Summary of Findings table which was reported in the published paper (121), we also assessed the abstracts and included studies for each the 57 reviews that had a Summary of Findings table as additional information for the thesis.

The result for the assessment of abstracts was similar to that of the Summary of Findings tables, with mainly the domain of pain intensity reported in either single- or multiple-domain instruments using a variety of scales. The visual analogue scale (VAS) was reported in over

half of the unidimensional measures of pain intensity (Table 4). One instrument encompassing both pain intensity and pain interference was reported in an abstract. Thirty-nine percent (18/46) of abstracts did not report the pain intensity outcome instrument and a further 10 abstracts reported only a SMD with no re-expression of the SMD.

There were a total of 617 included studies and 143 of these (23%) did not report a measure of pain (Table 5). A total of 758 scales were reported. As expected given the results of the Summary of Findings tables and abstracts, pain intensity was the most frequently reported pain domain and the VAS for pain intensity was the most frequent scale used in the included studies. Only 32/498 (6%) of trials reported NRS. Specific subdomains of the pain measure, for example, pain at night, pain at rest, pain on activity, were occasionally reported, but these specifics were not reported in the abstract or Summary of Findings table. The domain of 'pain interference' included instruments that had a measure of pain as well as a measure of whether pain impacted on function, range of motion, limitations to activities of daily living, etc. Systematic review authors often reported these scales under the domain of 'function' in the Summary of Findings tables. Only three of the included studies reported a unidimensional measure of pain interference. There were a few examples of where the systematic review authors included an important outcome in the Summary of Findings table (e.g. ASAS40, ASAS Partial Remission, QUALEFFO) but none of the studies included in the review measured that important outcome.

Table 4: Pain outcome domains and instruments reported in abstracts of Cochrane Reviews

Pain Outcomes Reported in Abstracts (N=57 reviews)*			
Outcome Domain/ Subdomain	*N abstracts*	*Outcome Instrument*	*N*
Pain intensity	46	***Unidimensional pain intensity scales:***	
		VAS (0-10 cm or 0-100 mm)	
		0-10 scale	11
		0-20 scale	3
		100 point scale	1
		Verbal rating score (0-10)	3
		10-point Likert scale	1
			1
		Multidimensional pain intensity scales:	
		WOMAC pain subscale score	
		Dichotomous outcomes – instrument not reported	1
		At least 50% reduction in pain	
		Instrument not reported (only SMD reported)	2
		Pain reported in abstract as an outcome, but not measured in included studies	28 (10)
			3
Pain intensity/tender joints	2	Number of tender joints	2
Multi-domain outcomes including pain intensity	13	ACR20 response criteria	1
		ACR50 response criteria	5
		Composite including BASDAI	1
		ASES Shoulder Score	1
		Disease Activity Score (DAS28)	2
		Lequesne Index	1
		SPADI	2
Multi-domain/dimension outcomes including pain	1	DASH	1

intensity and pain interference			
Multi-domain/dimension outcomes including pain interference	0		
Multi-domain/dimension outcomes including pain intensity, pain frequency and pain interference	0		
Pain not reported in abstract	3	Not applicable	

*Note: More than one pain outcome can be reported per abstract; 1 review had no included studies but had an SoF table for the major outcomes. Five abstracts used the term 'pain relief' but the corresponding SoF table used 'pain intensity'. We extracted those outcomes under the domain of pain intensity.

ACR 50=American College of Rheumatology 50% response criteria; ASES=American Shoulder and Elbow Surgeons Shoulder score; DASH=Disabilities of the Arm, Shoulder, and Hand ;VAS=visual analogue scale; WOMAC=Western Ontario and McMaster Universities Arthritis Index; SMD=standardized mean difference; SPADI: Shoulder Pain and Disability Index

Table 5: Pain outcome domains and instruments reported in included studies of Cochrane Reviews

Included studies N=617 studies from 57 reviews*			
Domain	*N*	*Scale*	*N*
Pain intensity	500	***Unidimensional pain scale***	
		VAS (0-10, 0-100, modified)	243
		NRS	32
		VRS	9
		Pain rating scale	4
		Pressure Pain Threshold	2
		Pain free function index	4
		HAQ Pain	15
		Patient Experience Diary 24-h recall pain score	5
		Regional pain score	1
		0-2 scale	3
		0-3 scale	11
		0-4 scale	4
		3-point scale	3
		4-point scale	12
		5-point scale	3
		6-point scale	2
		7-point scale	2
		9-point scale	6
		10-point scale	4
		11-point scale	1
		0-4 scale for PI on activity	4
		Change in perception of PI (5pt scale)	1
		Multidimensional pain intensity scales	
		Hospital for Special Surgery pain score	1
		WOMAC pain	23
		Dichotomous pain intensity(no scale)	
		>30% improvement in pain	9
		> 50% improvement in pain	16
		Pain experienced in rotation (Y/N)	1

		Not reported	79
Pain relief	15	4-point scale	2
		5-point scale	3
		6-point scale	3
		VAS (0-100)	1
		Verbal rating scale (4 & 5 point)	2
		>30% pain relief (scale not given)	1
		>50% pain relief (scale not given)	1
		Not reported	2
Pain intensity/tender joints	70	Tender joint count	59
		Tenderness (dolorimeter)	1
		Ritchie Index	8
		Not reported	2
Pain frequency	10	5-point scale ('never' to 'very often')	1
		4-point scale ('no pain' to '4-7 days')	1
		4-point scale ('never' to 'constant')	1
		Prevalence per week	1
		# days per month	2
		VAS (0-10)	1
		NRS	1
		Not reported	2
Pain duration	1	Time to recurrence	1
Overall/global pain	15	3-point scale	1
		VAS	3
		Not reported	11
Pain behaviour	1	Not reported	1
Pain level (duration x intensity)	1	Not reported	1
Multi-domain outcomes including pain intensity	160	ACR20/50/70	121
		ASES Shoulder score	1
		BASDAI	2
		Constant Shoulder score	5
		Disease Activity Score (DAS20)	20
		Lequense Index	5
		Roles and Maudsley score	1
		UCLA Shoulder rating scale	2
		Whiplash Disability Questionnaire	1

		SPADI	2
Multi-domain/dimension outcomes including pain intensity and pain interference	49	AIMS Brief Pain Inventory CPGS SF-36 Bodily Pain subscale Neck Disability Index Oswestry Disability Index Roland-Morris Disability Index DASH**	1 10 1 9 8 9 10 2
Multi-domain outcomes including pain intensity and pain quality	14	McGill Pain Questionnaire	14
Multi-domain outcomes including pain interference and pain location	1	Nordic Musculoskeletal Questionnaire	1
Multi-domain outcomes including pain intensity, pain interference, pain affect	3	West Haven-Yale Multidimensional Pain Inventory	3
Multi-domain outcomes including pain intensity, pain interference, pain duration	1	Northwick Park Neck Pain Questionnaire (NPQ)	1
Pain interference	3	***Unidimensional pain interference*** 7 point scale Sleep disturbance due to pain Pain interference with work or leisure activities (VAS 0-100)	 1 1 1
Pain extent	1	Pain drawing	1
Pain not reported in included study	153	Not applicable	

* Note: More than one pain outcome can be reported per study; 1 review had no included studies.

**In one review, DASH was reported in the SoF table but the included study reported SPADI

AIMS=Arthritis Impact Measurement Scales; ASES=American Shoulder And Elbow Surgeons; AS=ankylosing spondylitis; BASDAI=Bath Ankylosing Spondylitis Disease

Activity Index; CPGS=Chronic Pain Grading Scale; DASH=Disabilities of the Arm, Shoulder, and Hand; MPI= Multidimensional Pain Inventory; PPT=Pain Pressure Threshold; VAS=visual analogue scale; SMD=standardized mean difference; SPADI=Shoulder Pain and Disability Index; VAS=visual analogue scale; WOMAC=Western Ontario and McMaster Universities Arthritis Index

4.3 Results for Study 2: Survey and interviews of key stakeholders regarding presentation of pain outcomes in Cochrane Summary of Findings tables

Forty-five individuals were invited to participate in a more in-depth discussion via survey and/or telephone interview. Thirty-six completed an interview and/or the survey; 10 completed both an interview and a survey. Therefore, responses were obtained from 36/45 (80%) invited individuals. All 24 interviews were conducted by the same person (LJM). Reasons for non-participation were not obtained but assumed to be availability. Those involved in either the interview or survey included: patients with painful chronic musculoskeletal conditions (n=3); health care practitioners and/or researchers with expertise in outcomes measurement, representing the following fields: rheumatology (n=12), occupational therapy (n=2), physiotherapy (n=1), neurology (n=1), pain management (n=4), pain psychotherapy (n=2), and statisticians and methodologists with expertise in outcomes measurement (n=11). The majority of health care practitioners also conduct outcomes research and thus fulfilled more than one stakeholder role.

Each respondent was active in one or more of the following initiatives: ACTTION/IMMPACT; COMET; COSMIN; Cochrane (the Editorial Unit; Back, Musculoskeletal, Neuromuscular Disorders, and Pain, Palliative and Supportive Care Review Groups and Applicability and Recommendations and Patient Reported Outcomes Methods Groups); OMERACT; or VAPAIN (Validation and Application of a core set of patient-relevant outcome domains to assess the effectiveness of multimodal PAIN therapy) (122).

Tables 6 to 9 describe the four key themes derived from respondents along with examples of issues that were raised. In some areas there was good agreement, such as pain intensity is an important domain to measure in a chronic painful musculoskeletal condition. Twenty-four of 36 respondents raised the issue of the importance of incorporating the patient perspective in these discussions. Analysis of other topics resulted in a range of responses, which were

occasionally contradictory. For example, most respondents agreed that analysis by 'responder' is ideal and when available should be presented. In cases when only means are available, a few respondents noted that the presentation of absolute change (e.g. treatment group improvement by 2 points more than a control group on a scale of 0 to 10) is easily interpretable and a useful way of presenting results, while some respondents felt that these means should not be presented. Explanations for the latter included: the argument that presenting a mean change is not useful because the distribution response is often bimodal making an 'average' change meaningless; or that mean change is not easily interpretable by patients. Others noted the importance of ensuring that treatment groups are similar at baseline to interpret the absolute change. Appendix D provides the completed COnsolidated criteria for Reporting Qualitative studies (COREQ) checklist.

Table 6. Theme 1: Which concepts (or 'domains') of chronic pain should be included as 'core' in Cochrane SoF tables?

Key issues raised by respondents:	Number of respondents (denominator=36)
Pain Intensity (PI) is an important outcome to present in SoF tables for chronic conditions	32
• a direct measure of pain, describes the pain experience; the first issue of communicating with HCP	
• there is clear and consistent evidence that improving pain results in improvements in fatigue depression, health-related quality of life and function, and work	
• existing consensus on this by IMMPACT (PI measured on a 0-10 NRS)	
A one-dimensional measure of PI alone does not capture the complexity of pain impact	10
• "This is to me more important: whether it [pain] stopped me from what I wanted to or needed to do rather than something that was just there. Rating the intensity of the pain might be impacted by whether it is preventing me from doing what I want/need to do" (quote from patient)	
• The best measure for a trial because it has the best sensitivity to change [ie intensity] doesn't necessarily reflect a meaningful improvement in the patient experience	
Consideration of the phrasing and standardization of questions about PI with respect to:	7
• time frame (e.g current, last 24 hours, last month, change from previous time point)	
• type of pain (e.g. average, least, worst)	
• specification of activity (e.g. on movement, on walking, at rest)	
• location (overall or global pain, pain targeted to a joint);	
• recall bias concerns	
Difficulties in capturing and measuring the concept of PI	5
• it is framed by individual experience and tolerance	
• it is a qualitative construct that we are trying to quantify	

Importance of pain frequency	
• is an important outcome to include in an SoF table	5
• 'it depends' on condition , e.g. important to describe for recurrent/periodic/intermittent pain	11
Importance of pain interference with function	
• is an important outcome to include in an SoF table	28
• how does it link or overlap with a measure of function alone?	2
• oversimplification that improving pain improves function	1
Consideration of whether generic or disease-specific pain measures should be reported	
• both	7
• prefer generic ("pain is pain")	4
• prefer condition-specific	3
• depends on the question and goal of the systematic review	5
• generic helps to make comparisons across conditions but a field may prefer to use condition-specific	4
Other pain-related domains for consideration:	15
• Pain duration, pain relief, pain behaviour, pain quality, and the impact of pain on fatigue, activities of daily living, worker productivity, health-related quality of life, sexual activities, impact on partners/caregivers	
• Should consider both the etiology of the pain condition and the nature of the intervention	
Important to include patient perspective in the discussions	24
• Link with existing OMERACT Pain Working group and their discussions on pain domains and key issue: is chronic non-cancer pain a disease in and of itself?	
• Consider OMERACT Filter 2.0 framework	

Legend: PI=pain intensity; HCP=health care practitioner

Table 7. Theme 2: Criteria for acceptable clinimetrics/psychometrics for core endpoints for inclusion in Cochrane SoF Tables

Key issues raised by respondents:	Number of respondents (denominator=36)
Must establish congruent language about measurement properties	5
• Terminology used in various groups is not consistent (e.g. meaning of 'discrimination' differs in OMERACT and COSMIN contexts)	
Need clear distinction between WHAT to measure and HOW to measure	5
• First, what is the most important construct to measure and then to discuss what is the best instrument to measure this construct. • Need outcome instruments with acceptable clinimetric criteria before we can have a discussion on how best to express treatment response	
Consideration of assumptions that instruments like NRS or VAS have underlying operational metrics	3
• Concern of use of non-linear scales to try to quantify a percent improvement and impact on MID/MCID calculations • Suggest attention to use of Rasch methods	
Important to consider the instrument in terms of the intervention	2
• Where you expect to see variation in the scale as a result of the intervention is the place on the scale that needs to be the most sensitive • perhaps different scales might be needed depending on severity of pain and where we expect the intervention to act?	

NRS=numerical rating scale; VAS=visual analogue scale

Table 8. Theme 3: Which 'Threshold of Meaning' should be presented in the SoF table?

Key issues raised by respondents:	Number of respondents (denominator=36)
There should be a presentation of the proportion of people reaching a certain threshold (e.g proportion patients achieving a 50% change from baseline). How to define the threshold?:	26
• MCID	3
• "collective 'minimum important change' can [not] be defended scientifically or logically"	1
• 50% is a very good pain reduction and recommended by IASP	5
• Pain responses tend to be bimodal – good relief or very little –an easy discriminating point is 50%	1
• What patients want is ≥50% pain intensity reduction	5
• 50% is a less realistic target	2
• Interested in empirical data re bimodal response	3
Show results for various thresholds	5
• e.g. 20%, 50%, 70% responders	3
• Report all percentage improvements in cumulative frequency distribution	5
• If concerned about statistical power; might find a statistically significant difference with mean change but not in a responder analysis	4
• Want to determine a reliable way to dichotomize continuous data	3
• Why limit to one way of presentation? Consider offering web-based automatic calculation	1
Concern that a fixed proportion like 50% will bias against those with low/better scores at baseline	2
• unless you have similar baselines, meaning is different	

There should be a presentation of proportion of people achieving a state. e.g. Patient Acceptable State, Low or Minimum State; a state of 'no worse than mild pain (NWTMP)';	17
• Status/state' is much more important than 'change' to a patient	8
• the important question for patients "is your pain at a level now where you can function and do what you want/have to do without the pain being an issue?"	2
• it might be considered the ultimate goal of treatment as in reality NWTMP is what patients want - a manageable point vs not manageable point;	3
• for many people in chronic conditions associated with pain, they will not be completely pain-free	3
• keep magnitude and value separate and focus on clear ways to present the data	1
Suggestions of thresholds for defining a 'state':	
• Magnitude of change: below 4 on 0-10 NRS or less than 3 on 0-10 scale	2
• Based on patient response: can ask patient at end of study if they are in an 'acceptable state'	6

IASP =International Association for the Study of Pain; MCID=minimal clinically important difference

Table 9. Theme 4: Establishing a hierarchy of pain outcome instruments

Key issues raised by respondents:	Number of respondents (denominator=36)
Important for systematic review authors	4
• To reduce bias we need a systematic method to inform which pain outcome instrument to choose when more than one is reported in a trial	
Different methods have been used to develop hierarchies for pain outcome instruments in OA	3
• Methods include expert opinion and responsiveness of pain outcome instruments in OA trials.	
• What other criteria than responsiveness should be considered?	
What is the patient perspective on this hierarchy?	2
• Could use concept mapping approach to get input from patients	
Important to distinguish the hierarchy of constructs from hierarchy of instruments	2

OA=osteoarthritis

This chapter contains text from the two published qualifying manuscripts submitted as per the TRIBE PhD requirements: (1) Maxwell, L. J. et al: Current State of Reporting Pain Outcomes in Cochrane Reviews of Chronic Musculoskeletal Pain Conditions and Considerations for an OMERACT Research Agenda. J Rheumatol 2015;42 (10);1934-1942. All rights reserved; (2) Maxwell L, Singh J. Abatacept for rheumatoid arthritis. The Cochrane Database of Systematic Reviews 2009, Issue 4. Art. No.: CD007277. DOI: 10.1002/14651858.CD007277.pub2

5 DISCUSSION

The result of the systematic review and the two projects described above provides compelling evidence of the disparate use of pain outcomes and underpins the need to establish dialogue between stakeholders in the fields of pain measurement and outcome methodology. Such partnerships can advance the development of guidance on best practices for expressing the pain response to an intervention in a way that is most meaningful to decision makers.

The reporting of pain outcomes in the seven trials included in the abatacept for rheumatoid arthritis Cochrane review (95) demonstrated some challenges that systematic review authors face. Out of the seven included studies, an outcome of pain intensity was not separately reported in two included studies though data on pain was collected during the trials since the American College of Rheumatology (ACR) responder index, which includes pain as one of the individual variables, was a reported outcome in both trial publications. As well, across the trials, different instruments were used to measure pain. In one study, the instrument used for measuring pain was not reported, and a measure of variance was not presented in three studies. This demonstrates how the lack of detailed reporting and use of harmonized outcome instruments makes it difficult to synthesize evidence.

The Summary of Findings table is the hallmark of current Cochrane reviews, and while it is reassuring that all but one (1/57) of the included reviews of interventions in chronic musculoskeletal pain conditions provided an outcome of pain, the results were not presented in a consistent manner. A variety of scales, cut points, and transformations were reported in the domains of pain intensity, pain frequency, and pain interference, making it difficult for the readers of Cochrane reviews to make sense of the evidence across reviews. Different pain conditions were included in this analysis, ranging from inflammatory to degenerative conditions, and as noted by the survey/interview respondents, it is not entirely clear whether chronic painful conditions of different etiologies can be reported similarly. As well, the nature of an intervention may impact the choice of key outcome domains. Future work is needed to explore these issues.

The lack of reporting of the outcome instrument used seriously limits the interpretation of results; this was a concern in almost a third of Summary of Findings tables (18/57) that did not report the outcome instrument. For the domain of pain intensity in 498 included studies, 79 (16%) did not clearly report the outcome instrument, thus making it difficult for

systematic reviewers to incorporate this information. IMMPACT has published consensus recommendations for using a numerical rating scale (NRS) to measure pain intensity in chronic pain trials (32) and to report the proportion of patients who achieve reductions in pain intensity of ≥ 30% and ≥ 50% (reflecting what are proposed as moderate and substantial clinically important differences, respectively) (58). For back pain, one consensus paper suggests a 30% reduction in pain as minimally important to patients (123). In spite of these recommendations, it is notable that the NRS was not reported in a single Summary of Findings table and that visual analogue scale (VAS) use was much more prevalent in the included studies. Responder analyses for pain intensity were reported in only 4% (2/57) of Summary of Findings tables and 5% (25/498) of included studies. The lack of reporting of responder analyses in primary studies limits their use in systematic reviews. However, the timing of the publication of the recommendations (2005 and 2008) may be a reason for the lack of their use as the uptake of this information into clinical trials and subsequent systematic reviews can take time.

A limitation of our study is that we used the 'Characteristics of Included Studies' table to obtain the data on the pain outcome domains and instruments reported in the included studies. We did not obtain the information directly from the included studies due to the feasibility of obtaining full reports of all the 617 included studies. This method relies on the systematic review author to have accurately and fully reported the measured outcomes. In most cases, the 'Characteristics of Included Studies' table provided clear, detailed information and we believe that the data we obtained provides a good representation of the outcomes reported in the included studies.

The between-group difference of pain was reported only as a SMD (without re-expression of the SMD in another format) in six (6/57, 11%) of Summary of Findings tables. Due to the difficulty of interpreting SMDs, the Cochrane Handbook and GRADE recommends re-expression to other formats such as odds ratio or a familiar instrument to facilitate better understanding (50, 54). Review authors should be encouraged to present re-expressed results of SMD, preferably in both absolute and relative terms, as suggested by Johnston et al after their examination of the understanding of different presentations formats to clinicians (55). The judgement of decision makers can be affected when only relative intervention effects, which are numerically larger than absolute effects, are presented. A recent systematic survey of Cochrane and non-Cochrane reviews found that absolute effects were reported much less frequently, and often insufficiently, than relative effects (73/202; 36.1%). Again it was

recommended that to enhance interpretability of findings, both absolute and relative effects should be presented in systematic reviews (124).

The effects from continuous outcomes, compared to dichotomous outcomes, are more challenging to communicate. As discussed in the background section of this thesis, the GRADE Working Group has laid out different options for presenting results from continuous outcomes (50). Recently, a new statistical metric for presenting results from continuous outcomes when a clinically meaningful threshold is available has been proposed. The $NNT_{threshold}$ can be calculated from summary data from meta-analyses which is useful for systematic reviewers. Similar to the 'number needed to treat (NNT) for a dichotomous outcome, $NNT_{threshold}$ describes the number of people who need to be treated in order for one additional person to exceed the clinically meaningful threshold (125). Given that pain is often measured used a continuous scale, it would be useful to investigate further whether using this statistic in Summary of Findings tables improves interpretation.

Seven global pain domains (intensity, frequency, interference, location, affect, quality, and factors associated with pain) were identified when chronic pain patients were asked to describe their pain in their own words (126). An IMMPACT survey of patients with a variety of chronic pain conditions found that within the concept of pain interference, patients identified 19 aspects of pain interference with daily life (e.g. sleep, social relationships, employment, emotional well-being, etc.) as being important (127). We found that the majority of Summary of Findings tables reported on pain intensity (though it was not often explicitly identified that intensity was the domain of interest), with few assessing pain interference or frequency or any other of the pain (sub) domains identified by IMMPACT. A majority of survey and interview participants in this study raised the importance of including the patient perspective and that a more complex measure of the impact of pain, in addition to intensity, should be considered when reporting the evidence for the effectiveness of a treatment for chronic pain. They also noted that, although the burden on respondents must be taken into account, a more complex measure of the effect of pain, in addition to intensity, should be considered when reporting evidence for the effectiveness of a treatment for a chronic painful condition. The majority of included studies assessed only the domain of pain intensity with 8% assessing a measure of pain interference in either a single- or multiple-domain instrument, and very few assessing other domains like pain frequency, quality, location, or affect. A strength of the new Intermittent and Constant Osteoarthritis Pain (ICOAP) instrument (128), which was developed based on focus groups with patients from

four countries and used modern psychometric approaches as recommended in the OMERACT Filter 2.0 (20), is that it captures information about both constant pain and intermittent pain. People with osteoarthritis identified the importance of both types of pain. We look forward to seeing this instrument in the updates of Cochrane osteoarthritis systematic reviews.

In a brief assessment conducted in May 2013, 36 Cochrane Review Groups reported an outcome of pain in at least one of their plain language summaries. Our analysis of the reporting of pain outcomes was limited to Cochrane Reviews from three Cochrane Review Groups and the focus was on chronic musculoskeletal painful conditions; this is rather limited given the wide range of conditions included across Cochrane Reviews. However, a recently published descriptive survey of outcomes reported in Cochrane Reviews (129) confirms our findings in that it also found large variability in the types of outcomes reported within pre-defined outcome categories (e.g. pain, quality of life) as well as a considerable amount (37%) of outcomes which were pre-specified in Cochrane review methods sections but which were not reported in the results section. This further calls for the need for discussion about the use of standardized core outcome sets for Cochrane Summary of Findings tables.

The survey and interviews allowed us to generate themes for future research, based on input from a broad group of stakeholders. Our survey and interview response rate was reasonable (80%) and we were able to obtain participation from the key organizations with expertise in outcome measures methodology and who form necessary partners in our efforts to obtain consensus.

Survey and interview participants generally agreed on some topic areas while others, such as the preferred method for the presentation of results, highlighted differences of opinion. The importance of ensuring the perspective of patients and caregivers is captured in discussions about what core domains should be reported in Summary of Findings tables was raised and should be adhered to in further research projects on developing core outcome sets.

The four themes we generated from the survey and interviews formed the basis of discussions at a pre-conference meeting at OMERACT12 designed to identify the requirements for moving towards consensus on how best to express and measure pain outcomes in intervention trials and resulting systematic reviews for chronic musculoskeletal conditions (130). Forty-two individuals representing key international outcome organizations attended the meeting which was organized around a workshop for each of the following themes: (1) Pain Domains,

(2) Clinimetric considerations, (3) Thresholds for presenting results, and (4) Establishing hierarchies of outcomes. The resulting discussions and proposed research agendas were published in five manuscripts (two separate manuscripts from the Pain Domains workshop) in the Journal of Rheumatology (131-135).

Following the success of the pre-conference meeting, we applied to the Canadian Institutes of Health Research (CIHR; Canada's federal funding agency for health research), for a Knowledge and Dissemination Grant to further the efforts of this project. With the funding received, we held a second meeting at the Cochrane Colloquium in October 2015 to discuss the proposal of establishing a Patient Reported Outcomes (PRO) Alliance within the Cochrane Collaboration. The purpose of this Alliance is to bring together experts in outcome methodology from both within Cochrane (through the Cochrane Editorial Unit, Consumer Network, Review Groups, Methods Groups, Fields, and other Alliances) as well as those from other international outcome organizations such as ACTTION/ IMMPACT, COMET, COSMIN, and the Patient Reported Outcomes Measurement Information System (PROMIS) to develop partnerships, share resources and knowledge, advance research agendas, and provide harmonized guidance to Cochrane authors and others interested in outcomes research. The PRO Alliance was established at the end of the meeting and one of the deliverables is that members will contribute to updating and revising the 'Patient-reported outcomes' chapter of the Cochrane Handbook (21) for discussion at the next Cochrane Colloquium in October 2016.

6 CONCLUSION

The assessment of pain outcomes reported in 57 Cochrane Reviews of interventions for chronic musculoskeletal conditions confirmed our hypothesis that pain outcomes are not reported in a consistent manner. We found a variety of pain outcome domains, outcome instruments, and methods of presentation which hampers efforts to synthesize and interpret the evidence base.

The survey and interviews of stakeholders, including patients, health care providers, and methodologists, identified four key themes which have guided discussions on developing explicit processes for deciding which outcomes, which outcome instruments, and which effect size metrics should be reported in Summary of Findings tables, with the aim of harmonizing the reporting of evidence across Cochrane Reviews.

Through fostering partnerships with key organizations involved in pain and outcome measurement, and building on OMERACT's standard of actively including patients and relevant stakeholders in the consensus process, there is now a strong opportunity to achieve consensus and develop guidance on best practices for reporting pain outcomes, and other patient-reported outcomes, in Cochrane Summary of Findings tables.

7 SUMMARY

Cochrane systematic reviews aim to synthesize all available evidence on a health care intervention in order to provide high-quality, unbiased information to patients, providers, policymakers or other decision makers. A unique feature of a Cochrane Review is a 'Summary of Findings' table. It is a succinct and transparent summary of the magnitude of the effect size and quality of the evidence for each important outcome. The outcomes chosen for presentation in the Summary of Findings table should be ones that are important to patients and demonstrate a balance of benefit and harm.

Pain is a key outcome in studies of interventions designed to improve symptoms of chronic musculoskeletal conditions. We found that in a sample of Cochrane reviews on chronic musculoskeletal conditions, both individual intervention studies and the systematic reviews of these studies report various domains of pain and outcome instruments for measuring those domains. The methods of analysis, thresholds for minimally important and clinically important differences, and presentation format also differed. These differences provide a challenge in both synthesizing the evidence in meta-analyses within systematic reviews and ensuring comparability of results across different systematic reviews.

As OMERACT stated in 1993, clinical trials are "only as credible as their endpoints". Thus, it is imperative that valid, responsive, and feasible outcome measures are utilized in intervention studies. Use of a standardized core set of outcomes in every trial will help ensure comparability across studies. There are various international organizations with expertise in outcome measures methodology. Our survey and interviews of individuals representing these organizations, as well as patients, health care providers, and methodologists, identified four themes: (1) Pain Domains, (2) Clinimetric considerations, (3) Thresholds for presenting results, and (4) Establishing hierarchies of outcomes. Early buy-in from this wide group of stakeholders and eliciting their opinions on key methodological issues will help strengthen the partnerships needed to achieve consensus on best practices for reporting pain outcomes in Cochrane Summary of Findings tables.

8 REFERENCES

1. Committee on Comparative Effectiveness Research Prioritization, Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington: The National Academies Press; 2009; 252 pages

2. Tugwell P, Boers M. OMERACT Conference on Outcome Measures in RA Clinical Trials: Introduction. J Rheumatol 1993;20:528-30.

3. Smith R. The Cochrane Collaboration at 20 [editorial]. BMJ 2013;347.

4. Tovey D. The impact of Cochrane Reviews [editorial]. Cochrane Database Syst Rev. 2010; (7):10.1002/14651858.ED000007

5. Bunn F, Trivedi D, Alderson P, Hamilton L, Martin A, Pinkney E, et al. The impact of Cochrane Reviews: a mixed-methods evaluation of outputs from Cochrane Review Groups supported by the National Institute for Health Research. Health Technol Assess 2015;19(28):1-99, v-vi.

6. Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. Med Decis Making 2010;30:566-77.

7. Rosenbaum SE GC, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. J Clin Epidemiol 2010;63:620-6.

8. Carrasco-Labra A B-PR, Santesso N, Neumann I, Mustafa RA, Mbuagbaw L, et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. J Clin Epidemiol 2016; Jan 11. pii: S0895-4356(15)00583-1 [Epub ahead of print].

9. Guyatt GH, Oxman AD, Akl E, Kunz R, Vist G, Brozek J, et al. GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383-94.

10. Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-26.

11. Guyatt GH Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. Rating quality of evidence and strength of recommendations: What is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995-8.

12. Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic

Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org.

13. Ghogomu E, Maxwell L, Buchbinder R, Rader T, Pardo Pardo J. Editorial Board of the Cochrane Musculoskeletal Group et al. Updated method guidelines for Cochrane musculoskeletal group systematic reviews and metaanalyses J Rheumatol 2014; 41:194-205.

14. Cummings SR KD, Harris F, Genant HK, Ensrud K, LaCroix AZ, et al. Improvement in spine bone density and reduction in risk of vertebral fractures during treatment with antiresorptive drugs. Am J Med 2002;112:281-9.

15. Sarkar S MB, Wong MM, Stock JL, Black DM, Harper KD. Relationships between bone mineral density and incident vertebral fracture risk with raloxifene therapy. J Bone Miner Res 2002;17:1-10.

16. Schünemann H, Brożek J, Guyatt G, Oxman A (editors). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Chapter 5: Quality of evidence. 2013; Available from: http://gdt.guidelinedevelopment.org.

17. US Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. [Monograph on the Internet, Accessed August 12, 2015]. Rockville, MD: Food and Drug Administration, Center for Drug Evaluation and Research (CDER); 2009. Available from: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf

18. Frank L, Basch E, Selby JV, For the Patient-Centered Outcomes Research I. The PCORI perspective on patient-centered outcomes research. JAMA 2014;312:1513-4.

19. Johnston BC, Patrick DL, Busse JW, Schünemann HJ, Agarwal A, Guyatt GH. Patient-reported outcomes in meta-analyses–part 1: assessing risk of bias and combining outcomes. BMC Health Qual Life Outcomes 2013;11.

20. Boers M KJ, Wells G, Beaton D, Gossec L, d'Agostino MA et al. Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. J Clin Epidemiol 2014 67:745-53.

21. Patrick D, Guyatt G, Acquadro C. Chapter 17: Patient-reported outcomes. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org.

22. Kirwan JR HS, Heiberg T, et al. Incorporating the patient perspective into outcome assessment in rheumatoid arthritis — progress at OMERACT 7. J Rheumatol 2005;32:2250-6

23. Kirwan JR M, P, Adebajo A, Bresnihan B, Choy E, de Wit M, Hazes M et al. Patient Perspective: Fatigue as a Recommended Patient Centered Outcome Measure in Rheumatoid Arthritis. J Rheumatol 2007;34:1174-7.

24. IASP Task Force on Taxonomy. IASP Pain Terminology [changes to] Part III: Pain Terms, A Current List with Definitions and Notes on Usage (pp 209-214) in Classification of Chronic Pain. Seattle, WA: IASP Press; 1994. Available from: http://www.iasp-pain.org/Taxonomy.

25. IASP. Unrelieved pain is a major global healthcare problem: fact sheet2004 [Accessed March 21, 2016]. Available from: http://www.iasp-pain.org/GlobalYear/RighttoPainRelief.

26. Torrance N, Elliott AM, Lee AJ, Smith BH. Severe chronic pain is associated with increased 10 year mortality. A cohort record linkage study. Eur J Pain 2010;14:380-6.

27. Mannion AF, Balague F, Pellise F, Cedraschi C. Pain measurement in patients with low back pain. Nat Clin Pract Rheum 2007;3:610-8.

28. Furlan AD SJ, Mailis-Gagnon A, Tunks E. Opioids for chronic noncancer pain: a meta-analysis of effectiveness and side effects. Can Med Assoc J 2006;174:1589-94.

29. Turk D, Melzack R. The Measurement of Pain and the Assessment of People Experiencing Pain. In: Turk D, Melzack R, editors. Handbook of Pain Assessment. 3rd ed. New York, NY: The Guilford Press; 2011. p. 3-13.

30. Hannan M, Felson D, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000;27:1513-7.

31. Jensen M, Karoly P. Self-Report Scales and Procedures for Assessing Pain in Adults. In: Turk D, Melzack R, editors. Handbook of Pain Assessment. 3rd ed. New York, NY: The Guildford Press; 2011. p. 19-44.

32. Dworkin R, Turk D, Farrar J, i sur. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005:9-19.

33. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. Journal of Clinical Nursing 2005;14:798-804.

34. Smith S, Hunsinger M, McKeown A, i sur. Quality of Pain Intensity Assessment Reporting: ACTTION Systematic Review and Recommendations. The Journal of Pain 2015;16:299-305.

35. Price D, Staud R, Robinson M. How should we use the visual analogue scale (VAS) in rehabilitation outcomes? II: Visual analogue scales as ratio scales: An alternative to the view of Kersten et al. J Rehabil Med 2012;44:800-4.

36. Schünemann H, Oxman A, Higgins J, Vist G, Glasziou P, Guyatt G. Chapter 11: Presenting results and ‘Summary of findings' tables. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org.

37. Williamson P, Altman D, Blazeby J, Clarke M, Devane D, Gargon E. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.

38. Busse JW BS, Dougados M, Johnston BC, Guyatt GH, Kirwan JR, Kwoh K, Maxwell LJ et al. Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop. J Rheumatol 2015;42:1962-70.

39. Boers M BP, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.

40. Mokkink LB TC, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported ouctomes. J Clin Epidemiol 2010;63:737-45.

41. Mokkink L, Terwee C, Patrick D, i sur. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research 2010;19:539-49.

42. Dickersin K MY, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. JAMA 1992;263:374-8.

43. Hopewell S LK, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009;1.

44. Dwan K, Gamble C, Kirkham J, Williamson P. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias ? An Updated Review . PLoS One 2013; 8:e66844.

45. Kirkham JJ DK, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010;340:c365.

46. Kirkham JJ AD, Williamson PR. Bias Due to Changes in Specified Outcomes during the Systematic Review Process. PLoS ONE 2010;5:e9810.

47. Chalmers I, Glasziou P, Godlee F. All trials must be registered and the results published. BMJ 2013;346:f105.

48. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials 2007;8:39.

49. Kirkham JJ GE, Clarke M, Williamson PR Can a core outcome set improve the quality of systematic reviews? – a survey of the Co-ordinating Editors of Cochrane review groups. Trials 2013;14:21.

50. Guyatt GH, Thorlund K, Oxman AD, i sur. GRADE guidelines: 13. Preparing summary of findings tables: continuous outcomes. J Clin Epidemiol 2013;66:173-83.

51. Johnston BC, Patrick DL, Thorlund K, i sur. Patient-reported outcomes in meta-analyses -Part 2: methods for improving interpretability for decision-makers. BMC Health Qual Life Outcomes 2013;11:1-9.

52. Moore RA, Straube S, Aldington D. Pain measures and cut-offs – 'no worse than mild pain' as a simple, universal outcome. Anaesthesia 2013;68:400-12.

53. Cohen J. Statistical Power Analysis in the Behavioral Sciences. Hillsdale (NJ): Lawrence Erlbaum Associates, Inc; 1988.

54. Schünemann H, Oxman A, Vist G (editors). Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org.

55. Johnston BC A-CP, Friedrich JO, Mustafa RA, Tikkinen KAO, Neumann I et al. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. Can Med Assoc J 2016;188:25-32.

56. Deeks JJ HJ, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org..

57. Dworkin R, Turk D, McDermott M, i sur. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009;146:238-44.

58. Dworkin R, Turk D, Wyrwich K, Beaton D, Cleeland C, Farrar J. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations J Pain 2008;9:105-21.

59. Mease P AL, Choy EH, Clauw DJ, Crofford LJ, Glass JM, et al. Fibromyalgia Syndrome Module at OMERACT 9: Domain Construct. J Rheumatol 2009;36:2318-29.

60. Furlan A, Pennick V, Bombardier C, van Tulder M. Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group. Spine 2009;34:1929–41.

61. Maxwell L, Santesso N, Tugwell P, i sur. Method guidelines for Cochrane Musculoskeletal Group systematic reviews. J Rheumatol 2006;33:2304-11.

62. Moore R, Eccleston C, Derry S, i sur. "Evidence" in chronic pain – establishing best practice in the reporting of systematic reviews. Pain 2010;150:386-9.

63. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. Trials 2007 8:38.

64. Boers M, Tugwell P, Felson D, van Riel P, Kirwan J, Edmonds J. World Health Organization and International League of Associations for Rheumatology Core Endpoints for Symptom Modifying Antirheumatic Drugs in RA Clinical Trials J Rheumatol 1994;20:86-9.

65. Bellamy N, Kirwan JR, Boers M, Brooks, P, Strand V, Tugwell P. Recommendations for a Core Set of Outcome Measures for Future Phase III Clinical Trials in Knee, Hip, and Hand Osteoarthritis.Consensus Development at OMERACT III. J Rheumatol 1997;24:799-802.

66. Sambrook P, Cummings S, Eisman J, Genant H, Eastell R, Reginster J. Guidelines on osteoporosis trials. J Rheumatol 1997;24:1234-36.

67. van der Heijde D, van der Linden S, Bellamy N, Calin A, Dougados M, Khan M. Which Domains Should Be Included in a Core Set for Endpoints in Ankylosing Spondylitis? Introduction to the Ankylosing Spondylitis Module of OMERACT IV. J Rheumatol 1999;26:946-7.

68. Gladman D, PJ M, Strand V, i sur. Consensus on a Core Set of Domains for Psoriatic Arthritis. J Rheumatol 2007;34:1167-70.

69. Kirkham JJ, Boers M, Tugwell P, Clarke M, Williamson P. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. Trials 2013;14:324.

70. Turk D, Dworkin R, Allen R, i sur. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106:337-45.

71. (ACTTION) Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks: Home Page. [Accessed March 21, 2016]; Available from: http://www.acttion.org/.

72. Gargon E, Gurung B, Medley N, i sur. Choosing Important Health Outcomes for Comparative Effectiveness Research: A Systematic Review. PLoS One 2014;9:e99111.

73. Mokkink L, Terwee C, Knol D, i sur. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. . BMC Med Res Methodol 2010;10:22.

74. Terwee C, Mokkink L, Knol D, Ostelo R, Bouter L, de Vet H. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. . Qual Life Res 2012;21:651–7.

75. Schumacher Jr H. Rheumatoid Arthritis. In: Schumacher Jr H, Klippel J, Koopman W, editors. Primer on the Rheumatic Diseases. 10th ed. Atlanta, GA: Arthritis Foundation; 1993.

76. Badley E, DesMeules M. Arthritis in Canada. An ongoing challenge. Health Canada. Ottawa 2003.

77. Woolf A, Åkesson K, Compston J. European action towards better musculoskeletal health - a public health strategy to reduce the burden of musculoskeletal conditions. Lund, Sweden: Department of Orthopedics, University Hospital; 2004. Available from: http://ec.europa.eu/health/ph_projects/2000/promotion/fp_promotion_2000_frep_15_en.pdf.

78. National Audit Office. Services for People with Rheumatoid Arthritis: HC 823, Report by the Comptroller and Auditor General, Session 2008-2009. UK: TSO (The Stationery Office); 2009.

79. Helmick CG, Felson DT, Lawrence RC, i sur. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. Arthritis & Rheumatism 2008;58:15-25.

80. Birnbaum H, Pike C, Kaufman R, Maynchenko M, Kidolezi Y, Cifaldi M. Societal cost of rheumatoid arthritis patients in the US. Cur Med Res Opin 2010;26:77-90.

81. Lopez-Olivo MA, Siddhanamatha HR, Shea B, Tugwell P, Wells GA, Suarez-Almazor ME. Methotrexate for treating rheumatoid arthritis. Cochrane Database Syst Rev 2014 Jun 10;6:CD000957

82. Osiri M, Shea B, Welch V, i sur. Leflunomide for the treatment of rheumatoid arthritis. Cochrane Database Syst Rev. 2003;(1):CD002047.

83. Suarez-Almazor ME, Belseck E, Shea B, Homik J, Wells GA, Tugwell P. Antimalarials for treating rheumatoid arthritis. Cochrane Database Syst Rev. 2000;(4):CD000959

84. Suarez-Almazor ME, Belseck E, Shea B, Tugwell P, Wells GA. Sulfasalazine for treating rheumatoid arthritis. Cochrane Database Syst Rev. 2000;(2):CD000958

85. Kirwan JR, Bijlsma JWJ, Boers M, Shea B. Effects of glucocorticoids on radiological progression in rheumatoid arthritis. Cochrane Database Syst Rev. 2007 Jan 24;(1):CD006356

86. Lethaby A, Lopez-Olivo MA, Maxwell LJ, Burls A, Tugwell P, Wells GA. Etanercept for the treatment of rheumatoid arthritis. Cochrane Database Syst Rev. 2013 May 31;5:CD004525

87. Navarro-Sarabia F, Ariza-Ariza R, Hernandez-Cruz B, Villanueva I. Adalimumab for treating rheumatoid arthritis. Cochrane Database Syst Rev. 2005 Jul 20;(3):CD005113

88. Blumenauer BBTB, Judd M, Wells GA, i sur. Infliximab for the treatment of rheumatoid arthritis. Cochrane Database Syst Rev. 2002;(3):CD003785

89. Ruiz Garcia V, Jobanputra P, Burls A, i sur. Certolizumab pegol (CDP870) for rheumatoid arthritis in adults. Cochrane Database Syst Rev. 2014;9:CD007649.

90. Singh JA, Noorbaloochi S, Singh G. Golimumab for rheumatoid arthritis. Cochrane Database Syst Rev. 2010 Jan 20;(1):CD008341

91. Lopez-Olivo MA, Amezaga Urruela M, McGahan L, Pollono EN, Suarez-Almazor ME. Rituximab for rheumatoid arthritis. Cochrane Database Syst Rev. 2015 Jan 20;1:CD007356

92. Singh JA, Beg S, Lopez-Olivo MA. Tocilizumab for rheumatoid arthritis. Cochrane Database Syst Rev. 2010 Jul 7;(7):CD008331

93. Bristol-Myers Squibb. Highlights of prescribing information: Orencia. Princeton, NJ: 2015 Report No.: 1292618A5 / 1294018A4.

94. Patented Medicine Prices Review Board, Canada. Report on New Patented Drugs - Orencia. 2007. Available from: http://www.pmprb-cepmb.gc.ca/CMFiles/Regulating%20Prices/Price%20Review/New_Pat_Meds_Human/Orencia-e_Oct_07__New_38IYX-11162007-2541.pdf

95. Maxwell L, Singh JA. Abatacept for rheumatoid arthritis. Cochrane Database Syst Rev. 2009 Oct 7;(4):CD007277

96. Arnett FC, Edworthy SM, Bloch DA, i sur. The American rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.

97. Felson DT, Anderson JJ, Boers M, i sur. American college of rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.

98. van Gestel AM, Prevoo MLL, van't Hof MA, van Rijswijk MH, van de Putte LBA, van Riel PLCM. Development and validation of the European league against rheumatism response criteria for rheumatoid arthritis: Comparison with the preliminary American college of rheumatology and the world health organization/international league against rheumatism criteria. Arthritis Rheum 1996;39:34-40.

99. van der Heijde D, van Hof M, van Riel P, van de Putte L. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol 1993;20:579-81.

100. DAS-Score.nl. Home of the DAS. Disease Activity Score in rheumatoid arthritis. [Accessed 24 July 2009]; Available from: http://www.das-score.nl/www.das-score.nl.

101. Higgins JPT, Green S (editors) Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration; 2008. Available from: www.cochrane-handbook.org.

102. The Nordic Cochrane Centre. Review Manager (RevMan) version 5.0. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2008.

103. Kremer JM, Genant HK, Moreland LW, i sur. Effects of Abatacept in Patients with Methotrexate-Resistant Active Rheumatoid Arthritis-A Randomized Trial. Ann Intern Med 2006;144:865-76.

104. Kremer JM, Westhovens R, Leon M, i sur. Treatment of Rheumatoid Arthritis by Selective Inhibition of T-Cell Activation with Fusion Protein CTLA4Ig. N Engl J Med 2003;349:1907-15.

105. Higgins J, Thompson S, Deeks J, Altman D. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.

106. Cates C. Visual Rx 2.0 NNT Calculator. Dr Chris Cates EBM Website; 2004.

107. Tubach F, Ravaud P, Beaton D, i sur. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188-93.

108. Farrar JT, Young JPJ, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149-58.

109. Dieppe P. Practice. ABC of Pain. Chronic Musculoskeletal Pain. BMJ 2013;346:bmj.f3146.

110. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15:1277-88.

111. Genovese MC, Becker J-C, Schiff M, i sur. Abatacept for Rheumatoid Arthritis Refractory to Tumor Necrosis Factor α Inhibition. N Engl J Med 2005;353:1114-23.

112. Moreland LW, Alten R, Van Den Bosch F, i sur. Costimulatory blockade in patients with rheumatoid arthritis: A pilot, dose-finding, double-blind, placebo-controlled clinical trial evaluating CTLA-4Ig and LEA29Y eighty-five days after the first infusion. Arthritis Rheum 2002;46:1470-9.

113. Weinblatt M, Combe B, Covucci A, Aranda R, Becker JC, Keystone E. Safety of the selective costimulation modulator abatacept in rheumatoid arthritis patients receiving background biologic and nonbiologic disease-modifying antirheumatic drugs: A one-year randomized, placebo-controlled study. Arthritis Rheum 2006;54:2807-16.

114. Weinblatt M, Schiff M, Goldman A, i sur. Selective costimulation modulation using abatacept in patients with active rheumatoid arthritis while receiving etanercept: a randomised clinical trial. Ann Rheum Dis 2007;66:228-34.

115. Schiff M, Keiserman M, Codding C, i sur. Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. Ann Rheum Dis 2008;67:1096-103.

116. Emery P, Kosinski M, Li T, i sur. Treatment of rheumatoid arthritis patients with abatacept and methotrexate significantly improved health-related quality of life. J Rheumatol 2006;33:681-9.

117. Kremer JM, Dougados M, Emery P, i sur. Treatment of rheumatoid arthritis with the selective costimulation modulator abatacept: Twelve-month results of a phase iib, double-blind, randomized, placebo-controlled trial. Arthritis & Rheumatism 2005;52:2263-71.

118. Russell AS, Wallenstein GV, Li T, i sur. Abatacept improves both the physical and mental health of patients with rheumatoid arthritis who have inadequate response to methotrexate treatment. Ann Rheum Dis 2007;66:189-94.

119. Westhovens R, Cole JC, Li T, i sur. Improved health-related quality of life for rheumatoid arthritis patients treated with abatacept who have inadequate response to anti-TNF therapy in a double-blind, placebo-controlled, multicentre randomized clinical trial. Rheumatology 2006;45:1238-46.

120. Cole JC, Li T, Lin P, MacLean R, Wallenstein GV. Treatment impact on estimated medical expenditure and job loss likelihood in rheumatoid arthritis: re-examining quality of life outcomes from a randomized placebo-controlled clinical trial with abatacept. Rheumatology 2008;47:1044-50.

121. Maxwell LJ, Wells GA, Simon LS, i sur. Current State of Reporting Pain Outcomes in Cochrane Reviews of Chronic Musculoskeletal Pain Conditions and Considerations for an OMERACT Research Agenda. J Rheumatol 2015;42:1934-42.

122. Deckert S, Kaiser U, Kopkow C, Trautmann F, Sabatowski R, Schmitt J. A systematic review of the outcomes reported in multimodal pain therapy for chronic pain. Eur J Pain 2016;20:51-63.

123. Ostelo R, Deyo RA, Stratford P, Waddell G, Croft P, von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90-4.

124. Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, i sur. Systematic reviews experience major limitations in reporting absolute effects. J Clin Epidemiol 2016;72:16-26.

125. Valentine JC, Aloe AM. How to communicate effect sizes for continuous outcomes: a review of existing options and introducing a new metric. J Clin Epidemiol 2016;72:84-9.

126. Jensen M, Johnson L, Gertz K, BS G, Gammaitoni A. The words patients use to describe chronic pain: Implications for measuring pain quality. Pain 2013;154:2722-8.

127. Turk D, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: An IMMPACT survey of people with pain. Pain 2008;137:276-85.

128. Hawker G, Davis A, French M, Cibere J, Jordon J, March L. Development and preliminary psychometric testing of a new OA pain measure – an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16.

129. Smith V CM, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. J Clin Epidemiol 2015;68:237-45.

130. Tugwell PS, Maxwell LJ, Beaton DE, i sur. Dialogue on Developing Consensus on Measurement and Presentation of Patient-important Outcomes, Using Pain Outcomes as an Exemplar, in Systematic Reviews: A Preconference Meeting at OMERACT 12. J Rheumatol 2015;42:1931-3.

131. Phillips K, Taylor A, Mease PJ, i sur. Harmonizing Pain Outcome Measures: Results of the Pre-OMERACT Meeting on Partnerships for Consensus on Patient-important Pain Outcome Domains Between the Cochrane Musculoskeletal Group and OMERACT. J Rheumatol 2015;42:1943-6.

132. Taylor AM, Phillips K, Taylor JO, i sur. Is Chronic Pain a Disease in Its Own Right? Discussions from a Pre-OMERACT 2014 Workshop on Chronic Pain. J Rheumatol 2015;42:1947-53.

133. Busse JW, Bartlett SJ, Dougados M, i sur. Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop. J Rheumatol 2015;42:1962-70.

134. Christensen R, Maxwell LJ, Jüni P, i sur. Consensus on the Need for a Hierarchical List of Patient-reported Pain Outcomes for Metaanalyses of Knee Osteoarthritis Trials: An OMERACT Objective. J Rheumatol 2015;42:1971-5.

135. Beaton DE, Terwee CB, Singh JA, i sur. A Call for Evidence-based Decision Making When Selecting Outcome Measurement Instruments for Summary of Findings Tables in Systematic Reviews: Results from an OMERACT Working Group. J Rheumatol 2015;42:1954-61.

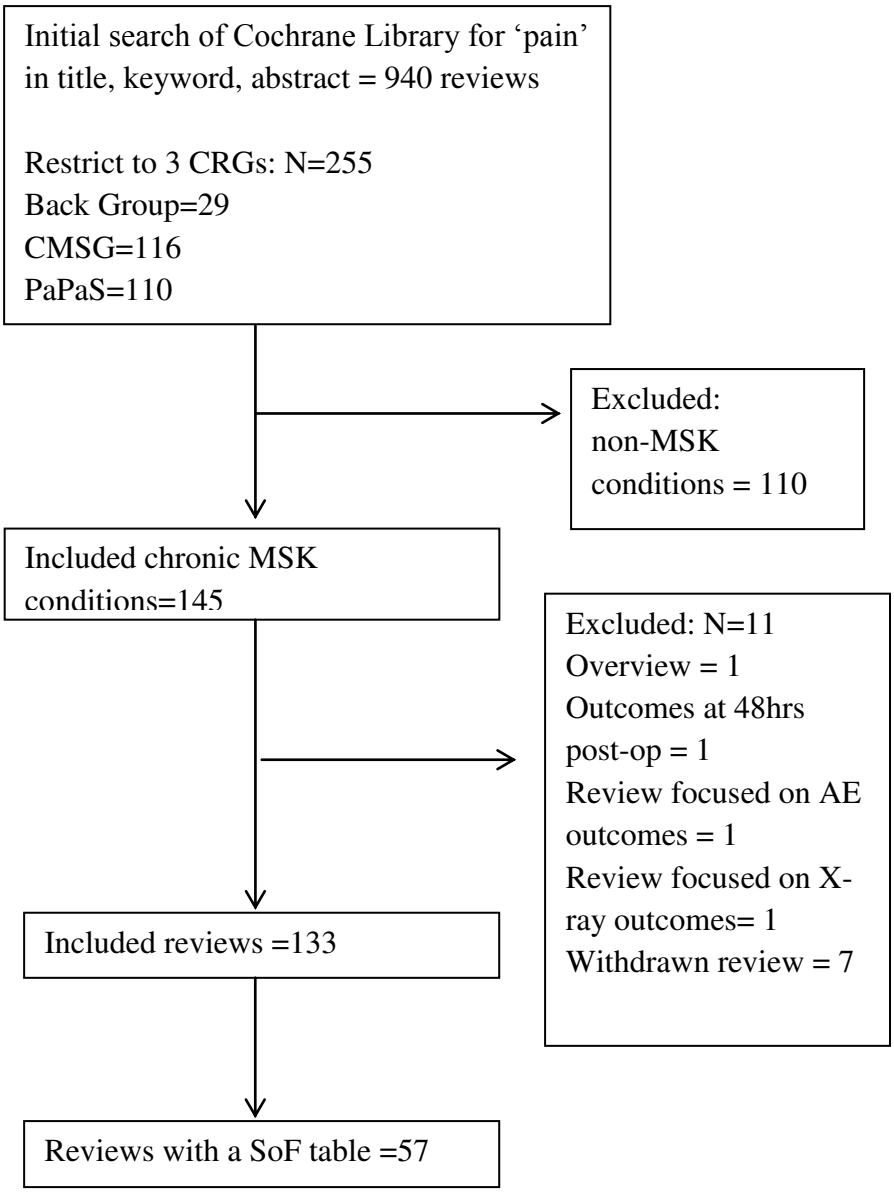
9 APPENDICES

9.1 Appendix A: Survey questions

In our initial thinking about which domains of pain should be measured, we came up with the following suggestions: pain intensity, pain frequency, and pain interference with function. In the following questions, we will ask you to provide comments on each one separately and to add any other pain domains of interest to you that are not on this list. Please comment on whether information on pain intensity should be presented in a Cochrane 'Summary of findings' table and discussed further.
Please comment on whether information on pain frequency should be presented in a Cochrane 'Summary of findings' table and discussed further.
Please comment on whether information on pain interference with function should be presented in a Cochrane 'Summary of findings' table and discussed further.
Please comment on any other pain domains that you feel should be discussed.
In our initial thinking about potential topics for discussion on the question, which measures of change (or 'discrimination') should be presented, we came up with the following suggestions: individual patient response, definitions of 'state', and between-group differences. In the following questions, we will ask you to provide comments on each one separately and to add any other topics about discrimination that are of interest to you. Please comment on whether you think information on the following should be presented in a Cochrane 'Summary of findings' and discussed further: Patient response: analyses and decision on a definition of thresholds for minimum and major clinical importance for change (e.g. 50% improvement in pain intensity from baseline as measured on a NRS)
Please comment on whether you think information on the following should be presented in a Cochrane 'Summary of findings' table and discussed further: Definition of thresholds for minimum and major clinical importance for state (Patient Acceptable State, Low or Minimum State; e.g. achieving a state of 'no worse than mild pain')
Please comment on whether you think information on the following should be presented in a Cochrane 'Summary of findings' table and discussed further: Group difference: analyses and decision on a definition of thresholds for minimum and major clinical importance for change between treatment and control groups (e.g. a 20%? or 30%? difference in mean pain scores between groups)
Please comment on any other important topics about 'discrimination' that you feel should be discussed.
Please comment on whether you feel it is important that a responder index with pain as one of multiple domains be included in a Cochrane 'Summary of findings' table? (e.g. OMERACT-OARSI responder index).

In our initial thinking about important topics to discuss under 'feasibility of measurement', we came up with the following suggestions: sensibility, respondent burden, monetary and other costs, and interpretability of results . In the following questions, we will ask you to provide comments on each one separately and to add any other items about 'feasibility of measurement' of interest to you that are not on this list. Please comment on whether you feel that 'sensibility' (comprehensiveness, understandability, length, and suitability of response options) is an important topic to consider about an outcome presented in a Cochrane 'Summary of findings' table.
Please comment on whether you feel that 'respondent burden' is an important topic to consider about an outcome presented in a Cochrane 'Summary of findings' table.
Please comment on whether you feel that 'monetary and other costs' is an important topic to consider about an outcome presented in a Cochrane 'Summary of findings' table.
Please comment on whether you feel that 'interpretability of results' is an important topic to consider about an outcome presented in a Cochrane 'Summary of findings' table.
Please comment on any other important topics about 'feasibility of measurement' that you feel should be discussed.
Please comment on whether you feel it is important to present generic pain measures, condition-specific pain measures, or both, in a Cochrane 'Summary of findings' table?
We propose using osteoarthritis as the example for our discussions. Are there other conditions you feel important to include in the discussions?
Which aspects of domains of pain, discrimination, and feasibility do you feel are the most important to achieve consensus on?
Please suggest which methods you feel are best to achieve consensus; i.e. Delphi, surveys, face-to-face meetings, background papers for workshops, others?

9.2 Appendix B: Flow chart for selecting included Cochrane Reviews



Legend: CMSG=Cochrane Musculoskeletal Group; CRG=Cochrane Review Group; MSK=musculoskeletal; PaPaS=Pain, Palliative, and Supportive Care; SoF=Summary of Findings

9.3 Appendix C: List of included interventions and disease conditions in Cochrane review assessment

1	Abatacept for rheumatoid arthritis
2	Acupuncture for treating fibromyalgia
3	Antidepressants for pain management in rheumatoid arthritis
4	Arthroplasty versus fusion in single-level cervical degenerative disc disease
5	Assistive technology for rheumatoid arthritis
6	Balance training (proprioceptive training) for patients with rheumatoid arthritis
7	Behavioural treatment for chronic low-back pain
8	Botulinum toxin for shoulder pain
9	Botulinum toxin for subacute/chronic neck pain
10	Cemented, cementless or hybrid fixation options in total knee arthroplasty for osteoarthritis and other non-traumatic diseases
11	Certolizumab pegol (CDP870) for rheumatoid arthritis in adults
12	Continuous passive motion following total knee arthroplasty in people with arthritis
13	Diacerein for osteoarthritis
14	Doxycycline for osteoarthritis of the knee or hip
15	Dynamic exercise programs (aerobic capacity and/or muscle strength training) in patients with rheumatoid arthritis
16	Electrotherapy for neck pain
17	Etanercept for the treatment of rheumatoid arthritis
18	Exercise for improving outcomes after osteoporotic vertebral fracture
19	Exercises for adolescent idiopathic scoliosis
20	Exercises for mechanical neck disorders
21	Febuxostat for treating chronic gout
22	Folic acid and folinic acid for reducing side effects in patients receiving methotrexate for rheumatoid arthritis
23	Gabapentin for chronic neuropathic pain and fibromyalgia in adults
24	Glucosamine therapy for treating osteoarthritis
25	Golimumab for rheumatoid arthritis
26	Herbal therapy for treating rheumatoid arthritis
27	Image-guided versus blind glucocorticoid injection for shoulder pain
28	Interventions for treating osteoarthritis of the big toe joint
29	Joint lavage for osteoarthritis of the knee
30	Leflunomide for the treatment of rheumatoid arthritis
31	Lifestyle interventions for chronic gout
32	Methotrexate for ankylosing spondylitis
33	Methotrexate monotherapy versus methotrexate combination therapy with non-biologic disease modifying anti-rheumatic drugs for rheumatoid arthritis
34	Monoamine oxidase inhibitors (MAOIs) for fibromyalgia syndrome
35	Muscle relaxants for pain management in rheumatoid arthritis
36	Neuromodulators for pain management in rheumatoid arthritis
37	Non-invasive brain stimulation techniques for chronic pain

38	Non-steroidal anti-inflammatory drugs (NSAIDs) for treating lateral elbow pain in adults
39	Non-surgical interventions for paediatric pes planus
40	Opioid therapy for treating rheumatoid arthritis pain
41	Oral or transdermal opioids for osteoarthritis of the knee or hip
42	Psychological therapies for the management of chronic and recurrent pain in children and adolescents
43	S-Adenosylmethionine for osteoarthritis of the knee or hip
44	Serotonin and noradrenaline reuptake inhibitors (SNRIs) for fibromyalgia syndrome
45	Spinal manipulative therapy for chronic low-back pain
46	Stretch for the treatment and prevention of contractures
47	Surface neuromuscular electrical stimulation for quadriceps strengthening pre and post total knee replacement
48	Surgery for lateral elbow pain
49	Surgery for shoulder osteoarthritis
50	Surgery for thumb (trapeziometacarpal joint) osteoarthritis
51	Surgical interventions for the rheumatoid shoulder
52	Therapeutic ultrasound for osteoarthritis of the knee or hip
53	Tocilizumab for rheumatoid arthritis
54	Topical glyceryl trinitrate for rotator cuff disease
55	Topical herbal therapies for treating osteoarthritis
56	Transcutaneous electrostimulation for osteoarthritis of the knee
57	Workplace interventions for neck pain in workers

9.4 Appendix D: Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist

Developed from:
Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Completed for Study 2: Survey and interviews of key stakeholders regarding presentation of pain outcomes in Cochrane Summary of Findings tables (Maxwell et al. J Rheumatol 2015;42 (10);1934-1942)

No. Item	Guide questions/description	Reported on Page #
Domain 1: Research team and reflexivity		
Personal Characteristics		
1. Inter viewer/facilitator	Which author/s conducted the interview or focus group?	1936
2. Credentials	What were the researcher's credentials? E.g. PhD, MD	1936 The survey was drafted by 2 researchers (MD/MSc: rheumatologist/journal editor/systematic reviewer and senior outcomes researcher; MSc: managing editor and systematic reviewer)
3. Occupation	What was their occupation at the time of the study?	1936 The survey was drafted by 2 researchers (MD/MSc: rheumatologist/journal editor/systematic reviewer and senior outcomes researcher; MSc: managing editor and systematic reviewer)
4. Gender	Was the researcher male or female?	Female interviewer
5. Experience and training	What experience or training did the researcher have?	1936 The interviews were conducted by a researcher trained in

		conducting semi-structured interviews.
Relationship with participants		
6. Relationship established	Was a relationship established prior to study commencement?	Interviewer had some previous professional involvement with 11 of the interviewees
7. Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	Knew that it was part of PhD and that interviewer worked for the Cochrane Collaboration
8. Interviewer characteristics	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	Knew that it was part of PhD and that interviewer worked for the Cochrane Collaboration
Domain 2: study design		
Theoretical framework		
9. Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	1936 "directed approach of qualitative content analysis"
Participant selection		
10. Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	1936 'A purposive, expert sampling technique…"
11. Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	1936 "Prospective participants were sent a link to the survey via email"
12. Sample size	How many participants were in the study?	1936 Forty-five people were invited to participate in a survey and interview. Thirty-six completed an interview and/or a survey; 10 people completed both an interview and a

		survey. Therefore, responses were obtained from 36/45 (80%) invited individuals."
13. Non-participation	How many people refused to participate or dropped out? Reasons?	1937 Therefore, responses were obtained from 36/45 (80%) invited individuals." " Reasons for non-participation were not obtained but assumed to be availability."
Setting		
14. Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	1936 Email survey and telephone interviews; did not record where participants were located
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	One-on-one telephone interview
16. Description of sample	What are the important characteristics of the sample? e.g. demographic data, date	1937-8; organizations represented; interviewee occupations
Data collection		
17. Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	1936 and Appendix A
18. Repeat interviews	Were repeat inter views carried out? If yes, how many?	no
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	1936
20. Field notes	Were field notes made during and/or after the inter view or focus group?	1936
21. Duration	What was the duration of the inter views or focus group?	1936
22. Data saturation	Was data saturation discussed?	Our goal was to "obtain representation from Cochrane Review Groups, international

		initiatives involved in outcome measures methodology, patients with painful musculoskeletal conditions, health care practitioners, and methodologists." As we had an 80% response rate, we felt we had reached a reasonable response rate. We didn't discuss data saturation specifically in the manuscript but as we identified both areas of agreement and differences in opinion, we feel we have adequate saturation.
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No
Domain 3: analysis and findings		
Data analysis		
24. Number of data coders	How many data coders coded the data?	1936; The text data were analyzed by one researcher (LJM) and checked by a second (PT) for themes using a directed approach of qualitative content analysis
25. Description of the coding tree	Did authors provide a description of the coding tree?	1936; Interpretation of the results started from the responses to these ideas in the survey/interview guide and then ascertained and counted new topics

		identified by the participants
26. Derivation of themes	Were themes identified in advance or derived from the data?	1936; This directed approach was used as the survey was developed around existing ideas for themes as identified by OMERACT Executive Committee members.
27. Software	What software, if applicable, was used to manage the data?	Excel
28. Participant checking	Did participants provide feedback on the findings?	No
Reporting		
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Tables 2 to 5; quotes used but participants not identified
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Relationship to existing knowledge – page Discussion 1939-40
31. Clarity of major themes	Were major themes clearly presented in the findings?	Results – Tables 2 - 5
32. Clarity of minor themes	Is there a description of **diverse** cases or discussion of minor themes?	Results p.1939 Analysis of other topics resulted in a range of responses, which were occasionally contradictory… Discussion – page 15 Participants were generally agreed in some topic areas while others, such as methods for the presentation of results, highlighted differences of opinion.

10 RESUME

PIVLJANINA BAJA 19, BELGRADE, SERBIA
PHONE: (063) 623-185 E-MAIL: LMAXWELL@UOTTAWA.CA

LARA MAXWELL

WORK EXPERIENCE

2004 – Present University of Ottawa Ottawa, ON
(note: since August 2012 I have been working remotely from Belgrade, Serbia)

Co-Managing Editor, Cochrane Musculoskeletal Group (2004 to present)

The Cochrane Collaboration is an international network of individuals and organizations committed to preparing, maintaining and disseminating systematic reviews of health care interventions. Its purpose is to help patients, clinicians, and other interested parties make well-informed decisions about health care. The Cochrane Musculoskeletal Group (CMSG) focuses specifically on reviews of interventions for musculoskeletal conditions. As the Managing Editor my responsibilities include:

- Coordinating the development of systematic reviews from the initial receipt of a proposed title for a review to the publication of the completed review in the Cochrane Library; providing advice on systematic review methodology to internationally-based review authors; editing reviews; liaising with other Cochrane entities, researchers, and editors throughout the peer review process.
- Arranging and orchestrating editorial meetings; corresponding with review teams; answering queries from prospective members of the public and other interested parties about the Musculoskeletal Group. Supervising the Assistant Managing Editor who provides support for these tasks.
- Working with a specialist on the development and implementation of a research/knowledge transfer strategy for Musculoskeletal Group reviews, including writing articles, preparing newsletters and other resource materials for both review authors and users of musculoskeletal reviews. Preparing and delivering presentations and workshops on systematic review methodology and use of reviews.
- Contributing to methodological research on: advancing the quality and user-friendliness of systematic reviews and overviews of systematic reviews, developing validated outcome measures in rheumatology and incorporating equity considerations into systematic reviews by organizing and participating in meetings and writing articles.
- Preparing grant applications and internal and external budgets, completing funding reports, progress reports and Cochrane Collaboration monitoring reports.

Coordinator, Cochrane Equity Field (2004 – 2006)

- Prepared application for approval of the Field from the Cochrane Collaboration with input from international equity experts. Developed list of international members, wrote newsletters, liaised with other Cochrane groups to promote equity-relevant issues, and prepared presentations.

Coordinator, OMERACT Secretariat (2004 – 2006)

- Arranged OMERACT teleconferences and meetings; prepared minutes for the executive committee; assisted OMERACT members with manuscript submissions; organized OMERACT8 – major biennial conference held in Malta in May 2006

2002 - 2003 Ministry of Health & Social Security Commonwealth of Dominica

Pan American Health Organization Consultancy *April 2003 – June 2003*

- Managed the development and implementation of hospital information system plan.
- Organized a workshop of local and international experts to determine a national minimum data set of indicators on ageing to be used for determining the national health status of older persons and resource planning within the Ministry of Health.

2002 - 2003 Canadian Society for International Health Internship
Commonwealth of Dominica

Asst. Health Information Officer, Ministry of Health *Nov. 2002 – Mar. 2003*

- Chaired the Information Systems Steering Team, responsible for upgrading the Princess Margaret Hospital information system. The hospital is the only tertiary care facility on the island.
- Managed all project plan requirements including research and creation of a detailed budget, project task list, implementation timeframe, risk analysis, identification of a software vendor, hardware technical specifications, and effective procurement methods.
- Provided computer training to hospital staff for the successful implementation of the information system project; organized computer maintenance and networking fundamentals workshops

Assistant, Dominica Council on Ageing *November 2002 – March 2003*

- Research assistant in a Help-Age International study on the Situation of Older Persons in Dominica. Developed and established a website to improve communications between the Council, its member organizations and the international community; trained staff in FrontPage 2000.

2000-2002 EDS Canada Ottawa, ON

Team Lead *August 2001 – November 2002*

- Supervised a team of 20 technical analysts. Responsible for distributing and monitoring work assignments, addressing problems and providing resolutions to management, and administering performance reviews.
- Responsible for ensuring the smooth transition of business operations of 150 client businesses from the US to Ottawa. Developed a successful employee coaching program to improve customer service and team morale.

Technical Analyst *June 2000 – August 2001*

- Used strong problem-solving and analytical techniques to perform technical troubleshooting in a fast-paced, high-pressure environment. Provided support to a client base of 12,000 on application, operating system, hardware and connectivity problems.

EDUCATION

2013-current **University of Split** **Split, Croatia**

- PhD candidate, Translational Research in Biomedicine [TRIBE] Program

2006 - 2011 **University of Ottawa** **Ottawa, ON**

- M.Sc. Epidemiology; Thesis: 'Assessment of intra- and inter-individual variability of outcome measures in ankylosing spondylitis and the efficacy and adverse effects of anti-TNF therapy'

1991-1995 **McMaster University** **Hamilton, ON**

- Honours B.Sc. Biochemistry; Minor in Anthropology

AWARDS

2009 **Canadian Cochrane Review of the Year (*Abatacept for rheumatoid arthritis)*** **Ottawa, ON**

CONFERENCES/PRESENTATIONS

Workshop: Basic Training in Systematic Reviews & Meta-Analyses - Jan 2016 **Houston, USA**

Assessing risk of bias; Interpreting and presenting results; SoF tables and GRADE.

Croatian Cochrane Branch 5th Symposium – April 2013 **Split, Croatia**

Workshop for Cochrane authors – SoF tables and GRADE. Maxwell L (presented), Tugwell P

Annual Review Academic Rheumatology Unit, University Hospitals, Bristol Nov.2012 **Bristol, UK**

Annual Rheumatology Lecture: Making the Most of the Evidence Tugwell P (presenter),Boers M, Buchbinder R , Ghogomu E, **Maxwell L**, McIlwain C, Pardo J, Rader T, Welch V

EULAR - June 2013 **Madrid, Spain**

Risk of cancer, serious lung infections and death with biologics: A systematic review and network meta-analysis of randomized controlled trials (RCTs). (Poster). Singh JA, Wells G, Christensen R, Ghogomu E, MacDonald J, **Maxwell L**, Tarp S, Buchbinder R, Tugwell P and the Cochrane Biologics Study Group

EULAR - June 2012 **Berlin, Germany**

Adverse effects of biologics: A network meta-analysis and Cochrane overview (Poster). Singh JA, Wells G, Christensen R, Tanjong E, **Maxwell L**, Tugwell P and Buchbinder R and the Cochrane biologics Network Meta-analysis Group

Clinical Epidemiology Rounds, Ottawa General Hospital – April 2012 **Ottawa, Canada**

OMERACT: developing core sets of outcomes for trials. **Maxwell L** (presented), Tugwell P, Ghogomu E, Pardo Pardo J, Rader T, Toupin-April K, Welch V

ASAS Annual Meeting – January 2012 **Amsterdam, The Netherlands**
Presentation: Disease activity fluctuation in a cohort of AS patients (based on M.Sc. thesis). **Maxwell L**, Wells G, Tugwell P, Boonen A (presenter), Landewé R, van der Heijde D

Croatian Cochrane Branch Meeting – June, 25 2010 **Split, Croatia**
Presentations: "How do you critically appraise a systematic review? AMSTAR" and "Communicating the Evidence: Cochrane Summary of Findings Tables". Tugwell P (presenter); Foote M, Ghogomu E, **Maxwell L**; Shea B.

6th Canadian Cochrane Symposium – March 2008 **Edmonton, AB**
- Is health equity considered in systematic reviews of the Cochrane Musculoskeletal Group? Tugwell, P, **Maxwell L** (presenter), Welch V, Kristjansson, E, Petticrew M, Wells G, Buchbinder R, Suarez-Almazor, M, Nowlan, M-A, Morris E, Khan M,; Shea B, Tsikata S

5th Canadian Cochrane Symposium – February 2007 **Ottawa, ON**
- Cochrane Author Training Workshop (facilitator)

ACR (American College of Rheumatology) Meeting - Nov. 2006 **Washington, DC**
- The Clinical Meaning of Improvement in Fatigue and Sleep Quality in Patients with Rheumatoid Arthritis (Poster). Wells G, Li T, **Maxwell L**, Bahrt K, Tugwell P

- Sensitivity of Measures of Patient Reported Outcomes Following Treatment with Abatacept in Patients with Rheumatoid Arthritis (Poster). Wells G, Li T, **Maxwell L**, Maclean R, Tugwell P

- Minimal Clinically Important Differences in Activity, Fatigue and Sleep Quality in studies of Rheumatoid Arthritis (Poster). Wells G, Li T, **Maxwell L**, Maclean R, Tugwell P

IMHA – 2nd Annual Meeting of STIHR – May 2006 **Ottawa, ON**
- Knowledge Translation: Why and How. Tugwell P, Grimshaw J, **Maxwell** L, Santesso N

IMHA – Knowledge Exchange Task Force - April 2006 **Ottawa, ON**
- Evidence on Pain and Evidence-based Messaging. **Maxwell L**, Santesso N, Tugwell P (presented)

4th Canadian Cochrane Symposium – December 2005 **Montreal, QC**
- Cochrane Author Training Workshop (facilitator)

ACR (American College of Rheumatology) Meeting - Nov. 2005 **San Diego, CA**
- Sensitivity of Measures of Function and Patient Reported Outcomes Following Treatment with Abatacept in Patients with Rheumatoid Arthritis (Poster). Wells G, Li T, **Maxwell L**, Maclean R, Tugwell P

- Determining the Minimal Clinically Important Differences in Activity, Fatigue and Sleep Quality for Rheumatoid Arthritis(Poster). Wells G, Li T, **Maxwell L**, Maclean R,Tugwell P

Cutting Edge Debates in Evidence-Informed Public Health – Oct. 2005 Melbourne, AU

- Cochrane/Campbell Health Equity Field. Tugwell P, Petticrew M, **Maxwell L** (presenter),Robinson

13th Cochrane Colloquium – Oct. 2005 Melbourne, AU

- A tale of four review groups – experiences of Cochrane review authors (Poster). Pennick V, Gillespie L, **Maxwell L**, Mayhew A

Cochrane Workshop – Nov. 2004 Toronto, ON

- Cochrane Author Training Workshop (facilitator)

12th Cochrane Colloquium – Oct. 2004 Ottawa, ON

- Making the results of Cochrane reviews more accessible and friendly. **Maxwell L** (presenter), Judd M, Santesso N, Robinson V, Tugwell P, Wells G.

- Consumer priority survey - Bridging the gap between producers and users & of reviews. Judd M, Walker J, Qualman A, **Maxwell L**, Santesso N, Tugwell P and the Cochrane Musculoskeletal Group

10th Canadian Conference on International Health – Oct. 2003 Ottawa, ON

- 'The Right to Health – A Dominican Perspective'. **Maxwell L** (presenter)

MEMBERSHIPS

2005 - Present
Cochrane Equity Methods Group (since 2005); Cochrane Risk of Bias Group (since 2008); Cochrane Comparing Multiple Interventions Methods Group (since 2008)

- Member

EXTERNAL REVIEWER

2013 – Pain

2012 – Journal of Clinical Epidemiology

2011 – Agency for Healthcare, Quality and Research (AHRQ)

2010 – Therapeutic Advances in Drug Safety

PUBLICATIONS

1. Chiarotto A, **Maxwell LJ**, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in non-specific low back pain? A systematic review and meta-analysis. Phys Ther. April 2016 (in press)

2. **Maxwell LJ**, Wells GA, Simon LS, Conaghan PG, Grosskleg S, Scrivens K, Beaton D, Bingham III CO, Busse JW, Christensen R, Goel N, Juni P, Kaiser U, Lyddiatt A, Mease PJ, Ostelo R, Phillips K, Sapunar D, Singh JAS, Strand V, Taylor AM, Terwee CB, Tugwell P. Current State of Reporting Pain Outcomes in Cochrane Reviews of Chronic Musculoskeletal Pain Conditions and Considerations for an OMERACT Research Agenda. J Rheumatol October 2015 42(10):1934-1942; published online before print September 15, 2015, doi:10.3899/jrheum.141423

3. Phillips K, Taylor A, Mease PJ, Simon LS, Conaghan PG, Choy EH, Singh JAS, Strand VS, Gossec L, Kaiser U, de Wit M, Ostelo R, **Maxwell L**, Tugwell P. Harmonizing Pain Outcome Measures: Results of the Pre-OMERACT Meeting on Partnerships for Consensus on Patient-important Pain Outcome Domains Between the Cochrane Musculoskeletal Group and OMERACT. J Rheumatol October 2015 42(10):1943-1946; published online before print August 1, 2015, doi:10.3899/jrheum.14138

4. Toward Ensuring Health Equity: Readability and Cultural Equivalence of OMERACT Patient-reported Outcome Measures. Petkovic J, Epstein J, Buchbinder R, Welch V, Rader T, Lyddiatt A, Clerehan R, Christensen R, Boonen A, Goel N, **Maxwell LJ**, Toupin-April K, de Wit M, Barton J, Flurey C, Jull J, Barnabe C, Sreih AG, Campbell Willemina, Pohl C, Duruöz TM, Singh JA, Tugwell P, Guillemin F. J Rheumatol published online before print June 15, 2015,doi:10.3899/jrheum.141168

5. Busse JW, Bartlett S, Dougados M, Johnston BC, Guyatt GH, Kirwan J, Kwoh K, **Maxwell LJ**, Moore A, Singh JA, Stevens R, Strand V, Suarez-Almazor ME, Tugwell P, Wells G. Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from a 2014 OMERACT Workshop. J Rheumatol. October 2015 42 (10):1962-70 published online before print May 15, 2015, doi:10.3899/jrheum.141440

6. Christensen R, **Maxwell LJ,** Jüni P, Tovey D, Williamson PR, Boers M, Goel N, Buchbinder R, March L, Terwee CB, Singh JA, Tugwell P. Consensus on the Need for a Hierarchical List of Patient-reported Pain Outcomes for Metaanalyses of Knee Osteoarthritis Trials: An OMERACT Objective. J Rheumatol. 2015 May 6. DOI: 10.3899/jrheum.141384 [Epub ahead of print]

7. **Maxwell LJ**, Zochling J, Boonen A, Singh JA, Veras MMS, Tanjong Ghogomu E, Benkhalti Jandu M, Tugwell P, Wells GA. TNF-alpha inhibitors for ankylosing spondylitis. Cochrane Database of Systematic Reviews 2015, Issue 4. Art. No.: CD005468. DOI: 10.1002/14651858.CD005468.pub2.

8. Singh JA, Noorbaloochi S, MacDonald R, **Maxwell LJ**. Chondroitin for osteoarthritis. Cochrane Database of Systematic Reviews 2015, Issue 1. Art. No.: CD005614. DOI: 10.1002/14651858.CD005614.pub2.

9. Hansen JB, Juhl CB, Boutron I, Tugwell P, Ghogomu EAT, Pardo Pardo J, Rader T, Wells GA, Mayhew A, **Maxwell L**, Lund H, Christensen R, The Editorial Board of the Cochrane Musculoskeletal Group. Assessing bias in osteoarthritis trials included in Cochrane reviews: protocol for a meta-epidemiological study. BMJ Open 2014;4:e005491.doi:10.1136/bmjopen-2014-00549

10. Fidelix TS, Macedo CR, **Maxwell LJ**, Fernandes Moça Trevisani V. Diacerein for osteoarthritis. Cochrane Database of Systematic Reviews 2014, Issue 2. Art. No.: CD005117. DOI: 10.1002/14651858.CD005117.pub3

11. Updated Method Guidelines for Cochrane Musculoskeletal Group Systematic Reviews and Meta-Analyses. Ghogomu E, **Maxwell LJ**, Buchbinder R, Rader T, Pardo Pardo J, Johnston R, Christensen R, Singh J, Wells GA, Tugwell P and The Editorial Board of the Cochrane Musculoskeletal Group. *J Rheumatol 2014 Feb;41(2):194-205. doi: 10.3899/jrheum.121306. Epub 2013 Dec 1.*

12. Strategies to translate evidence Cochrane Musculoskeletal Group systematic reviews for use by various stakeholders: clinicians, patients, policy makers and the public: An update. Rader T, Pardo Pardo J, Stacey D, Ghogomu E, **Maxwell L.J**., Singh J.A., Buchbinder R, Légaré F, Santesso N, Winzenberg T, Tugwell P and The Editorial Board of the Cochrane Musculoskeletal Group. *J Rheumatol. 2014 Feb;41(2):206-15. doi: 10.3899/jrheum.121307. Epub 2013 Dec 1.*

13. Singh JA, Wells GA, Christensen R, Tanjong Ghogomu E, **Maxwell L**, MacDonald JK, Filippini G, Skoetz N, Francis DK, Lopes LC, Guyatt GH, Schmitt J, La Mantia L, Weberschock T, Roos JF, Siebert H, Hershan S, Cameron C, Lunn MPT, Tugwell P, Buchbinder R. Adverse effects of biologics: a network meta-analysis and Cochrane overview. Cochrane Database of Systematic Reviews 2011, Issue 2. Art. No.: CD008794. DOI: 10.1002/14651858.CD008794.pub2.

14. **Maxwell L**, Singh J. Abatacept for Rheumatoid Arthritis: A Cochrane Systematic Review. *J Rheumatol* 2010; 37:2doi:10.3899/jrheum.091066

15. **Maxwell L**, Singh J. Abatacept for rheumatoid arthritis. *The Cochrane Database of Systematic Reviews* 2009, Issue 4. Art. No.: CD007277. DOI: 10.1002/14651858.CD007277.pub2

16. Is health equity considered in systematic reviews of the Cochrane Musculoskeletal Group? Tugwell, P, **Maxwell L**; Welch V; Kristjansson, E, Petticrew M, Wells G, Buchbinder R, Suarez-Almazor, M; Nowlan, MA; Morris E, Khan M, Shea B; Tsikata, S. *Arthritis & Care Research*. 2009 Volume 59 Issue 11, Pages 1603 – 1610

17. Occupational therapists should be more involved in the Cochrane Collaboration: The example of the Australian Cochrane Musculoskeletal Review Group; Martin E, Baggaley K,

Buchbinder R, Johnston R, Tugwell P, **Maxwell L**, Santesso N. *Australian Occupational Therapy Journal. 2008; 55 (3): 207–211*, doi: 10.1111/j.1440-1630.2007.00711.x

18. Responsiveness of patient reported outcomes including fatigue, sleep quality and activity limitation and quality of life following treatment with abatacept in patients with rheumatoid arthritis; Wells G, Li T, **Maxwell L**, Maclean R, Tugwell P. *Ann Rheum Dis.* 2008;67:260–265; doi:10.1136/ard.2007.069690

19. Determining the Minimal Clinically Important Differences in Activity, Fatigue, and Sleep Quality in Patients with Rheumatoid Arthritis. Wells G, Li Tracy, **Maxwell L**, MacLean R, Tugwell P. *J Rheumatol 2007*;34:280–9.

20. Evidence Based Chiropractic Care: Systematic Reviews from the Cochrane Musculoskeletal Group. Santesso N, **Maxwell L**, Tugwell PS, Buchbinder R, Johnston R, and the Editorial Board of the Cochrane Musculoskeletal Group. *Journal of the Canadian Chiropractic Association* 2006; 50(4).

21. Method guidelines for Cochrane Musculoskeletal Group systematic reviews. **Maxwell L**, Santesso N, Tugwell PS, Wells GA, Judd M, Buchbinder R and the Editorial Board of the Cochrane Musculoskeletal Group. *Journal of Rheumatology* 2006; 33:2304-11.

22. Knowledge Transfer to Clinicians and Consumers by the Cochrane Musculoskeletal Group. Santesso N, **Maxwell L**, Tugwell PS, Wells GA, O'Connor AM, Judd M, Buchbinder R, and the Editorial Board of the Cochrane Musculoskeletal Group. *Journal of Rheumatology* 2006;33:2312-8

23. Tugwell P, Petticrew M, Robinson V, Kristjansson E, **Maxwell L**, Cochrane Equity Field Editorial Team. Cochrane and Campbell Collaborations, and health equity. Lancet 367(9517):1128-30, 2006 Apr 8.

24. Towheed TE, **Maxwell L**, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. *The Cochrane Database of Systematic Reviews* 2006, Issue 1. Art. No.: CD004257. DOI: 10.1002/14651858.CD004257.pub2.

25. Towheed TE, **Maxwell L**, Anastassiades TP, Shea B, Houpt J, Robinson V, Hochberg MC, Wells G. Glucosamine therapy for treating osteoarthritis. *The Cochrane Database of Systematic Reviews* 2005, Issue 2. Art. No.: CD002946.pub2. DOI: 10.1002/14651858.CD002946.pub2. (updated Issue 1, 2009)

26. Garner SE, Fidan DD, Frankish RR, **Maxwell L**. Rofecoxib for osteoarthritis. *The Cochrane Database of Systematic Reviews* 2005, Issue 1. Art. No.: CD005115. DOI: 10.1002/14651858.CD005115.

27. **Maxwell L**, Tugwell P. High-intensity exercise for rheumatoid arthritis was associated with less joint damage of the hands and feet than physical therapy. ACP J Club. 2005 May-Jun;142(3):73