ACR OA Guideline Development Process

Knee and Hip

1. Literature searching frame work

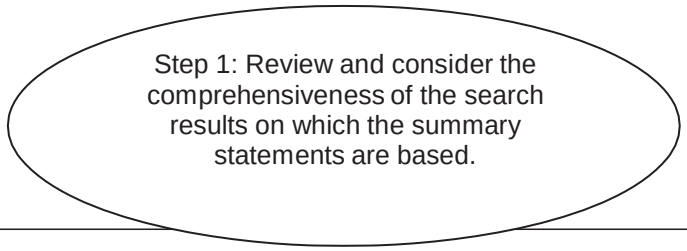
Literature searches were developed based on the scenarios. A comprehensive search strategy was used to guide the process for identifying the relevant literature. The searching approach was systematic and followed a process to allow for transparency and reproducibility. A minimum systematic search for relevant evidence included the medical databases MEDLINE, CINAHL and The Cochrane Library. References for each question were imported into a Reference Manager database and all data were independently screened by 2 reviewers.

The literature was searched to identify the most recent systematic review (SR) or meta-analysis (MA) for each outcome and comparison, with the appropriate data for analysis. If no systematic review was found, we searched for the most recent RCT which met the CONSORT guidelines.

2. How to read the Summary of Findings documents

There are 4 steps in reading a "Summary of Findings" document, using acupuncture for patients with OA of the knee as an example.

Step 1: Search results



We found one meta-analysis (Manheimer, 2007), which pooled results from two RCTs (Berman, 2004; Scharf, 2006) that compared acupuncture to usual care. Other systematic reviews were found: some had an older search (Jamtvedt, 2008; Bjordal, 2007), two were written in German (Joos, 2007; Schuller, 2008), one looked at laser acupuncture (Schuller, 2008), one only looked at acupuncture compared to sham (White, 2007) and one was an exploratory review (Vas, 2007).

An "Intervention description" is also included to provide further detail about the intervention.

Step 2: GRADE Summary of findings table

Step 2: The GRADE table provides a summary of the risk and the quality of evidence corresponding to the benefits (pain and function) and harms of the treatment that have been derived from the evidence.

Acupuncture versus usual care for knee OA

Patient or population: patients with knee OA
Intervention: Acupuncture versus usual care

Outcomes	Illustrative comparative risks [1] (95% CI)		Absolute difference	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	
	Assumed risk Usual care	Corresponding risk Acupuncture					
Benefits							
Pain WOMAC. Scale from: 0 to 100. (follow-up: median 6 months)	35%	56% (50% to 61%)	21%	2	1051 (2) [1]	⊕⊕⊕⊕ high [2]	5 (4 to 6)
Function WOMAC. Scale from: 0 to 100. (follow-up: 6 months)	36%	54% (49 % to 60%)	18%	1	1051 (2)	⊕⊕⊕⊕ high	5 (4 to 7)
Harms							
Hematoma number of patients reporting reaction (follow-up: 6 months)	0.3%	4% (0.5% to 27%)	3.7%	11.63 (1.52 to 88.93) [3]	642 (1) [3]	⊕⊕⊕⊕ high	31 (4 to 641) [4]

[1] 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).
[2] The meta-analysis pooled 9 studies which gave heterogeneous results. Only two studies compared acupuncture to usual care in knee OA at 6 months after treatment (Berman 2004 and Scharf, 2006)
[3] The study by Scharf was funded by a German insurance agency. However, it did not participate in the design, analysis or reporting of the study. Therefore, the score for the quality assessment was not reduced.
[4] Only 3 RCTs reported adverse events such as hematoma. Two studies reported the percentage of patients with hematoma and one of them against usual care (Scharf 2006)
[5] Chris Cates' calculator.

Step 2: A visual summary of the absolute number of patients that experience the benefits and harms are expressed using a 10x10 grid of 100 faces. Your judgment on these benefits and harms, quality of evidence, and values and preferences are requested; as well as your overall recommendation.

Acupuncture versus usual care for knee OA

Issue
Evidence from SRs and trials
Judgment (panel)

1. Balance between desirable and undesirable effects

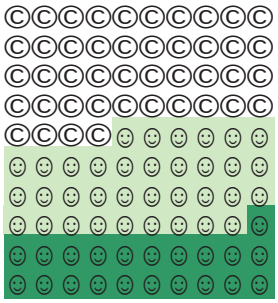
Chance: Improving pain (6 mos.)

NNT: 5

44% © Don't improve

35% ☺ Improve with or without Rx

21% ☺ Benefit with Rx-7



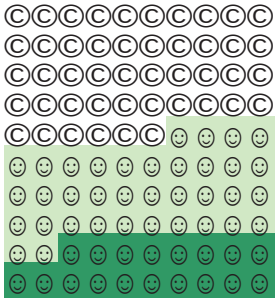
Chance: Improving function (6 mos.)

NNT: 5

46% © Don't improve

36% ☺ Improve with or without Rx

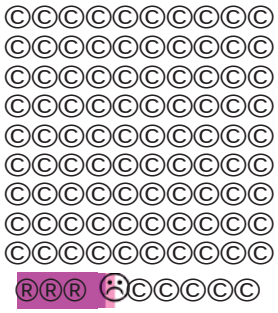
18% ☺ Benefit with Rx-7



We rounded up the % of improvement to put it in the faces for the outcomes of efficacy (pain, function).

The faces are shaded different colors to show the outcomes of 100 including adverse events, or harms. For the purposes of this visual aid, we have rounded up the percentages for the efficacy outcomes such as pain and function.

However, if the percentage is 0.5% or less for a **harms** outcome, percentages are **not** rounded up, resulting in partially colored faces

Chance: Hematoma (6 mos.)	
NNH: 31	
96% © Avoid bad outcome	
0.3% ® Bad outcome with or without Rx	
3.7% ® Harmed by Rx-7	

Step 3: GRADE Evidence profile

See Tables at the end of the document.

Step 4 Other recommendations

Group	Recommendation
EULAR	None
OARSI	Acupuncture may be of symptomatic benefit in patients with knee OA. ES (95% CI) = 0.51 (0.23 to 0.79) for pain and function and a NNT of 4. The strength of recommendation was 59% (47% to 71%).

Step 4: Consider the views of other recommendations that have been previously developed.

References

Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for osteoarthritis of the knee. Annals of Internal Medicine 2007; 146 (12):868-877.

Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. Ann Intern Med 2004; 141(12):901-910.

Scharf HP, Mansmann U, Streitberger K et al. Acupuncture and knee osteoarthritis: a three-armed randomized trial. Ann Intern Med 2006; 145(1):12-20.

Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, et al. Acupuncture in patients with osteoarthritis of the knee: a randomised trial. Lancet 2005;366:136-43.

References to all the studies used are provided.

Tables from Step 3: The GRADE evidence profile tables are included at the end of the document. They provide the details of the quality assessment and effect size that are summarized in the Summary of Findings Table in Step 2.

GRADE Evidence profile

Author(s): Jessie McGowan, Karine Toupin April
Date: 2008-10-20
Question: Should Acupuncture versus usual care be used for knee OA?
Bibliography: Manheimer 2007

Quality assessment							Summary of findings				Quality	Importance
							No of patients		Effect			
No of studies	Design	Limitations	Inconsis-tency	Indirectness	Imprecision	Other consider-ations	Acu-puncture	Usual care	Relative (95% CI)	Absolute		
Pain (follow-up median 6 months; measured with: WOMAC; range of scores: 0-100; Better indicated by less)												
2 [1]	randomised trial	no serious limitations[2]	no serious inconsis-tency	no serious indirectness	no serious imprecision	none[3]	520	531	1.58	SMD -0.52 (-0.66 to -0.39)	⊕⊕⊕⊕ HIGH	CRITICAL
Function (follow-up 6 months; measured with: WOMAC; range of scores: 0-100; Better indicated by more)												
2	randomised trial	no serious limitations	no serious inconsis-tency	no serious indirectness	no serious imprecision	none	520	531	1.49	SMD -0.45 (-0.59 to -0.32)	⊕⊕⊕⊕ HIGH	CRITICAL
Hematoma (follow-up 6 months; number of patients reporting reaction)												
1 [4]	randomised trial	no serious limitations	no serious inconsis-tency	no serious indirectness	no serious imprecision	none	12/326 (3.7%)	1/316 (0.3%)	11.63 (1.52 to 88.93)[5]	3% more (0.2% to 26% more)	⊕⊕⊕⊕ HIGH	IMPORTANT

[1] The meta-analysis pooled 9 studies which gave heterogeneous results. Only two studies compared acupuncture to usual care in knee OA at 6 months after treatment (Berman 2004 and Scharf, 2006).

[2] More patients discontinued treatment in the usual care group than in the acupuncture group. This may indicate that users of acupuncture have high expectations concerning its benefits according to the authors (placebo effect). The score for the quality assessment was not reduced.

[3] The study by Scharf was funded by a German insurance agency. However, it did not participate in the design, analysis or reporting of the study. Therefore, the score for the quality assessment was not reduced.

[4] Only 3 RCTs reported adverse events such as hematoma. Two studies reported the percentage of patients with hematoma and one of them against usual care (Scharf, 2006).

[5] We calculated the RR using the data in GRADE (ratio of number of events in exposed vs. number of events in non-exposed).

Appendix 1

GRADE approach explanation

Search

A systematic search was conducted for each intervention identified by EULAR. For efficacy and adverse events (AE), meta-analyses and systematic reviews were searched first, followed by randomized controlled trials. The highest level of evidence was prioritized. In the presence of uncommon and/or rare AEs, we also searched observational studies if necessary.

Grade quality assessment

Quality options are: high, moderate, low or very low for each outcome. RCTs start at high; observational studies start at low. A study downgraded for any reason cannot be upgraded.

Factors	Assessment	Effect on quality	Score
1. Study limitations	Limitations in design that suggest high likelihood of bias (factors may vary according to type of intervention, but may include, for example: allocation concealment, blinding for subjective outcomes, failure to use intention to treat analysis, loss to follow-up, trial stopped early).	None, -1 (serious) or -2 (fatal flaws)	
2. Consistency of results	Significant heterogeneity without explanation or problems with subgroup analyses.	None, -1	
3. Directness of evidence	Indirectness in population, intervention, control or outcomes that limits applicability because relative or absolute effects may differ across these factors.	None, -1, -2	
4. Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect.	None, -1	
5. Publication bias	High probability of publication bias.	None, -1	
Increasing quality of evidence (for observational studies only)			
6. Magnitude of effect	RR > 2 or RR <0.5 with no plausible confounders (+1) RR >5 or RR <0.2 with no threats to validity (+2)	None, +1, +2	
7. All plausible confounding would work to underestimate treatment effect	Only studies with no important threats to validity should be upgraded.	None, +1	
8. Dose-response gradient	The presence of a dose-response gradient may increase our confidence in the findings of observational studies and thereby increase the quality of evidence. Only studies with no threats to validity (not downgraded for any reason) can be upgraded.	None, +1	
TOTAL			

The strength of recommendation is based on judgment on 3 issues

Issue	Comment
1. Balance between desirable and undesirable effects	Larger the difference between the desirable and undesirable effects, more likely a strong recommendation warranted. Narrower the gradient, more likely weak recommendation warranted.
2. Quality of the evidence	Higher the quality of evidence, more likely a strong recommendation warranted.
3. Values and preferences	More variability in values and preferences, or more uncertainty in values and preferences, more likely weak recommendation warranted.

Appendix 2

Glossary (from GRADE)
This glossary is partially based on the glossary of Cochrane Collaboration terms and the Users' Guides to the Medical Literature.

Absolute risk reduction (ARR): Synonym of the risk difference (RD). The difference in the risk between two groups. For example, if one group has a 15% risk of contracting a particular disease, and the other has a 10% risk of getting the disease, the risk difference is five percentage points.

Confidence interval (CI): A measure of the uncertainty around the main finding of a statistical analysis. Estimates of unknown quantities, such as the RR comparing an experimental intervention with a control, are usually presented as a point estimate and a 95% confidence interval. This means that if someone were to keep repeating a study in other samples from the same population, 95% of the confidence intervals from those studies would contain the true value of the unknown quantity. Alternatives to 95%, such as 90% and 99% confidence intervals, are sometimes used. Wider intervals indicate lower precision; narrow intervals, greater precision.

Continuous outcomes: The term 'continuous' in statistics conventionally refers to data that can take any value in a specified range (such as Height, weight and blood pressure).

Dichotomous outcomes: Each individual's outcome may have one of only two possible categorical responses (e.g. dead or alive, myocardial infarction or no myocardial infarction, etc.).

Minimal Clinically important difference (MCID): The smallest effect of a treatment that patients perceive as beneficial and that (in the absence of unacceptable side effects, inconvenience, and costs), mandates that the treatment be given. (A dictionary of Epidemiology, Fifth Edition, Miquel Porta, 2008).

Mean difference (MD): the 'difference in means' is a standard statistic that measures the absolute difference between the mean value in the two groups in a clinical trial. It estimates the amount by which the treatment changes the outcome on average. It can be used as a summary statistic in meta-analysis when outcome measurements in all trials are made on the same scale. Previously referred to as weighted mean difference (WMD).

Minimally important difference (MID): The smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and that would lead the patient or clinician to consider a change in the management.

Number needed to treat (NNT): An estimate of how many people need to receive a treatment before one person would experience a beneficial outcome. For example, if you need to give a stroke prevention drug to 20 people before one stroke is prevented, then the number needed to treat to benefit for that stroke prevention drug is 20. It is estimated as the reciprocal of the risk difference.

Number needed to harm (NNH): A number needed to treat to benefit associated with a harmful effect. It is an estimate of how many people need to receive a treatment before one more person would experience a harmful outcome or one fewer person would experience a beneficial outcome.

Odds ratio (OR): The ratio of the odds of an event in one group to the odds of an event in another group. In studies of treatment effect, the odds in the treatment group are usually divided by the odds in the control group. An odds ratio of one indicates no difference between comparison groups. For undesirable outcomes an OR that is less than one indicates that the intervention was effective in reducing the risk of that outcome. When the risk is small, the value of odds ratio is similar to risk ratio.

Risk: The proportion of participants experiencing the event of interest. Thus, if out of 100 participants the event (e.g. a stroke) is observed in 32, the risk is 0.32. The control group risk is the risk amongst the control group. The risk is sometimes referred to as the event rate, and the control group risk as the control event rate. However, these latter terms confuse risk with rate.

The assumed risk: The proportion of participants experiencing the event of interest (such as a clinically important improvement in pain and function or some type of harm) in the control group.

The corresponding risk: The proportion of participants experiencing the event of interest (such as a clinically important improvement in pain and function or some type of harm) in the treatment group. It is based on the relative magnitude of an effect and assumed (baseline) risk.

Relative risk (RR): Synonym of risk ratio. The ratio of risks in two groups. In intervention studies, it is the ratio of the risk in the intervention group to the risk in the control group. A risk ratio of one indicates no difference between comparison groups. For undesirable outcomes, a risk ratio that is less than one indicates that the intervention was effective in reducing the risk of that outcome.

Review Manager (RevMan): Software used for preparing and maintaining Cochrane systematic reviews. RevMan allows you to write ad manage systematic review protocols, as well as complete reviews, including text, tables, and study data. It can perform meta-analysis of the data entered, and present the results graphically.

Standardized mean difference (SMD): The difference between two estimated means divided by an estimate of the standard deviation. It is used to combine results from studies using different ways of measuring the same continuous outcome, e.g., pain. By expressing the effects as a standardized value, the results can be combined since they have no units.
0.2 = small effect
0.5 = moderate effect
0.8 = large effect (Cohen 1988)

Strength of a recommendation: the degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

Appendix 3

Calculations for NNT and assumed and corresponding risks

For continuous outcomes (such as pain and function rated on a visual analogue scale 0-100 mm), the NNT and assumed/corresponding risks were calculated using the Wells calculator. This calculator takes into account the minimal important difference (MID) and the effect size (SMD). The MID is calculated using the ratio between the MCID (15%) and the standard deviation of the most representative trial in a review (or the only trial if we chose an RCT). The MID was used to determine which patients had a significant improvement and which did not, therefore dichotomizing the outcome of interest in order to find the NNT and assumed and corresponding risks.

For dichotomous outcomes (such as number of patients presenting a certain type of harm), the NNT was calculated using the Chris Cates calculator (formula: 1 / [control group event rate*(1-rr)], where control group event rate is taken as the event rate in the control group of the trial).

The NNT were calculated for the duration of each individual trial (see appendix for evidence profile).

Calculations for the relative risk (RR)

For continuous outcomes, the RR was calculated as the ratio between the corresponding risk and the assumed risk (which were previously calculated using the Wells Calculator).

For dichotomous outcomes, the RR was usually calculated in the reviews (number of events in the treatment group/number of events in the control group).