

Stakeholder feedback from the Cochrane Pain, Palliative and Supportive Care (PaPaS) Review Group on the National Institute of Health and Care Excellence (NICE) draft clinical guideline GID-NG10069 Chronic pain: assessment and management

10/9/20

Clinical guidelines for the management of chronic primary pain have the potential to improve the quality and consistency of care for a group who commonly feel neglected by a healthcare system that does not work for them and as such we at Cochrane Pain, Palliative, and Supportive Care Review Group (PaPaS) are committed to working collaboratively with guideline developers to facilitate the development of rigorous guidelines based on the best available evidence that the community of people living with pain, those that treat them clinically and policy makers can use with confidence. In this spirit we welcome the opportunity to offer feedback as a registered stakeholder and offer what we intend as a constructively critical commentary on the draft NICE guideline: GID-NG10069 Chronic pain: assessment and management.

As a group we have extensive experience of synthesising evidence in this field and through that experience we are very aware of the many substantial difficulties that can arise when trying to draw conclusions and develop workable recommendations from what is often a rather messy evidence base. This area is affected by challenges of clinical heterogeneity (in populations and interventions), diagnostic ambiguity, difficulty in capturing an elusive outcome (the experience of pain), heterogeneity in treatment response, the highly variable quality of relevant clinical studies and generally small average treatment effects.

These challenges result in substantial uncertainty and leave the findings of evidence reviews prone to being unhelpfully influenced by specific methodological choices and open to varied interpretations. In offering our feedback we hope to constructively raise concerns of this nature for the committee to consider.

Concerns with the scope

The draft guideline states that it covers "...assessing and managing chronic pain in people aged 16 years and over" and should be used alongside existing NICE guidance for "specific conditions" that cause pain, including headaches, low back pain and sciatica, rheumatoid arthritis, osteoarthritis, spondylarthritis, endometriosis and irritable bowel syndrome. It includes recommendations on managing chronic primary pain (as defined in ICD-11) for which there is no other NICE guidance.

The use of the word "specific" is worthy of attention here as some of the above listed conditions will include people who fit the ICD-11 definition. This is particularly true of low back pain, where non-specificity and diagnostic uncertainty is the norm; many people would fit the ICD-11 definition and also where there is the largest evidence base for clinical interventions. One possible unintended consequence of pragmatically excluding studies in populations for which there is existing NICE guidance is that it may exclude highly relevant evidence that may be of better quality than what is actually included.

The evidence base for many of the conditions that remain within the scope is relatively small and immature. As such the resultant evidence reviews for NICE mainly include small and relatively exploratory studies. This can have an important impact on the resulting evidence reviews and subsequent recommendations by introducing a study-level selection bias where larger more robust

trials are selectively excluded. The opposing risk of the scope is that of pooling data from heterogenous clinical populations and interventions which may result in failure to identify a uniquely effective intervention in a broader class or one that is that is effective for a specific patient group.

The decision to include people under 18 in the guideline raises issues. Although many children's hospitals cease intake of patients over the age of 16, many pain clinics continue to treat people as children beyond the age of 16. In fact, childhood was recently redefined as continuing to the age of 24 in 2018 (Sawyer 2018). Certainly, children between 16-18 years are often included in paediatric studies. Chronic pain presents challenges to children who are less autonomous, continue to live at home and rely heavily on their parents. Starting the guidance at 16 years results automatically in a void of evidence as children of this age-range are included mostly in paediatric studies.

Clearly decisions regarding scope were made a priori and cannot be changed at this stage but a clear recognition of how they may impact the evidence reviews is vital when drawing conclusions and forming recommendations.

Concerns with the methods

Pain is a field in which the choice of methods can profoundly affect the results obtained. At PaPaS we have produced a suite of high-quality systematic reviews across the full range of interventions that pertain to the population of interest and should inform clinical decision making. Due largely to the unique and restrictive scope of this guideline many of these were excluded or not considered, despite being highly relevant. Cochrane reviews, including PaPaS reviews, were excluded from consideration in the evidence reviews for pharmacological, psychological, manual therapy, exercise, acupuncture, electrophysical modalities (TENS) and pain management programmes. It is disappointing and inefficient that many of these were not formally considered in the process.

Defining clinical importance in the field of persistent pain is a difficult question that has received a lot of attention and some, though not total, consensus. For within-person change it is encouraging to see the use of widely accepted "responder" thresholds, though disappointing to see them take a low priority, and of course disappointing to see how little of such data there was available. We recognise the need to apply thresholds to aid consistent decision making. In persistent pain it is clear that nothing works well for most people but for some interventions a small number of people may derive important benefit. As such the threshold applied of ≥ 10 absolute risk difference (NNTB 10) presents a risk of excluding an intervention that may offer important benefit to a small number of people with pain. An example of where this may have occurred is presented in our discussion of pregabalin and duloxetine below. For average between-group differences in pain the decision to base judgement thresholds of clinical benefit as a function of baseline variance is more problematic as variance in the measurement of outcomes is not a function of clinical importance.

Synthesis: The separation of analyses of pain by different measurement tools, and HRQoL into the multiple subdomains/ scales of the included measures, creates a significant issue of multiple comparisons. The predominance of single or 2 trial analyses throughout is not the best use of the data and sacrifices the potential precision that can be afforded by pooling. There is nothing in the Appendix: Methods 2.3.2 on Methods of combining clinical studies that explains why there is so little combination of similar studies with similar or the same outcomes, generating instead tens of single trial meta-analyses that jettison the power of meta-analysis.

Beyond the issue of precision, the size of studies may have a profound impact on their results that this might lead to an overly positive picture for some interventions (Dechartres 2013; Nüesch 2010).

Dechartres (2013) demonstrated that trials with fewer than 50 participants, which reflects the majority of studies included in this review, returned effect estimates that were on average 48% larger than the largest trials and 23% larger than estimates from studies with sample sizes of more than 50. Similarly, in Cochrane Reviews of amitriptyline for neuropathic pain and fibromyalgia (Moore 2015a; Moore 2015b), smaller studies were associated with substantially lower numbers needed to treat for an additional beneficial outcome (NNTBs) for treatment response than larger studies. In their recommendations for establishing best practice in chronic pain systematic reviews, Moore (2010) suggest that study size should be considered an important source of bias, as have others (Fanelli 2017; Flather 1997; IntHout 2015; Ioannidis 2005; Moore 1998; Nguyen 2017; Pogue 1998; Roberts 2015; Thorlund 2011; Turner 2015; Zhang 2013). Recent examples of how small study size can influence results include a commentary to a recent JAMA paper (Oberoi 2020), and in postoperative pain (Brink 2018). The bottom line is that conclusions based only on small studies are often or usually incorrect, especially where methodological considerations indicate significant risk of bias. That situation applies to several parts of the evidence presented to the committee.

These dual issues of multiple comparisons and study size raise the risk of multiple false positives but more broadly a serious problem of imprecision. In this instance small differences in methodological approach and interpretation are prone to producing quite different conclusions which may influence the recommendations of the committee and, in turn, patient care.

The approach taken to the application of GRADE may result in overestimating the certainty of the evidence for some comparisons with potential impacts on the decisions of the GDG. Imprecision judgements were based on whether the effect sizes and 95% confidence intervals overlapped the MID threshold. This approach arguably undervalues the importance of study size in determining the certainty of evidence and a more cautious approach would be to consider any analysis based on a small number of participants to be downgraded on the basis of imprecision. As an example, the approach taken by NICE would allow a single small study with poor randomisation and at subsequent risk of serious bias to be rated as offering moderate-quality evidence, which would be inappropriate. In both the evidence review on psychological treatment and on pain management programmes (1.7.12) is a statement about downgrading all trials for lack of blinding. While blinding is important in randomised trials (but rarely checked, only assumed by the nature of design even for drugs with well-known side effects that unblind participants), where it is clearly not possible, as in psychological treatment trials, methodological features that partly mitigate it have been investigated and used: assessment by staff blind to treatment allocation; expectations of improvement taken from all participants at baseline; an attempt at equipoise in those who deliver control and comparison conditions, rather than clear therapist allegiance expressed in some publications. These design features were discussed in committee, but appear to have been ignored. Simply marking down all trials for lack of blinding, and therefore lowering the overall quality, is not a helpful approach to the problem. Nor was there recognition that some trials used some outcomes that were not self-report: a feature that could have been identified better as positive and recommended to future trials.

In the light of the low confidence expressed about self-report, not without reason, it is surprising to see the weight put on exact calculation of change in scales for minimally important difference. None of this took account of unreliability of scales, often around the same size as the MID identified. Additionally, a quantum improvement in outcome scale has different meaning according to the baseline, which is why pain reduction is usually expressed in percentages rather than absolute values. For those with high baseline levels of pain, small reductions can be trivial; for those with low scores at baseline, they may represent substantial change. None of this appears to be recognised.

There follows commentary on specific evidence reviews and recommendations where we feel these concerns require closer consideration.

Consideration of specific recommendations

Risk factors

In the evidence review of risk factors, what is called 'comorbid psychiatric disorder' consists of anxiety or depression scores on non-diagnostic questionnaires. It is inaccurate and misleading to refer to this as psychiatric disorder. Misunderstanding of use of common psychiatric scales designed for and standardised on physically well populations is noted on p55 in relation to discussion by committee, but appears to have been ignored in the summary.

Pharmacological interventions

These are comments restricted to the pharmacological interventions that may be made in treating people with primary chronic pain, as defined by this guideline. Most of the comments concern fibromyalgia, which affects many people [global mean prevalence of potential cases of fibromyalgia estimated as 2.7% (range 0.4% to 9.3%), usually older women]. The condition is associated with very considerable disability and reduced quality of life, as well as severe and long-lasting pain that is difficult to treat. The few treatments known to be effective help no more than about 10% of people with the condition, but reduced pain is associated with improvements in associated symptoms, much improved quality of life, and ability to work.

Pregabalin

The guideline combines gabapentin, pregabalin, and (possibly) mirogabalin together under the generic term 'gabapentinoids'. It is not entirely clear why this is. The drugs **may** have similar mechanisms of action, but based on animal experiments that is increasingly being questioned. Moreover the evidence on gabapentin is relatively weak, and the excellent evidence on mirogabalin demonstrates that it is ineffective at the doses used. This commentary therefore sticks to a specific drug, and mostly a specific dose, specific patient-valued outcomes, and specific duration of trial.

Exclusions

The guideline has chosen to exclude a number of large, high-quality, randomised, double-blind trials that have been used to judge evidence of pregabalin efficacy and safety in fibromyalgia by, inter alia, the FDA, EMEA, and Cochrane reviews. There are four:

Arnold 2008, Mease 2008, and Pauer 2011 were excluded because they had "incorrect study design (placebo run-in phase)". Entry criteria for these studies was as follows (from Arnold, but they were all very similar designs):

*"Patients were considered eligible for the study if they were at least 18 years of age, male or female (were nonpregnant and nonlactating), met the American College of Rheumatology classification criteria for fibromyalgia, 34 and had a pain score of at least 40 mm on the 100-mm pain visual analog scale (VAS) at screening (visit 1) and random assignment (visit 2). **In addition, patients also had to complete a minimum of 4 of 7 daily entries in pain diaries during the 1-week, single-blinded run-in period, with average mean pain score ≥ 4 .**"*

The placebo run in was used to ascertain that these participants genuinely had moderate or severe pain at baseline. This is not only not an incorrect study design, but rather essential in establishing a sensitive assay. The requirement of moderate or severe pain in ascertainment of analgesic efficacy has been established for at least 75 years. In these trials, the ascertainment of at least moderate pain was even better established than usual. Almost all trials have a one-week assessment period

for establishing initial pain and when current treatments have been discontinued. For example, Arnold 2019, which is included, says that participants should have “ADPS of ≥ 4 on the 11-point numeric rating scale (NRS) over the past 7 days prior to randomization (based on completion of at least 4 daily pain diaries during the 7-day baseline period prior to randomization)”. There is little or no difference between a week on no drugs or a week on no drugs plus placebo. These three trials, with data on almost 2,250 people with fibromyalgia, have been erroneously excluded.

Crofford 2005 was excluded because “Not review population. Excluded known non-responders”. The exclusion was actually: “Those who had failed to respond to previous treatment with gabapentin at dosages $\geq 1,200$ mg/day for pain associated with FMS were excluded.” But this was a trial of pregabalin, not gabapentin, and the discussion admits that “prior beneficial response to gabapentin was not systematically recorded, so it is not possible to determine whether these participants were more likely to respond to treatment with pregabalin.” Not only was this not an exclusion of known non-responders, and certainly **not** an exclusion of non-responders to pregabalin (the drug under test), but also the evidence is that this sort of partial enrichment has no effect on analgesic efficacy assessment with pregabalin, where the maximum enrichment was by about 12% (Straube 2008). As a result we consider that this trial, with over 500 people with fibromyalgia, has been erroneously excluded.

Crofford 2008 was excluded on the grounds: “Not review population. Only responders”. That is not exactly true: the participants screened and entering the initial open label phase of the study were exactly the same in terms of the inclusion and exclusion criteria used as participants in other trials, fulfilling ACR criteria for fibromyalgia, and having at least moderate pain at screening and baseline visits. After a six week open label phase to determine whether participants could both get adequate pain relief, and those with “ $\geq 50\%$ reduction in pain VAS score from OL baseline and a self-rating of overall improvement on the PGIC scale of “much improved” or “very much improved” were then randomised to continuing with their established dose, or placebo, for a six-month period.

This trial is an exemplar of how enriched enrolment randomised withdrawal (EERW) trials should be done (Pain 2015 156:1382–1395) and mimics real world conditions. Although EERW designs cannot be combined with studies of conventional design, they can inform in just the same way. This trial, with over 1,000 people with fibromyalgia, has been erroneously excluded.

In total five large, high-quality, randomised and double blind trials of pregabalin in fibromyalgia have been erroneously excluded: totalling information on over 3,700 people.

A consequence of this policy is that most of the analyses performed for antiepileptics have data from only a few, rather small, trials for pain. Figure 4 has about 500 participants in total, Figure 5 54, and only Arnold 2007 and Arnold 2019 contribute data for fibromyalgia, the former with 117 participants and the latter with 1,903 participants, but only a single (different) pain outcome with each.

Duloxetine

Exclusions and inclusions

Arnold 2004 is excluded because it had a one-week placebo run in to establish minimum pain requirement. As already explained, that is an error, and it leads to the improper exclusion of data from 205 people with fibromyalgia.

However, Arnold 2012 is included. This trial used a suboptimal dose of 30 mg pregabalin daily, at least half that used in all other trials, and used clinically. Including this trial (with zero treatment effect) in an analysis of effective doses was an error.

Outcomes analysed

Good clinical trials are data rich, and Cochrane reviews of pregabalin in fibromyalgia and neuropathic pain, and those on duloxetine in a range of pain conditions, offer many different ways of expressing analgesic results. Often forgotten is the patient perspective – what do participants with pain want of therapy? The answer is consistent across all acute and chronic pain, and headache – large degrees of pain relief, and quickly. A recent systematic review demonstrates this clearly (Moore 2013a). Another demonstrates that people with pain rate their pain very differently from their carers, who typically downgrade the patient experience, roughly by the magnitude of some of the best analgesics known (Seers 2018). This is why pain studies have moved significantly in reporting, so that at least 50% (or at least 30%) pain intensity reduction has become the standard. These values are available in the excluded, and in at least one of the included studies (though neither sought nor used in this evidence assessment). Importantly with this form of outcome analysis, patient response is bimodal – either very large benefit is seen, or very little, with very few participants experiencing an ‘average’ response; this has important consequences for other symptoms.

An alternative analysis of pregabalin and duloxetine trials

Using outcomes important to participants with pain, our Senior Editor Andrew Moore has for fibromyalgia performed an analysis combining the data in the Cochrane review of pregabalin 300 mg daily for fibromyalgia with the three studies in Arnold 2019 using WebPlotDigitizer to abstract the relevant numbers from graphs for the three Daiichi trials.

Pregabalin results from seven trials and 3,278 patients using 300 mg daily for at least three months are shown in Figures 1 and 2, for at least 50% and at least 30% pain intensity reduction respectively. The magnitude of the risk difference is 8% and 9% respectively, significantly better than placebo, and with no measurable heterogeneity in these large, high-quality, long duration studies. Moreover, a large (1,000 patient) EERW trial lasting six months confirms the degree of benefit, and that it continues in the long term for pregabalin.

Figure 1: Pregabalin 300 mg daily: at least 50% pain intensity reduction

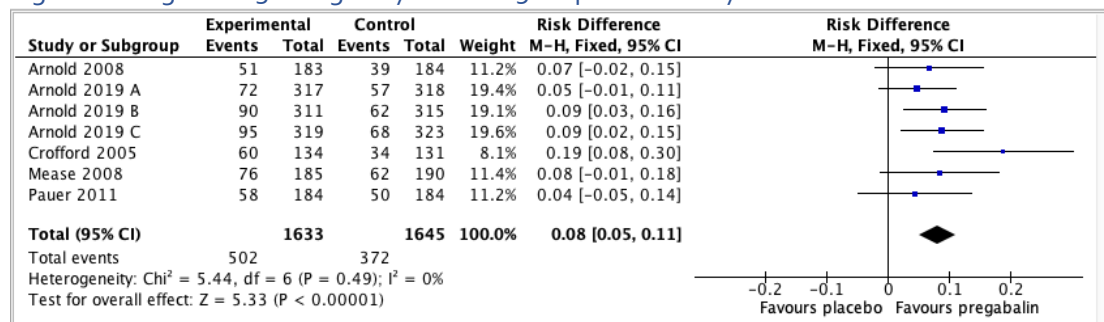
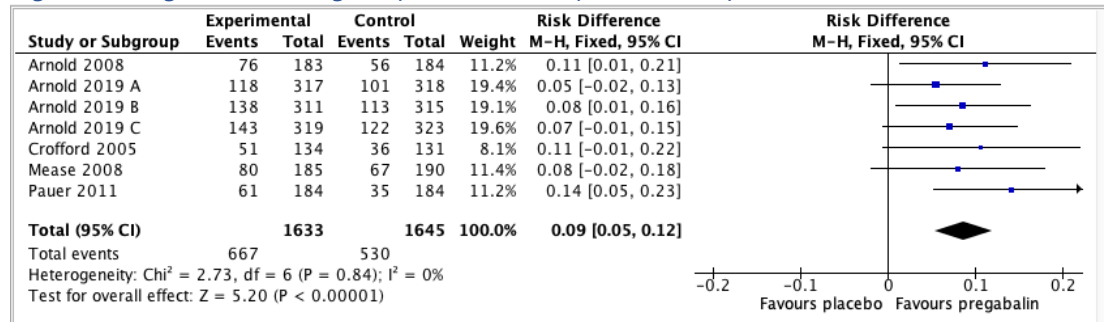


Figure 2: Pregabalin 300 mg daily: at least 30% pain intensity reduction



Duloxetine results from six trials and 2,246 patients using 60 or 120 mg daily for at least three months are shown in Figures 3 and 4, for at least 50% and at least 30% pain intensity reduction respectively. The magnitude of the risk difference is 9% and 11% respectively, significantly better than placebo, and with limited heterogeneity in these large, high-quality, long duration studies.

Figure 3: Duloxetine 60/120 mg daily: at least 50% pain intensity reduction

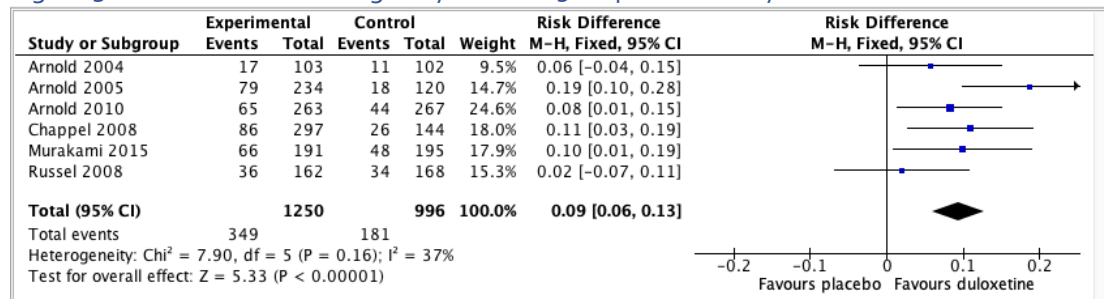
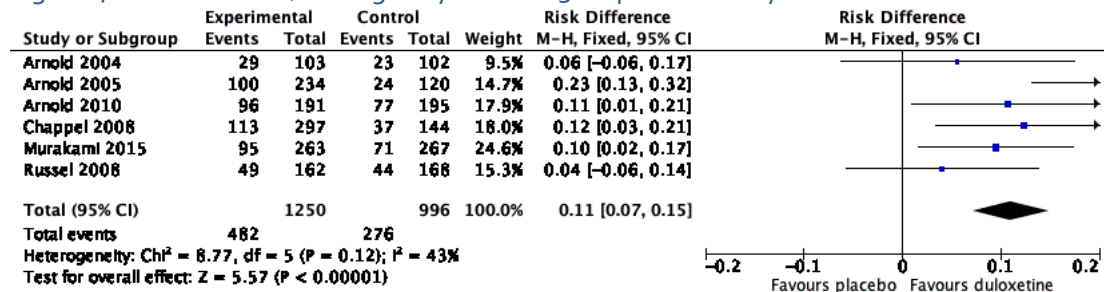


Figure 4: Duloxetine 60/120 mg daily: at least 30% pain intensity reduction



Comparing efficacy with pregabalin and duloxetine

Table 1 shows the comparison between the summary analyses of these two interventions for fibromyalgia. Each uses:

- the same dose of the drug under test in a comparison with placebo,
- using essentially similar patient populations with at least moderate pain relief (typical mean initial pain scores were in the range 6 to 7.5 out of 10, indicating most had severe pain),
- the same study duration of around three months,
- the same or very similar methods of ascertainment of pain by the patient,
- the same patient-centered outcomes,
- the same method of analysis,
- using all available data (at least all immediately available at short notice).

The table includes data on over 5,500 participants, and, for the each of the two outcomes, percentages with treatment and placebo achieving the outcome is very similar. For each outcome, about 10% more of the participants treated had the outcome with treatment than with placebo.

Table 1: Comparison between analyses of pregabalin and duloxetine

Drug daily dose	Number of		Percent with outcome		Risk difference (95% CI)	p for difference
	Trials	Patients	Active	Placebo		
At least 30% pain intensity reduction						
Pregabalin 300 mg	7	3278	41	32	0.09 (0.05 to 0.12)	z=0.858 p=0.39
Duloxetine 60/120 mg	6	2246	40	31	0.11 (0.07 to 0.15)	
At least 50% pain intensity reduction						
Pregabalin 300 mg	7	3278	31	23	0.08 (0.05 to 0.11)	z=0.693 p=0.49
Duloxetine 60/120 mg	6	2246	28	18	0.09 (0.06 to 0.13)	

Comparison of the responses between the two drugs at each outcome indicates no significant difference in effect size, using 2-tailed z-test.

The committee found that duloxetine had no benefit on pain in the long term, despite this overwhelming evidence, far larger in quantity and longer in duration than data for other antidepressant drugs. The committee found little or no benefit of antiepileptic drugs in terms of pain in shorter or longer term, and that *“there was insufficient evidence to justify the routine use of gabapentinoids for chronic primary pain.”*

As presented here, for fibromyalgia there is a large amount of high-quality data with little uncertainty, and with confidence intervals that include the threshold of absolute difference of ≥ 10 set as a limit by NICE. The evidence presentation to the committee omitted very large amounts of directly relevant evidence, by failing to follow established evidence for patient-centred outcomes, and by presenting the evidence in a way that precluded the committee from making a proper, evidence-based decision.

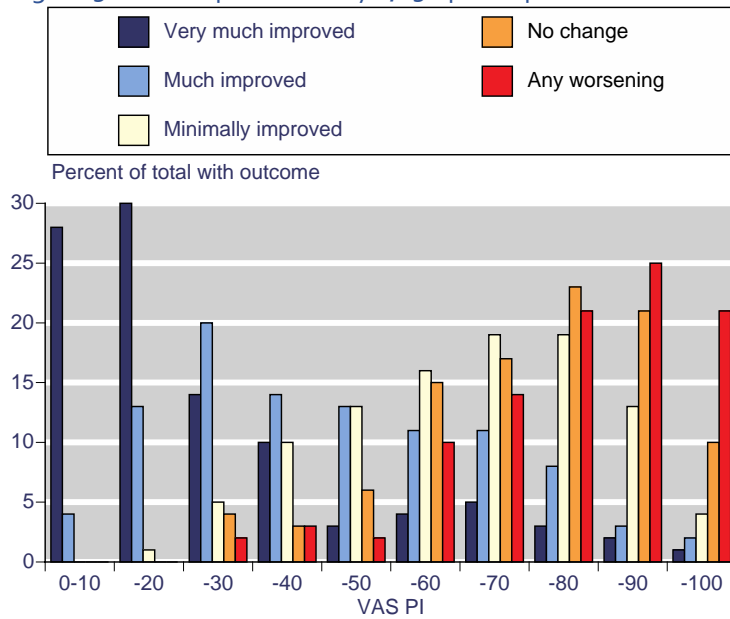
The individual patient experience

Clinical trials of pregabalin used for the treatment of fibromyalgia have examined the individual experience of pain, and have linked their pain experience to the experience of concomitant symptoms (fatigue, depression, sleep, etc). The experience of people with fibromyalgia who are successfully treated – their pain is reduced by a satisfactory degree – is similar to those with other pain conditions. Those who have good pain relief experience significant clinical benefit in all the other symptoms, and their quality of life improves dramatically.

Pain

How participants express their experience in terms of a global impression of change is associated with their pain at the end of a three-month trial, as shown in Figure 5. Those much or very much improved typically have low pain scores (at worst mild pain), while those reporting minimal change, no change, or worsening report typically moderate or severe pain.

Figure 5: PGIC experienced by 1,858 participants with fibromyalgia completing 8-14 week trials



The pain experience at the end of these trials is typically determined early in the trial and then sustained. One trial used a 1-100 mm VAS scale with intermediate reporting at 5 and 9 weeks. Figures 6-8 show the pain scores in the individual participants according to their pain intensity reduction at the end of the trial.

Most of those who did not respond never responded at any time, though there were some who had an early response but who withdrew from the trial principally because of intolerable adverse events. This was the largest group of participants, about 50%. Their pain scores throughout the trial were predominantly in the range of severe pain.

Most of those who had an intermediate response, between 15% and 50% pain intensity reduction had a similar response throughout the period. This was the smallest group, about 20%. Their pain scores throughout the trial were predominantly in the range of moderate pain, though some were severe and some were mild at the end of the trial.

Most of those who had a good response responded early, typically maintained that response throughout the trial, and had final pain scores of mild pain at the end of the trial. This was about 30% of the total.

Figure 6: People with pain intensity reduction 0-15% at end of trial, where withdrawal uses initial pain score

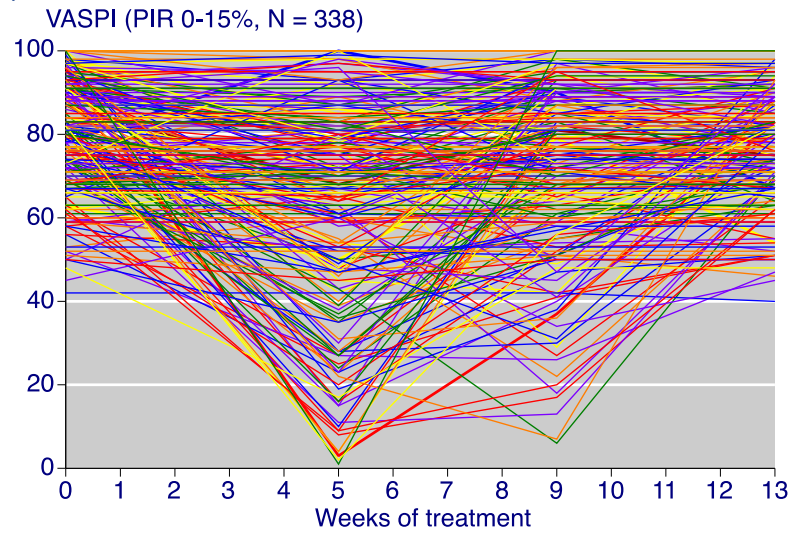


Figure 7: People with pain intensity reduction 15-49% at end of trial

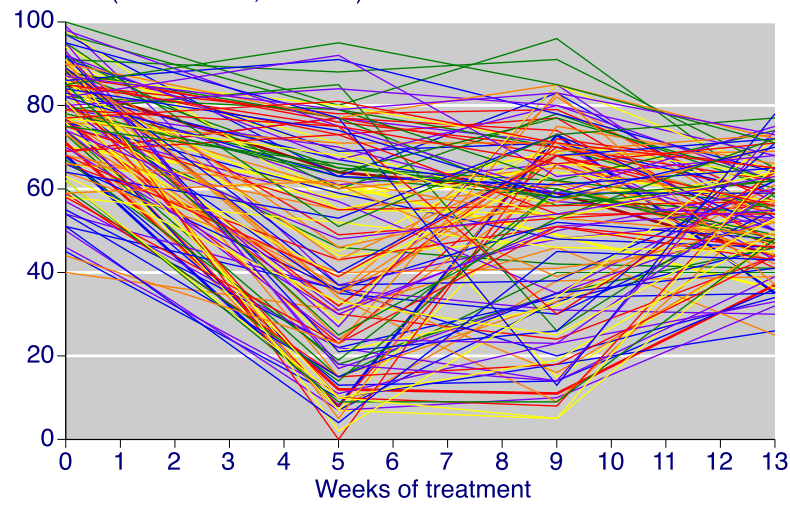
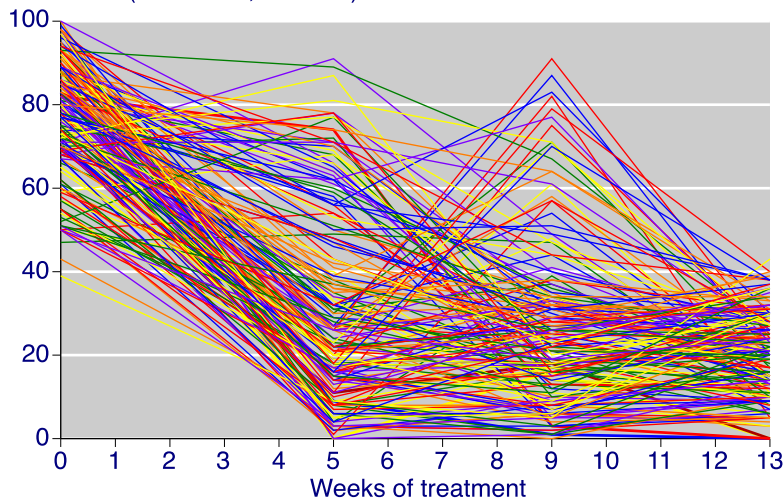


Figure 8: People with pain intensity reduction $\geq 50\%$ at end of trial
VASPI (PIR $\geq 50\%$, N = 177)



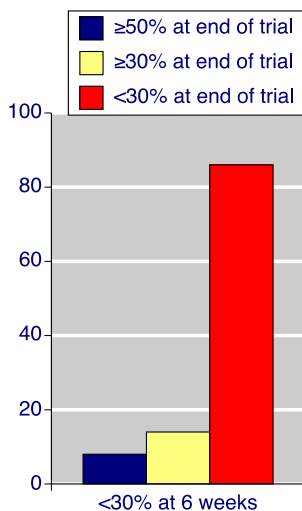
Stopping rules

This information can be used to test the potential for a “stopping rule” of value for clinical practice. A “stopping rule” is a point where we can be pretty sure that further treatment is futile. A stopping point is reached when:

- a patient stops treatment because of adverse or other event
- a patient experiences an inadequate level of pain relief to justify further treatment, *in the knowledge that further treatment will NOT bring good pain relief*

Stopping treatment prevents treatment when there are risks and costs, but no benefit. Using the data from the 645 participants described above, and using a pain intensity reduction of less than 30% from that at the beginning of treatment at 5 weeks as a stopping rule, we can test how efficient it would be. Figure 9 shows that 86% would not have achieved any useful pain relief, 14% may have achieved $\geq 30\%$ pain intensity reduction, and 8% $\geq 50\%$ pain intensity reduction.

Figure 9: End of trial result in participants with pain intensity reduction of less than 30% after five weeks of treatment



Quality of Life

The committee rightly identified throughout the document that quality of life was of key importance. However, the presentation of evidence was such that only **average** quality of life data were presented. It is obvious that when the magnitude of a difference in effect size in pain is small, there is unlikely to be much benefit in concomitant measures such as quality of life or sleep. For pain there is abundant evidence that large degrees of pain relief are associated with large benefits in concomitant areas, especially in terms of quality of life. For fibromyalgia there is good evidence from individual patient-level analysis that those with good pain relief have large benefits in quality of life and concomitant symptoms (sleep, depression), and their ability to work.

For example, analysis by degree of pain relief demonstrates stepped benefits in terms of quality of life (Figure 10) and days missed from work (Figure 11) (Straube 2011a) using data from almost 2,000 participants enrolled in clinical trials of fibromyalgia. Even better results are obtained for those with at least 50% pain intensity reduction and pain score below 3/10 at the end of the trial, in whom almost four days per week of work are gained (Straube 2011b).

Figure 10: Quality of life and pain
1-year QALY gain

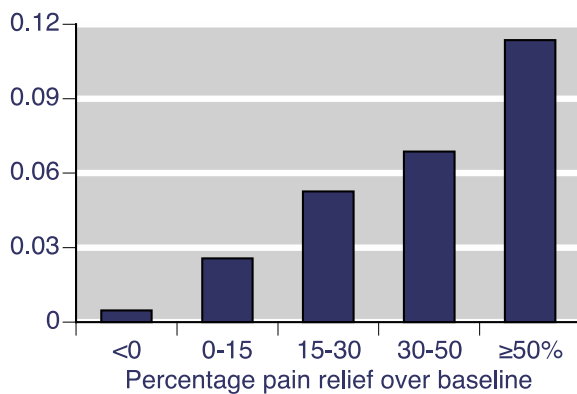
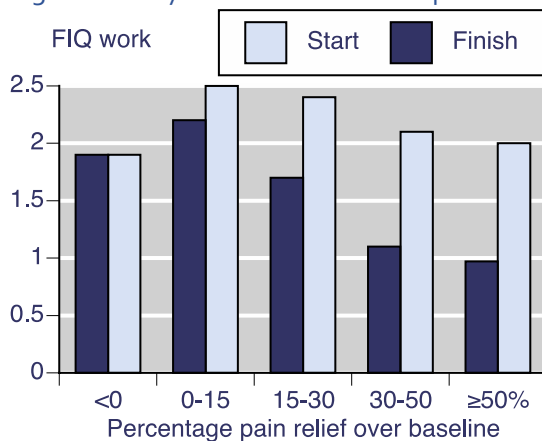


Figure 11: Days lost from work and pain



Benefits go further, and include fatigue (Figure 12), sleep disturbance (Figure 13), depression (Figure 14), disability (Figure 15), or all components of the SF-36 (Figure 16). All show large benefits in those people with the greatest degree of pain relief.

Figure 12: Fatigue measures from Global Fatigue Index (GFI) according to pain intensity reduction

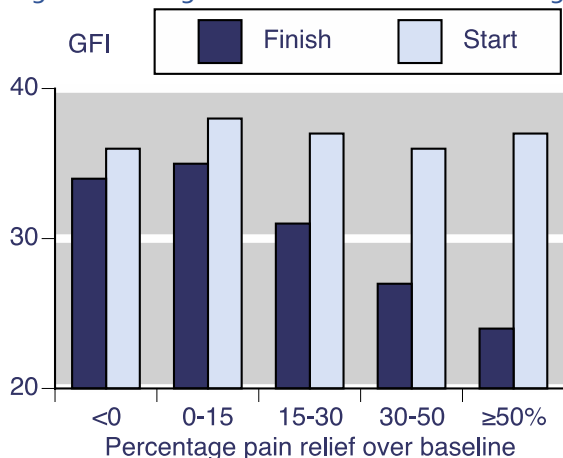


Figure 13: Sleep disturbance (SD) according to pain intensity reduction

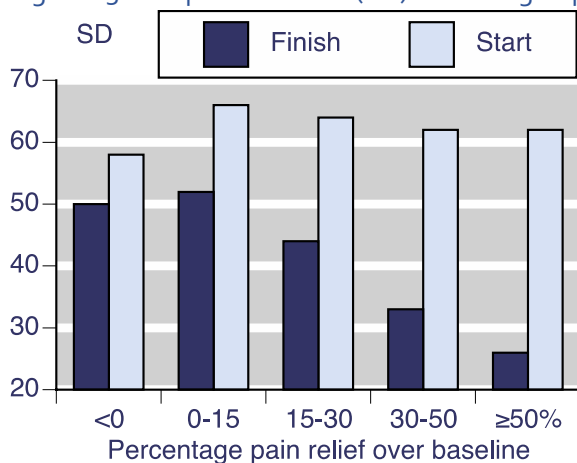


Figure 14: HADS Scale according to pain intensity reduction

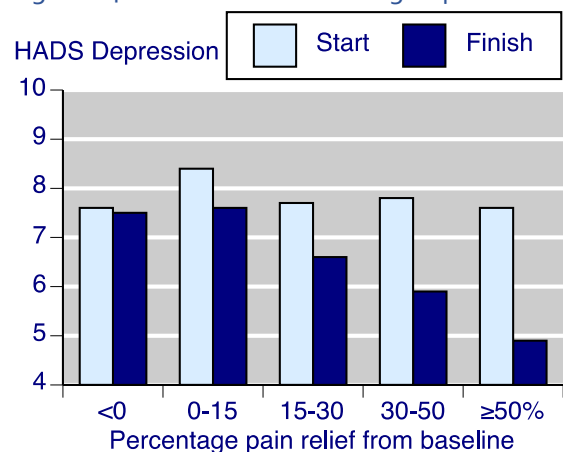


Figure 15: Work disruption from Sheehan Disability Scale according to pain intensity reduction

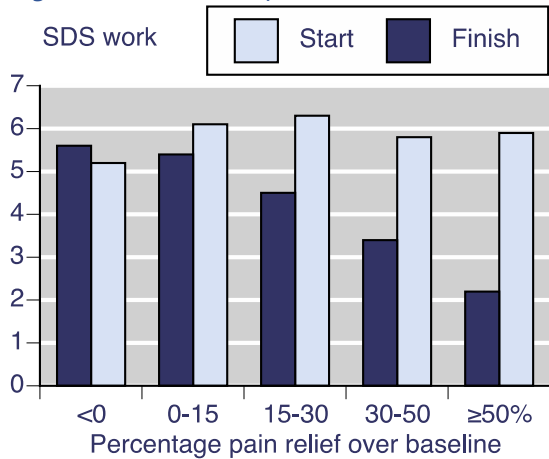
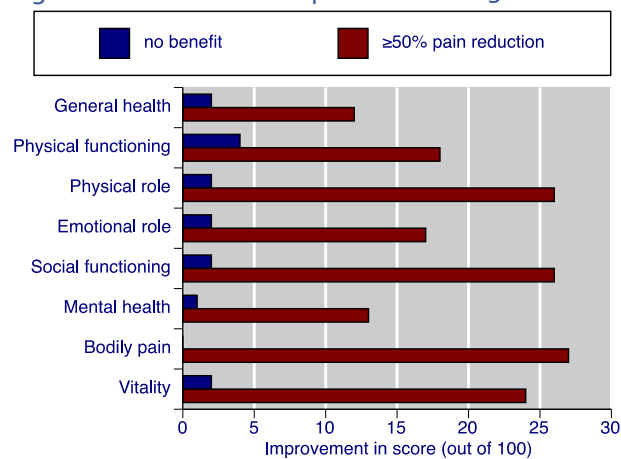


Figure 16: Individual components of SF-36 according to pain intensity reduction



This evidence undermines the conclusions of the guideline that there is little or no quality of life benefit from the use of antiepileptics in these pain conditions. On the contrary, those with good pain relief have values for quality of life that approach normal, including the ability to work.

Conclusion

These comments relate only to some aspects of pharmacological therapy for one pain condition in this guideline. The conclusion is that the methods of evidence collection, analysis, and presentation used in the guideline were flawed, consequently undermining the committee's ability to make an informed judgement.

There is an established evidence base demonstrating that, for fibromyalgia at least, there is good evidence that pregabalin not only has a similar effect size for pain as duloxetine, but also that those patients with good pain relief derive large benefits across all their concomitant symptoms, their quality of life, and their ability to work.

Acupuncture

The recommendation to consider acupuncture is interesting in that it deviates from the recommendations of the most recent NICE guidance on osteoarthritis and low back pain, both of which gave a "do not offer" recommendation on the basis of evidence of a lack of efficacy.

It is notable that the certainty of the evidence around the efficacy of acupuncture (vs sham) for pain (VAS) is very low in the largest analysis. The studies that comprise this comparison are generally small and at risk of multiple important biases that might be expected to exaggerate any true effect. There are major threats to clinician and patient blinding in the included studies (participant blinding is frequently suboptimal and from the clinician perspective most studies are effectively open-label, though reasonable double-blind methods are available for many acupuncture approaches) as well as issues with randomisation and allocation concealment, selective outcome reporting, incomplete outcome data and very high statistical heterogeneity (inconsistency). Similar issues of study quality impact the other comparisons. With the focus (appropriately) on subjective self-reported outcomes these multiple biases have great potential to create falsely positive results. It is also here that the scope may have an influence as the exclusion of studies in some conditions (for example in low back pain where numerous larger scale trials exist) introduces a study-level selection bias that broadly limits the analysis to smaller poorer quality studies. Contrasted with recent more inclusive synthesis of the efficacy of acupuncture (Vickers et al 2018) which include larger and more rigorous trials and found very small, clinically trivial differences between acupuncture and sham acupuncture, notwithstanding similar blinding issues, these results are incongruous and should be considered very carefully. There is no theoretical reason to explain this contrast. The effect sizes seen in some of the included studies are extreme for this clinical field against any benchmark. Examples here include a number of studies which present mean differences in pain intensity of greater than 3 points on a 0-10 NRS. While that might superficially appear to be a positive, it should raise concerns regarding the veracity of those results that go beyond issues of blinding. Results from the multiple analyses for HRQoL are highly inconsistent.

Some detailed comments indicate that the conclusions as presented to the committee may not be reliable, including the fact that only 379 participants were in larger studies with group sizes above 50.

- Not all the treatments called acupuncture are necessarily deliverable. Couto 2014 used “deep dry needling combined with paraspinal deep intramuscular stimulation with needle rotation” Also, from this study, actual pain scores at 4 weeks were available in Figure 2, and better than the 4-week averages used in the calculation. This may be relevant to some other forms of acupuncture described.
- In Harris 2005, the week 9 timepoint chosen happened to be the point at which pain scores in the acupuncture group were lowest.
- Lee 2011 is problematical on two counts. Firstly, the initial pain scores were so low as to make for an insensitive assay of analgesia (note that moderate or severe initial scores are needed for sensitivity). Secondly, Lee 2011 is a republication with additional data from Lee 2008, but the data on pain is inconsistent between the two.
- Vas 2016 uses data of percentage change in pain score as if they were the change in pain in absolute measures. They are not. By calculating the actual changes they should be reductions of 3.2 and 2.1 (not 4.1 and 2.7), and the MD would be 1.1 not 1.4. The largest trial (Molsberger 2011) is a large study comparing acupuncture with sham acupuncture and conservative orthopaedic therapy. It has broadly good methods, though did not blind clinicians and did not evaluate the success of participant blinding raising a substantial risk of performance bias. Attrition was substantial and imbalanced across groups at the primary endpoint of 3 months with 45% of the sham group lost and counted as non-responders compared to 17% in the verum group though surprisingly NICE rated the study at low risk of bias for incomplete outcome data for this timepoint.

The recommendation to consider acupuncture is therefore made using a highly uncertain evidence base and in the context of a broader, more robust and relevant evidence base that offers a substantially different answer with little uncertainty. While the committee offer reassurance that they “took into account the low quality in their interpretation of the evidence” this does not solve the problem that when we aggregate poor quality studies they cannot lead us to a reliable answer.

Pain Management Programmes and psychological interventions

Pain Management Programmes (PMPs) are always psychologically informed, with direct psychological content and other therapeutic components, most often physical activity, but also including analgesic reduction, occupational therapy, and sleep promotion, and are delivered in ways consistent with psychological methods and content. For this reason it makes little sense to distinguish pain management programmes from psychological interventions – usually CBT, ACT, or mindfulness – that are rarely delivered without any other components alongside except in trials where the pain management package is ‘dismantled’ to attempt to identify unique effects of particular components.

Further, it seems that many trials were sifted out at an early stage by the review teams, so that the guidelines group never had the chance to discuss whether they should be included or not. Sifting appears to have been rather insensitive to the varied ways in which psychological and other pain management content is described in many trials. PaPaS have just published a systematic review and meta-analysis of psychological interventions for chronic pain in adults (Williams et al. 2020), but during the guidelines process, Prof Amanda Williams, a member of our editorial board and a co-opted member of the guidelines group, raised a number of times the discrepancies between the 2020 review which was in process and the 2012 review of which it was an update, and the output of the search and sifting, using very similar PICOs and search terms, for the NICE guidelines.

Some of the trials that Williams et al. (2020) included appeared in exclusion lists for the NICE review of PMPs, with reasons, in the NICE documents, and this has allowed PaPaS to check. Most exclusions were because trial interventions were not deemed to be a PMP, defined by the committee in the protocol as any intervention that has two or more components including a physical and a psychological component delivered by trained people, with some interaction/coordination between the two. Exclusions were said to be usually because the intervention was either psychological or physical but not both, or included psychological and physical components but delivered in parallel with no interaction or coordination between them. A thorough check of the included and excluded trials for Williams et al. 2020 against the NICE review included and excluded trials showed very little overlap.

Two examples of trials incorrectly excluded as not pain management programmes, both found in Appendix I among the excluded trials follow. Here we provide the elaboration of reasons provided by NICE when Professor Williams queried the decisions. One is by Bliokas et al., published in 2007, whose title specifies “multidisciplinary chronic pain management groups”, and whose objectives in the Abstract also mention that the basis of the trial was a pain management programme to which a specific extra psychological component was added. The second is by Kole-Snijders et al., published in 1999, whose title and abstract do not mention pain management programmes, but were excluded on the basis of no physical intervention when participants each had 50 hours’ individual treatment by physical therapists.

These are just two examples where the rationale provided by NICE for exclusion does not seem clear or appropriate and they reflect a broader issue with this NICE review with numerous trials of

interventions that might reasonably be considered to be PMPs excluded. Further there are a number of examples of trials that were included by NICE but were excluded from the Cochrane review on the basis that they were too small or that the psychological component was delivered by non-psychologist professionals (10 trials) or laypersons (1 trial). Psychology is not common sense, and psychological therapy is not just talking. There are many studies showing the unsatisfactory nature of much communication between healthcare staff and patients (some reviewed in these guidelines), so to assume that any staff can teach psychology, when there is no suggestion that anyone could deliver medical care or physiotherapy or pharmaceutical advice, is problematic. It means that many of the NICE-included studies of psychological therapy, and of pain management programmes, are underpowered in terms of authentic delivery. The 2020 Cochrane review (Williams et al. 2020) required that a psychologist delivered the psychological content of the intervention, because there is very poor evidence that it can be adequately implemented without some training and this would reasonably be expected to impact effectiveness.

The overall result is that NICE excluded for incorrect reasons a large number of relevant trials. Many of these trials represent the kind of multicomponent pain management programmes delivered in many pain clinics and a few community settings in the UK. NICE instead included some trials that were underpowered either because of their size or the lack of suitable training of personnel. Where processes, particularly around early sifting of eligible studies, may not have been entirely transparent the result is that the committee is asked to make best sense of what was presented to them, without knowing what had been discarded or discounted at an earlier stage, having to take on trust that those presented were the most suitable trials on which to evaluate effectiveness. Further, the points made in the introductory section of this document also apply to the pain management programme section: about synthesis (splitting to an extreme rather than combining similar trials with similar outcomes for analysis), and about the prominence given to MID, a far more arbitrary quantity than is acknowledged by NICE, in evaluating efficacy.

The guideline says: "1.3.3 Consider acceptance and commitment therapy (ACT) or cognitive-behavioural therapy (CBT) for pain for people aged 16 years and over with chronic primary pain". The data on ACT are directly contradicted to the finding of the Williams 2012 review which comprehensively found that there was insufficient evidence to make such a claim and a recommendation for future research was only possible. Although there is significant enthusiasm for ACT based treatments in chronic pain that enthusiasm is not matched by the evidence. Perhaps of note is the reason for the difference. When robust criteria are applied the putative effects are missing. To illustrate this, in our recent review Williams (2020) reported: "For ACT, the finding of no evidence of efficacy or safety is at odds with several non-Cochrane reviews. Veehof 2011 combined 22 studies of ACT and mindfulness-based meditation, including non-randomised trials, and reported ACT to be "promising." In 2016, they updated this to 25 studies, all RCTs, and concluded "...that individuals with pain, in general, respond rather well to acceptance- and mindfulness-based interventions and that beneficial effects are retained after treatment" (Veehof 2016). Twenty-two of the studies included in that review did not meet our inclusion criteria. Twelve of the 25 are ACT studies. Nine of the 12 are not included here, seven because of small size, one because it was not delivered face-to-face, and one because it had no suitable control. One 2017 review included 11 RCTs (Hughes 2017). Their primary outcomes were acceptance of pain, quality of life and functioning. Their conclusions were for a positive effect of ACT on acceptance of pain and on functioning. Eight of the 11 are not included here, five because of small size, two because they were not delivered face-to-face, and one because it was a non-inferiority trial. A different 2017 review included 10 studies, had no accessible protocol, attempted no meta-analysis and simply reported on investigator-chosen endpoints (Simpson 2017). Their conclusions were positive for an effect on pain

acceptance. Seven of the 10 were not included here, four because of small size, two because they were not delivered face-to-face, and one because it was a non-inferiority trial.”

For those commissioning psychologically-based interventions for chronic pain in adults, or including such interventions in policy determinations, it is important to recognise that not all psychological treatments are the same. There is variety in the content, delivery, and clinical intentions of treatments, depending on their theoretical provenance. Interventions aim to reduce distress and disability, with or without a reduction in pain. The largest body of evidence we have supports the use, by trained psychologists, of CBT to produce benefits immediately after treatment and at follow-up of at least six months, rather than providing no treatment. The evidence is sufficient (i.e. large and of moderate quality) and unlikely to change with future studies. The overall effects are small or very small, meaning that the population benefit may be large, but more work is needed to identify which patients will individually benefit. There is development in other treatments such as ACT, emotional expression, and psychodynamic psychotherapy, but these remain experimental and monitoring of positive and negative outcomes is advisable.

Research recommendations

In key recommendations for research in the main guidelines document, p11, is the suggestion that optimum characteristics – by implication, the same for all chronic pain patients, an untenable assumption – of pain management programmes can be defined? There is an extensive empirical literature which has tried to do just that, using modelling, regression, dismantling trials and other methods. There is no ‘one size fits all’, as has been evidence for at least a decade. Encouraging further empty attempts to identify such a ‘one size’ is unhelpful.

Similarly, there is a body of existing literature which aims to identify risk factors that may represent barriers to successful management of chronic pain and on relaxation as a stand-alone treatment. Encouraging further simple attempts to answer these questions may lead to research waste (Glasziou & Chalmers 2018).

Concluding remarks

The above commentary outlines what we believe to be important specific limitations in the scope and methodological approach taken to the synthesis of the evidence used to develop this guideline. We have tried to specifically indicate where these issues may have unhelpfully impacted upon the decisions of the committee. The comments represent the combined expertise of our core editorial board. We recognise that it is challenging for the NICE committee and technical team to review or revise key methodological choices at this stage but we would encourage the committee to reflect carefully on how those choices may have influenced the evidence reviews and their subsequent recommendations when formulating the final guideline. PaPaS remains committed to working positively with NICE and other guideline developers to improve the standard of clinical care for people with persistent pain.

References

- Bliokas VV, Cartmill TK, Nagy BJ. Does systematic graded exposure In vivo enhance outcomes in multidisciplinary chronic pain management groups? *Clin J Pain* 2007;23:4: 361-74.
- Brinck EC, Tiippana E, Heesen M, et al. Perioperative intravenous ketamine for acute postoperative pain in adults. *Cochrane Database Syst Rev*. 2018;12:12:CD012033.
Pdoi:10.1002/14651858.CD012033.pub4
- Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 2013;346: f2304
- Fanelli D, Costas R, Ioannidis JP. Meta-assessment of bias in science. *Proc Natl Acad Sci U S A*. 2017 Apr 4;114:14:3714-3719. doi: 10.1073/pnas.1618569114.
- Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clin Trials*. 1997 Dec;18:6::568-79.
- IntHout J, Ioannidis JP, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-meta-analysis. *J Clin Epidemiol* 2015;68:860-9. DOI 10.1016/j.jclinepi.2015.03.017.
- Kole-Snijders AMJ, Vlaeyen JWS, Goossens MEJB, Rutten-van Mólken MPMH, Heuts PHTG, van Breukelen G, van Eek H. Chronic low-back pain: what does cognitive coping skills training add to operant behavioral treatment? Results of a randomized clinical trial. *J Cons Clin Psychol* 1999;67:6:931-44.
- Moore RA, Gavaghan D, Tramèr MR, Collins SL, McQuay HJ. Size is everything – large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. *Pain* 1998;78:209-16.
- Moore RA, Eccleston C, Derry S, Wiffen P, Bell RF, Straube S, et al. ACTINPAIN Writing Group of the IASP Special Interest Group on Systematic Reviews in Pain Relief, Cochrane Pain, Palliative and Supportive Care Systematic Review Group Editors. “Evidence” in chronic pain-establishing best practice in the reporting of systematic reviews. *Pain* 2010;1 50: 386–9.
- Moore RA, Straube S, Aldington D. Pain measures and cut-offs - 'no worse than mild pain' as a simple, universal outcome. *Anaesthesia* 2013a;68:4:400-12.
- Moore RA, Derry S, Aldington D, Cole P, Wiffen PJ. Amitriptyline for neuropathic pain in adults. *Cochrane Database Syst Rev* 2015; 7: CD008242.
- Moore 2015b, Moore RA, Derry S, Aldington D, Cole P, Wiffen PJ. Amitriptyline for fibromyalgia in adults. *Cochrane Database Syst Rev* 2015;:CD011824;
- Nguyen TL, Collins GS, Lamy A, Devereaux PJ, Daurès JP, Landais P, Le Manach Y. Simple randomization did not protect against bias in smaller trials. *J Clin Epidemiol*. 2017;84:105-113.
- Nüesch E, Trelle S, Reichenbach S, Rutjes AWS, Tschannen B, Altman DG, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;341:7766:241.
- Oberoi S, Yang J, Woodgate RL, et al. Association of Mindfulness-Based Interventions With Anxiety Severity in Adults With Cancer: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2020;3:8:e2012598. doi:10.1001/jamanetworkopen.2020.12598

Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet*. 1998 Jan 3;351:9095:47-52.

Roberts I, Ker K, Edwards P, Beecher D, Manno D, Sydenham E. The knowledge system underpinning healthcare is not fit for purpose and must change. *BMJ* 2015;350:h2463.

Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. *Lancet Child Adolesc Health*. 2018;2:3:223-228

Seers T, Derry S, Seers K, Moore RA. Professionals underestimate patients' pain: a comprehensive review. *Pain* 2018;159:5:811-818

Straube S, Derry S, McQuay HJ, Moore RA. Enriched enrollment: definition and effects of enrichment and dose in trials of pregabalin and gabapentin in neuropathic pain. A systematic review. *Br J Clin Pharmacol* 2008; 662:266-75

Straube S, Moore RA, Paine J, Derry S, Phillips CJ, Hallier E, McQuay HJ. Interference with work in fibromyalgia - effect of treatment with pregabalin and relation to pain response. *BMC Musculoskeletal Disorders* 2011a; 12:125

Straube S, Moore RA, Paine J, Derry S, Phillips CJ, Hallier E, McQuay HJ. Interference with work in fibromyalgia - effect of treatment with pregabalin and relation to pain response. *BMC Musculoskeletal Disorders* 2011;12:125

Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. *PLoS One*. 2011;610:e25491.

Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One*. 2013;8:3:e59202. doi: 10.1371/journal.pone.0059202.

Vickers AJ, Vertosick EA, Lewith G, et al. Acupuncture for Chronic Pain: Update of an Individual Patient Data Meta-Analysis. *Journal of Pain* 2018; 19:5:455-474

Williams AC de C, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database SystRev*2020;8: CD007407. DOI: 10.1002/14651858.CD007407.pub4.

Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. *Crit Care*. 201317:1:R2. doi: 10.1186/cc11919.