Stakeholders’ preferences for osteoarthritis interventions in health services: A cross-sectional study using multi-criteria decision analysis

Jason Chua, Paul Hansen, Andrew M. Briggs, Ross Wilson, David Gwynne-Jones, J. Haxby Abbott

* Corresponding author. E-mail address: haxby.abbott@otago.ac.nz (J.H. Abbott).

ARTICLE INFO

Keywords:
Knee osteoarthritis
Multi-criteria decision analysis
Conjoint analysis
Interventions
Health system

SUMMARY

Objectives: To combine cross-sectoral stakeholders’ preferences over interventions for knee osteoarthritis (OA) with guideline recommendations and evidence about interventions, and to investigate if these preferences differ by stakeholder group.

Design: A survey based on multi-criteria decision analysis was implemented whereby the stakeholders revealed the relative importance, represented as weights, of eight criteria for choosing or recommending knee OA interventions. Using data from an OA clinical guideline, 15 recommended interventions were rated on the criteria and ranked by their total scores, calculated by summing the corresponding weights. Associations between the weights and stakeholder groups were explored using regression analysis.

Results: Participants comprised 58 consumers with OA, 5 Māori health advocates, 79 healthcare providers, 24 policy-informants and 12 OA-researchers (N = 178; 63% female, [mean age ± SD] 54 ± 13 years). Mean weights on the eight criteria, in decreasing order of importance, are: recommendation: 19.0%; quality of evidence: 17.7%; effectiveness: 15.0%; duration of effect: 13.2%; risk of serious harm: 12.8%; risk of mild/moderate side-effects: 9.4%; cost: 6.6%; and accessibility: 6.3%. For first-, second- and third-line OA interventions respectively, all land-based exercise (total score = 71.7%), NSAIDs (topical) (74.2%) and total joint replacement (74.3%) were ranked first. At all care phases, the recommended core interventions of weight management and self-management education ranked between 11th and 15th (48.0%–56.0%). Regression analysis identified only small differences in weights (≤5.7%; p < 0.01) between stakeholder groups.

Conclusions: Not all recommended core interventions are preferred by cross-sectoral stakeholders, which may represent a barrier to their uptake. Stakeholders’ preferences do not appreciably differ by stakeholder group.

1. Introduction

Clinical practice guidelines (CPGs) for managing osteoarthritis (OA) consistently recommend exercise, education and weight loss (where indicated) as ‘core’ first-line interventions, followed by second- and third-line interventions such as drug therapies and other non-pharmacologic interventions and surgical interventions [1,2]. However, the recommended ‘core’ interventions are not systematically delivered to or taken up by patients [3–7], resulting in missed opportunities for potential health gains, a tendency to deliver low-value care and increased downstream health system costs without health gains [8,9]. One reason for poor delivery and uptake may be incompatibility between the interventions recommended in CPGs and the preferences of patients and other stakeholders with respect to interventions they want or would recommend [10,11].

Stakeholders’ preferences for health interventions [12], especially in primary care settings [13], play an important role in determining their uptake, highlighting the importance of widely engaging stakeholders in service co-design and care delivery recommendations [11]. And yet, when CPG recommendations are being developed, stakeholder engagement is often non-existent or, at best, very limited, with the preferences and contexts of stakeholders from across the sector often not adequately considered [14–19]. A better understanding of what matters to stakeholders, and which interventions more closely align evidence with...
stakeholders’ preferences for what they want or would recommend, may better support delivery of value-based care [8,9].

An important strategy in the co-design of models of service delivery may be the prioritisation of interventions based on the level of alignment between multi-disciplinary and cross-sectoral stakeholders’ preferences for criteria that matter to them, and the performance of interventions on those criteria. However, this approach has yet to be tested. Multi-criteria decision analysis (MCDA) is a robust methodology for revealing stakeholders’ preferences, with the potential to enhance the downstream implementation of evidence into policy and practice [20]. As the name implies, MCDA (i.e. ‘multi-criteria decision analysis’) is about decision-making based on considering multiple criteria (or objectives) together, in order to rank or prioritise the alternatives being evaluated (here, OA interventions). In effect, MCDA is a structured decision-making process that involves measuring the inevitable trade-offs when choosing between alternatives. Using choice-based surveys, stakeholders’ preferences for criteria can be quantified to reveal their relative importance (weight), as well as the value placed on the alternatives, by which they can be ranked relative to each other.

In recent times, the use of MCDA has become increasingly widespread in health care research [21,22]. MCDA has been used to explore OA patients’ preferences for physical activity [23], patients’ drug preferences, [24] and healthcare providers’ treatment choices for people with OA [25]. However, MCDA has not yet been used to explore stakeholders’ preferences for OA interventions across a health system, which may have the potential to assist in co-design of system-wide health service models. This study uses MCDA to: (i) discover the relative importance of criteria relevant to stakeholders when choosing or recommending knee OA interventions; (ii) use this preference information (criteria and weights) to rank (prioritise) a wide range of interventions from a recent CPG for first-, second- and third-line OA care; and (iii) to investigate if preferences differ by stakeholder group.

2. Methods

2.1. Design

This cross-sectional study followed six stages for conducting MCDA

![Flow diagram of the study by stage, primary activity and outcomes for each stage.](https://example.com/flowDiagram.png)
(Fig. 1), aligned with MCDA good practice guidelines [26]. Ethics approval was obtained from the Human Research Ethics Committees of the University of Otago (D16-329) and Curtin University (HRE2018-0276). The research was undertaken in New Zealand (NZ) and Australia between October 2017 and June 2018 and is reported here in accordance with the STROBE statement [27] (Supplement 1).

2.2. Sampling and recruitment

Convenience and snowball sampling were used to invite the survey participants: consumers (with a diagnosis of OA or symptoms consistent with the NICE criteria for OA [28]), healthcare providers (clinicians delivering care to people with OA; e.g. general practitioners, orthopaedic surgeons, physiotherapists), policy-informants (OA-related health policy, strategy, health service workforce coordination, delivery or funding for OA management, and consumer advocacy or representation for OA), Maori health (with an active interest in advocacy or consumer representation for Maori health) and OA researchers (having published at least one academic article related to OA). A characterisation of these groups is reported elsewhere [29].

Maori, the indigenous peoples of NZ, are recognised as a priority group with respect to enhancing healthcare equity and equality of health outcomes [30]. Because there is only a relatively small pool of eligible policy-informants and OA researchers within NZ, we included participants from Australia from these groups – justified on the basis of the two countries' proximity and the similarity of their health systems (public-private mix, with patient co-payments [31]) and socio-cultural characteristics [10]. The Qualtrics platform (Provo, USA) was used to screen and collect participants' demographic data.

2.3. Recruitment sources

Healthcare-provider participants from across the public and private health sectors were sampled from a NZ business directory and an online health-service database (https://healthpages.co.nz/). Health practitioner organisations, government and non-government organisations, healthcare delivery organisations and advocacy groups were asked to distribute invitations to participate to healthcare providers, policy-informants and Maori health advocates. OA researchers were initially identified using an online database (http://expertscape.com/) and screened for potential eligibility by three authors (JC, AMB, JHA), from which a convenience sample was invited to participate.

2.4. Stage 1: identifying the criteria and their levels for selecting OA interventions

Stage 1 was informed by our earlier mixed-methods study [32] whereby multi-disciplinary and cross-sectoral stakeholders identified nine criteria influencing their choice or recommendation of OA interventions in the NZ health system. These criteria were: Accessibility (travel or wait time to access the intervention), Cost (total financial costs relevant to the use or provision of healthcare for OA), Duration (duration of treatment effect), Effectiveness (magnitude of treatment effect), Recommendation (for using the intervention now), Risk of harm, Quality (quality of the evidence), Treatment Passivity and Immediacy of Treatment Effect. We excluded the last two criteria because in our previous study they were considered to be the least important to stakeholders [32]. After stratifying Risk of harm into Risk-Mild (risk of mild adverse effects) and Risk-Serious (risk of serious adverse effects), eight criteria were selected – which we deemed to be acceptable with respect to the time and cognitive burdens imposed on participants (in healthcare-related MCDA, the mean number of criteria is eight [33]).

Each criterion was specified with 2–4 levels of 'performance' – i.e. mutually-exclusive and exhaustive levels for differentiating between OA interventions in terms of their characterisation on each criterion. To support the definitions of the levels within each criterion, a literature search was undertaken to specify criteria performance (e.g. Cohen's d for effect size) and their intervals of performance, including plausible upper- and lower-bound levels [e.g. $d \leq 0.2$ (low); $0.2-0.5$ (moderate); $>0.5$ (high)]. The Accessibility criterion was considered to be context specific such that its levels were specified based on the judgement of three authors (JC, JHA, AMB). Key sources supporting the criteria specifications are reported in Supplement 2 (Table S1).

2.5. Stage 2: identifying the weights for the criteria and their levels

2.5.1. Choice-based survey

A choice-based survey administered by 1000minds software (www.1000minds.com) and implementing the PAPRIKA method [34] – an acronym for ‘Potentially All Pairwise Rankings of all Possible Alternatives’ – was used to determine the weights on the criteria and levels, representing their relative importance, for each participant and on average across all participants. The PAPRIKA method involves each participant being asked to answer a series of ‘pairwise-ranking questions’ based on choosing between two hypothetical OA interventions defined on just two criteria at a time and involving a trade-off (Fig. 2). The ‘pairwise-ranking questions’ are repeated with different combinations of the criteria, two at a time, until all possible questions are answered by each participant, either directly or indirectly. The consistency of each participant's answers was checked by three questions being repeated at the end of their survey. Real-time computer adaptation, based on applying the participant's previous answers and the logical property of 'transitivity' (e.g. if OA intervention 'X' is preferred to 'Y' which is preferred to 'Z', then 'X' must be preferred to 'Z'), serves to minimise the number of questions the participant is required to answer directly (with the remainder answered indirectly via transitivity). For technical details, see Hansen and Ombler (2008) [34].

From the questions answered directly by a participant, PAPRIKA uses quantitative methods to derive weights for the criteria and their levels, representing their relative importance to the participant. The weights for each participant were averaged across all participants to obtain mean weights for the sample. The weight for a level on a criterion represents the relative importance of the criterion overall and the level's degree of achievement or performance on the criterion [34]. The lowest level on a criterion represents the minimum/worst performance on the criterion and is assigned zero points. The highest level on a criterion represents the maximum/best performance of the criterion and the relative importance (weight) of the criterion overall. These weights sum across the criteria to 1 (100%).

To assist participants' understanding of the choice-based exercise and reduce their cognitive burden, two Supplementary materials, a 30-s YouTube instructional video and a definition sheet for the criteria, were included with the survey (Supplement 3). Participants were asked to complete the survey within two weeks, and reminders were sent to encourage completion.

2.5.2. Pilot-testing

Before being launched, the survey and accompanying Supplementary materials were pilot-tested with a convenience sample of 17 interviewees. The pilot-testing approach is included in Supplement 4.

A response verification and process evaluation was also undertaken to evaluate the extent to which survey participants’ results aligned with their overall expectations about the relative importance of the criteria and the usability characteristics of the choice-based survey (Supplement 7).

2.6. Stage 3: rating interventions on the criteria

2.6.1. Data extraction

Data for 75 OA interventions and evidence about their performance on the criteria established from Stage 1 were extracted from the 2018 Royal Australian College of General Practitioners guideline for hip and
knee OA (RACGP CPG) [35]. This information provided the most complete, rigorous, NZ-relevant and up-to-date evidence at the time to rate the interventions on six of the criteria: Duration, Effectiveness, Recommendation, Risk-Mild, Risk-Serious and Quality. Accessibility was estimated via a Delphi exercise involving a nationally representative panel of NZ OA researchers, independent from participants in our earlier study [32]. Cost was estimated using data and methods described in a systematic review [36]. A GRADE evaluation was conducted for total joint replacement (TJR), which was not included in the guideline evidence tables, to inform its performance on the criteria.

2.6.2. Rating performances

Each intervention was rated on the criteria and summarised into three ‘performance matrices’ for first-, second- and third-line OA care [37]. To align the CPG recommendations with first-, second- and third-line OA care, the authors (JC, AMB, JHA) developed a rubric to transform the guideline-assigned levels of recommendation (for any OA) into three categories for first-, second- and third-line OA care (Supplement 5; methods detailed in Supplement 2, page 14).

2.7. Stage 4: scoring the OA interventions and ranking them

A ‘total score’ for each intervention was calculated using a weighted-sum model [38]: the sum of the mean weights from Stage 2 corresponding to the intervention’s ratings on the criteria (Stage 3) for first-, second- and third-line care. The interventions were ranked (prioritised) according to their total scores, representing their alignment with participants’ preferences overall, for each phase of OA care.

Although we scored 75 interventions in the RACGP guideline [35], our analysis hereafter will focus on the 15 “recommended” interventions (p. 65): 3 first-line (core) interventions, 11 s-line (optional adjunctive and advanced pharmacological attempts) interventions, and 1 third-line (surgical) intervention (TJR). The interventions were scored and then ranked in decreasing order of priority for first-, second- and third-line care.

2.8. Stage 5: data analysis (criteria weights)

2.8.1. Response consistency

We assessed if inconsistent responses biased the weights by comparing the mean weights for the total sample and the mean weights of respondents who answered none of the three questions consistently.

2.8.2. Association with stakeholder group

To investigate if participants’ weights on the criteria differed by stakeholder group, fractional multinomial logistic regression (FMNL) [39–41] was performed using Stata (ver.15.1, StataCorp, TX), with the weights as dependent variables. The independent variables were stakeholder group (consumers, providers, policy-informants, OA researchers), controlling for NZ/Australian status, age, gender, working for a government agency, and years’ work experience in primary role. Model robustness was assessed using ordinary least squares (OLS) regression. Kendall’s W, ranging from no agreement to perfect agreement (0–1), was also used to assess if the relative importance of interventions differed by stakeholder group.

2.8.3. Selection bias

Selection bias in the criteria weights was explored in two ways. First, to determine if the FMNL regression results were influenced by unequal stakeholder group size, we performed an adjusted FMNL regression, weighting group size to achieve equal stakeholder representativeness. Second, we interrogated the relative importance of the interventions by assessing the level of agreement between the unadjusted and adjusted rank order of interventions weighted for equal representativeness using mean Spearman’s rank correlation.

2.9. Stage 6: uncertainty analysis (intervention scores)

2.9.1. Evaluating uncertainty in the intervention ratings

We explored the extent to which uncertainty in the ratings of the 15 guideline-recommended interventions for first-line care (Stage 3) on the criteria may have affected the interventions’ total scores and hence their ranking by examining the evidence used to assign ratings. We referred to the original studies cited in the RACGP CPG and determined plausible upper- and lower-uncertainty ratings on the criteria. The rules defining whether the criteria were up- or down-rated, on the basis of the evidence available are described in Supplement 2.

3. Results

3.1. Stage 2: identifying the criteria weights and process evaluation

3.1.1. Participants

Invitations were sent to 422 people, of whom 272 consented to participate; 178 (42.2%) completed the choice-based survey, and 147
Table 1  
Socio-demographic characteristics of the participants who completed the choice-based survey (N = 178).

| Socio-demographic characteristics | n (%) | Mean years experience ± SD [range] | Works in a government health agency n (%) |
|----------------------------------|-------|-------------------------------------|------------------------------------------|
| Gender                           |       |                                     |                                          |
| Male                             | 64 (36) | 18.1 ± 11.4 [1-42] | –                                        |
| Female                           | 114 (64) | 14.7 ± 11.0 [1-55] | –                                        |
| Region                           |       |                                     |                                          |
| Australia                        | 13 (7)  | 24.4 ± 12.1 [1-38] | –                                        |
| New Zealand                      | 165 (93)| 15.5 ± 11.1 [1-55] | –                                        |
| Age (years)                      |       |                                     |                                          |
| 18-34                            | 16 (9)  | 29 ± 2.9 [23-34] | –                                        |
| 35-54                            | 79 (45) | 45.7 ± 5.6 [35-54] | –                                        |
| 55 and over                      | 92 (52) | 63.7 ± 6.1 [55-82] | –                                        |
| Primary work area                |       |                                     |                                          |
| Consumers*                       | 58 (33) | 13.0 ± 11.7 [1-55] | 0 (0)                                    |
| Māori health advocates           | 5 (3)   | 17.6 ± 12.8 [1-31] | 3 (60)                                   |
| Providers                        | 79 (44) | 18.5 ± 10.8 [1-43] | 42 (54)                                  |
| Policy-informants†               | 24 (15) | 15.3 ± 11.5 [1-35] | 4 (17)                                   |
| OA researchers†                  | 12 (7)  | 13.8 ± 7.9 [2-30] | 10 (83)                                  |

*Years living with OA; Australian stakeholders n = 7; †n = 6.

The results of the Delphi exercise and the GRADE evaluation are summarised in Table 1. The weights for the criteria and levels are reported in Table 2. The relative importance of the criteria, in decreasing order of importance (weights in parentheses), are: Recommendation (19.0%), Quality (17.7%), Effectiveness (15.0%), Duration (13.2%), Risk-Serious (12.8%), Risk-Mild (9.4%), Cost (6.6%) and Accessibility (6.3%). Of the 178 participants who completed the survey, 145 (81%) answered at least two of the three repeated questions consistently. Participants spent a median of 4 s per question and answered a mean of 45 questions (range 20–92) each in total (median 15 min in total).

3.2. Stage 3: rating interventions on the criteria

The results of the Delphi exercise and the GRADE evaluation are summarised in Supplement 2. The assigned performance ratings across the criteria at each OA care phase are reported in the performance matrices (Supplement 6).

3.3. Stage 4: intervention scores and rankings

The total scores of the 15 guideline-recommended interventions are reported in Fig. 3, ranked in decreasing order of importance for first-, second- and third-line care. For first-, second- and third-line OA interventions respectively, ‘all land-based exercise’ (total score = 71.7%), ‘NSAIDs (topical)’ (74.2%), and ‘TJR’ (74.3%) were ranked first. Core interventions recommended in the CPG, ‘weight management’ and ‘self-management education’, were ranked in 11th to 15th place (48.0%–56.0%). The lowest ranked CPG-recommended intervention for first- and second-line care was ‘TJR’ and ‘self-management education’ (44.1% and 48.0% respectively); for third-line care, it was ‘mobilisation and manipulation’ (47.0%). Rating changes on the Recommendation criterion for second- and third-line care drove the change in total scores for ‘NSAIDs (topical)’ and ‘TJR’. The difference in total scores between the first- and seventh-ranked recommended interventions (the top half) at first-line care was 11.1%, while the difference in total scores between the eighth- and fifteenth-ranked interventions at first-line care was 15.8%.

Considering all 75 interventions (Supplement 7, Table S5), at first-, second- and third-line care, ‘Tai Chi’ was the highest ranked (total score = 76.9%), due to its strong performance on the Recommendation and Quality criteria. Several non-recommended interventions are more preferable to stakeholders than the core interventions ‘weight management’ and ‘self-management education’: e.g. nutraceuticals including ‘collagen’ (69.9%), ‘pycnonogenol’ (69.9%) and ‘curcuma’ (66.4%).

3.4. Stage 5: relationships between weights and stakeholder groups

3.4.1. Regression analysis

For the analysis of the weights on the criteria, we chose to combine the Māori health advocate group (n = 5) with the consumer stakeholder group due to a poor level of agreement previously reported for the Māori group [32]. Average partial effects (APEs) of the FMNL regression revealed weak evidence of associations between weights and stakeholder groups (Table 3). The APEs were relatively small after accounting for other socio-demographic characteristics (no more than 5.7%, aligning with the robustness check, Supplement 7 Table S11), suggesting that weights did not differ meaningfully by stakeholder group (or within consumer or healthcare provider groups – see Supplement 7, Tables S8–S10).

The level of agreement across groups by ranked interventions was very strong (N = 75, W = 0.990, p < 0.000; Supplement 7 Table S6).

3.4.2. Selection bias

The adjusted FMNL regression weighted for equal stakeholder group sample size, detected APEs that were statistically significant (p < 0.01). However, the APEs remained small (<5.1%), consistent with the unadjusted FMNL regression (Supplement 7, Table S7). We also calculated the correlation between the ranked interventions by importance (Stage 4), before and after adjusting weights for equal sample size; the correlation between the adjusted and unadjusted ranked interventions was very strong (r = 1.00, n = 75, p < 0.01). We interpret this to mean that selection bias had a negligible effect on the relative importance of the scored interventions.

3.5. Stage 6: uncertainty analysis

3.5.1. One-way sensitivity analysis

The uncertainty analyses at each phase of OA care (first-line care shown in Fig. 4) illustrate the aggregate effect of the uncertainty in the performance ratings assigned to the 15 guideline-recommended interventions. ‘All land-based exercise’ could plausibly achieve the highest score (relative to the other interventions) driven by the ratings on the Accessibility, Duration, Cost, Effectiveness and Quality criteria. For ‘NSAIDs (topical)’ the large uncertainty in its total score was driven by the neutral rating for the Recommendation criterion, and the evidence informing its ratings on the Risk-Serious and Effectiveness criteria. The uncertainty intervals for the interventions were the same for second- and third-line care, except for ‘TJR’ at second-line care, where the intervention’s total score varied by +4.2% and −6.5% – due to the uncertainty caused by the ‘neutral’ rating on the Recommendation criterion (dis-aggregated intervention scores are shown in Supplement 8).

4. Discussion

This study has systematically combined the preferences of stakeholders for OA interventions with CPG recommendations [35] and intervention performance data. Our main findings are that although the relative importance of the criteria differed by socio-demographic characteristics, these differences were small and did not translate to a meaningful effect on the relative importance of the interventions, and unequal group representation had little effect on the weights on the criteria. With respect to the first-line (core) interventions, ‘all land-based exercise’ aligned strongly with stakeholders’ preferences for first-line care; however, ‘weight loss’ and ‘self-management education’ are less preferred than most, if not all, recommended second-line interventions. ‘TJR’ is preferable but only for third-line care.

Our results show that stakeholders valued criteria often considered in systematic reviews of evidence, such as GRADE [42]. Participants valued...
Table 2
Criteria definitions and their sample mean criteria weights produced from the choice-based survey (N = 178), in decreasing order of relative importance. Criterion weights are at their best performance level sum to 1 (or equivalently, 100%).

| Criteria (most to least important) | Definition | Full sample mean weight (n = 178) | Mean weight by group |
|-----------------------------------|------------|----------------------------------|---------------------|
| Performance levels (worst to best) | Recommendation to use the intervention now |
| Strong against the intervention at first-line | 0.0 | 0.0 |
| Conditional against OA care. | 0.0647 | 0.0665 |
| Neutral | 0.1108 | 0.1073 |
| Conditional for | 0.1529 | 0.1462 |
| Strong for | 0.1904 | 0.1848 |
| Quality of the evidence about the intervention | The extent to which one can be confident that the effects of the treatment or service described are real. | 0.0 | 0.0 |
| Very low | 0.0587 | 0.0526 |
| Low | 0.1319 | 0.1145 |
| Moderate | 0.1765 | 0.1560 |
| High | 0.2077 | 0.1835 |
| Effectiveness of the intervention | The clinical effect of the intervention on pain. | 0.0 | 0.0 |
| Low (d < 0.2) | 0.0983 | 0.0865 |
| Moderate (d = 0.5) | 0.1501 | 0.1376 |
| High (d > 0.5) | 0.2558 | 0.2351 |
| Duration of the intervention effect | The duration of follow up demonstrating a meaningful effect on pain. | 0.0 | 0.0 |
| Short (up to 6hrs) | 0.0392 | 0.0421 |
| Short-medium (≤3 months) | 0.0748 | 0.0817 |
| Medium (3–12 months) | 0.1318 | 0.1506 |
| Cost of the intervention | Treatment side-effects that have significant medical consequences, e.g. lead to death, permanent disability or prolonged hospitalisation. | 0.0 | 0.0 |
| Risk of serious harm (Risk-Serious) | 0.0975 | 0.0864 |
| High (1 in 50 chance; >0.5%) | 0.1282 | 0.1325 |
| Medium (1 in 200 chance; 0.2%-0.5%) | 0.1765 | 0.1560 |
| Low (1 in 500 chance; <0.2%) | 0.2077 | 0.1835 |
| Risk of mild to moderate side effects (Risk-Mild) | Treatment side-effects that are not serious (see risk of serious harm). | 0.0 | 0.0 |
| High (3 in 4 chance; >50%) | 0.0527 | 0.0513 |
| Medium (2 in 4 chance; 25–50%) | 0.0941 | 0.0980 |
| Low (1 in 4 chance; ≤25%) | 0.1765 | 0.1560 |
| Cost of the intervention | Total financial costs relevant to be use or provision of healthcare for OA. | 0.0 | 0.0 |
| High (≥$1000 per month or ≥$15,000 one-off) | 0.0407 | 0.0450 |
| Medium ($100-$1000 per month or $1500-$15,000 one-off) | 0.0661 | 0.0735 |
| Low (<$100 per month or $1500 one-off) | 0.0313 | 0.0335 |
| Accessibility to the intervention | The extent to which the intervention can be accessed by people with OA. | 0.0 | 0.0 |
| Inconvenient travel, or wait time (>3 months) | 0.0627 | 0.0670 |
| Neither convenient or inconvenient travel, or wait time | 0.0618 |
| Convention travel, or wait time (<1 week) | 0.0526 |
| Mean weight of groups (n = 63) | 0.0 |
| Consumers (n = 79) | 0.0645 |
| Providers (n = 24) | 0.0618 |
| Policy-informants (n = 12) | 0.0720 |
| OA Researchers | 0.0697 |

* Interpolated criterion level using a Bézier curve; d = Cohen's d for effect size.
* Refer to Supplement 2 for a complete description of the criteria, including how interventions were rated on the criteria.
* The weights, multiplied by 100, are equivalent to per cent points and at their best level sum to 1 (100%).

Recommendation, Quality, Effectiveness, Duration and Risk-Serious approximately 5.7% Table 3) between weights and stakeholder group. Accounting for variance in group size made virtually no difference to the relative importance of the scored interventions (r<sub>c</sub> = 1, p < 0.01), while the level of agreement in intervention rankings across groups also confirmed that small weight differences were not meaningful (W = 0.990, p < 0.000). The outcomes of these different analytic approaches confirm our assertion that weights were not meaningfully heterogeneous with respect to the sample characteristics collected in this study. However, it would be prudent to re-evaluate these properties in a larger sample, as subgroup differences has been reported elsewhere.

Of the three core interventions, only ‘all land-based exercise’ aligned strongly with stakeholders’ preferences whereas ‘weight management’ and ‘self-management education’ did not (Fig. 3) due to poor performance on the Quality and Effectiveness criteria (Supplement 8 shows disaggregated total scores). This finding suggests that stakeholders’ preferences for the performance of the latter two interventions may contribute to their poor uptake in practice [43]. Weight management and the application of active self-management strategies for OA require substantial behaviour change for patients, which is often challenging to sustain [44]. We also note that the performance ratings on these interventions may not capture the broader benefits of engaging in them, such as reduction in the impact of other noncommunicable diseases which may feature alongside OA. Therefore, the value of these core interventions may be under-estimated in the current study.

For OA CPGs, broader stakeholder engagement is needed [2,17,45]. A number of studies have investigated consumer or provider preferences for the characteristics of OA interventions using MCDA methods [23–25, 46–48]. However, none has incorporated stakeholders’ preferences across a health system. Broader engagement may lead to more effective implementation strategies [19,49,50], particularly in primary care settings and in relation to policy change [13,51]. Yet, only about 2% of CPGs tailor guidelines to local health system user preferences. [52] Although stakeholders’ preferences did not meaningfully differ across the health system for OA interventions in the current study, the method used in this
Study may help cultivate more trustworthy decision-making and strengthen health systems by supporting decision-makers to focus on delivering what people value. Given that intervention success is influenced by interdependent factors across the health system, a multi-level approach to strengthening the health systems is needed [11,53]. For developers of health strategies, for example Models of Care [11] (currently absent in NZ [54]) or Models of Service Delivery [53], the approach outlined in this paper may help support better co-design and confirm consistency of cross-sectoral preferences, prior to upscaling such models nationally. Potential downstream effects could be realised through systemwide approaches such as better: health outcomes, patient experiences, and use of healthcare resources – the quadruple aim of value-based health care [55].

Strengths of our study include the mixed-methods design [32] used to inform the criteria selection and the independent source of evidence [35] – and provider experiences, and use of healthcare resources – and provider experiences, and use of healthcare resources (total joint replacement).

Fig. 3. RACGP guideline recommended OA interventions (N = 15) ranked by the full sample mean preference weights at first-, second- and third-line OA care.

Table 3
Average partial effects (APE) of the fractional multinomial logit model. APEs measure the change of a mean criterion weight, relative to the other criteria, given a change in the level of a socio-demographic characteristic. Negative coefficients indicate less importance. For example, healthcare providers, on average, place 4.3% (equivalently 0.043 APE) more importance on Recommendation, whereas policy-informants place 4.9% more importance on Quality and 4.7% less importance on Duration, relative to the other criteria and compared to consumers.

| Average Partial Effects | Socio-demographic characteristics | Recommend-ation to use the intervention now | Quality of the evidence | Effectiveness of the intervention | Duration of the intervention effect | Risk of serious harm | Risk of mild to moderate harm | Cost of the intervention | Accessibility to the intervention |
|-------------------------|---------------------------------|------------------------------------------|------------------------|----------------------------------|-----------------------------------|---------------------|-------------------------------|--------------------------|-----------------------------|
| Providers               |                                 | 0.043**                                 | 0.016                  | –0.003                           | –0.042*                           | –0.008              | 0.014                         | –0.010                   | –0.009                      |
| (ref: Consumers)        |                                 | (0.015)                                 | (0.012)                | (0.015)                          | (0.018)                           | (0.018)             | (0.014)                       | (0.01)                   | (0.012)                     |
| Policy-informants       |                                 | 0.028                                   | 0.049*                 | 0.016                            | –0.047**                          | –0.019              | –0.024                        | 0.010                    | –0.012                     |
| (ref: Consumers)        |                                 | (0.016)                                 | (0.019)                | (0.015)                          | (0.016)                           | (0.018)             | (0.017)                       | (0.01)                   | (0.015)                     |
| OA Researchers          |                                 | 0.057                                   | 0.007                  | –0.029                           | –0.030                            | 0.034               | –0.020                        | –0.010                   | –0.009                     |
| (ref: Consumers)        |                                 | (0.024)                                 | (0.02)                 | (0.019)                          | (0.026)                           | (0.026)             | (0.017)                       | (0.011)                  | (0.017)                     |
| Female                  |                                 | 0.018                                   | –0.015                 | –0.009                           | –0.014                            | 0.009               | 0.002                         | 0.000                    | 0.005                      |
| (ref: Male)             |                                 | (0.011)                                 | (0.01)                 | (0.009)                          | (0.012)                           | (0.011)             | (0.009)                       | (0.006)                  | (0.009)                     |
| Australian              |                                 | –0.028                                  | –0.012                 | 0.024                            | 0.025                             | 0.032               | –0.008                        | –0.020                   | –0.014                     |
| (ref: New Zealander)    |                                 | (0.019)                                 | (0.015)                | (0.015)                          | (0.015)                           | (0.018)             | (0.023)                       | (0.013)                  | (0.015)                     |
| Gov. employee           |                                 | –0.026*                                 | 0.003                  | 0.020                            | 0.014                             | 0.001               | –0.018                        | –0.004                   | 0.011                      |
| (ref: other employer)   |                                 | (0.013)                                 | (0.01)                 | (0.013)                          | (0.016)                           | (0.014)             | (0.01)                        | (0.006)                  | (0.009)                     |
| Age                     |                                 | 0.001*                                  | 0.001                  | –0.001                           | 0.000                             | 0.000               | 0.000                         | 0.000                    | 0.000                      |
| (at mean age 54yrs)     |                                 | (0.001)                                 | (0.000)                | (0.000)                          | (0.001)                           | (0.001)             | (0.000)                       | (0.000)                  | (0.000)                     |
| Work experience         |                                 | 0.001*                                  | 0.001                  | 0.000                            | 0.000                             | 0.000               | –0.001                        | 0.000                    | 0.000                      |
| (at mean exp. 16yrs)    |                                 | (0.000)                                 | (0.001)                | (0.000)                          | (0.001)                           | (0.001)             | (0.000)                       | (0.000)                  | (0.000)                     |
| Pairwise comparisons between stakeholder groups | Policy-informants | –0.014                                 | 0.033*                 | 0.019                            | –0.006                            | –0.011              | –0.038*                       | 0.020*                   | –0.003                     |
| (ref: Providers)        |                                 | (0.014)                                 | (0.016)                | (0.012)                          | (0.014)                           | (0.014)             | (0.017)                       | (0.008)                  | (0.012)                     |
| OA Researchers          |                                 | 0.014                                   | –0.009                 | –0.026                           | 0.012                             | 0.042               | –0.034**                      | 0.001                    | 0.000                      |
| (ref: Providers)        |                                 | (0.02)                                  | (0.019)                | (0.017)                          | (0.021)                           | (0.022)             | (0.012)                       | (0.008)                  | (0.014)                     |
| OA Researchers          |                                 | 0.029                                   | –0.042                 | –0.045**                         | 0.018                             | 0.053*              | 0.004                         | –0.020                   | 0.003                      |
| (ref: Policy-informants) |                                 | (0.023)                                 | (0.024)                | (0.016)                          | (0.024)                           | (0.025)             | (0.02)                        | (0.01)                   | (0.017)                     |

Standard errors are in parentheses.

Unadjusted * p < 0.05, ** p < 0.01; Gov = Government; exp = experience; yrs = years.
Separate regressions were run for the provider and policy–maker reference categories (italicised).

p = <0.001 ‘goodness-of-fit’ Wald Chi-square for each regression, indicating at least one of the coefficients has a significant impact on the criteria.
the criteria were not strictly non-overlapping and potentially non-independent; however, our criteria selection was informed by empirical data from local stakeholders, and we included pilot-testing and a response-verification and process evaluation to validate our choice-based survey. Our sample size was also modest, such that re-evaluation in a larger sample would be important to validate the findings to confirm that disease severity does not influence stakeholders’ preferences, or the associations in stakeholders’ sociodemographic characteristics and preferences we identified. The RACGP CPG also did not have an evidence-quality threshold for including evidence. This absence may have inflated the relative importance of some interventions, such as alternative medicines. The mean weights may also be at risk of bias due to the sampling method (which may underrepresent minority groups) and modest sample size. This study provides a framework for exploring cross-sectoral preferences for OA care in NZ due to the stakeholder-informed criteria selection, the representativeness of multi-level NZ stakeholders surveyed and the contextualised performance ratings for the Cost and Accessibility criteria. The framework is likely to be generalisable to other developed countries with similar health system funding schemes, access to health care and patterns of delivering lower-value OA care. However, the preference data should be interpreted cautiously due to the risk of sampling bias.

5. Conclusion

Stakeholders’ preferences for eight criteria influencing their choice of OA interventions in decreasing order of importance are: Recommendation, Quality, Effectiveness, Duration, Risk-Serious, Risk-Mild, Cost and Accessibility. Stakeholders’ weights did not appreciably differ by stakeholder group. Not all core recommended interventions are preferred by stakeholders; ‘all land-based exercise’ was highly valued for first-line OA care, but ‘weight-management’ and ‘self-management education’ are less preferred than most second-line interventions. The performance of TJR was most preferred for third-line OA care. These findings could help support greater delivery and uptake of value-based OA care across a health system.

Author contributions

JHA conceived the study idea; JC, PH, AMB & JHA conceived the study design; JC was responsible for acquisition of data, analysis, interpretation and initial draft manuscripts; RW & DGJ contributed to acquisition of data; JC, PH, AMB, RW, DGJ & JHA contributed to the interpretation of data, initial draft manuscripts, and approved the submitted version of the manuscript, for which they are accountable for its integrity.

Role of the funding source

This study is supported by the Health Research Council of New Zealand (grant 15/263). AMB was supported by a fellowship from Australian National Health and Medical Research Council (113548) and received travel support funding from the University of Otago (<$10,000).

Ethics approval and consent to participate

Approval for the research was obtained from the Human Research Ethics Committees of the University of Otago, New Zealand, (D16-329) and Curtin University, Australia, (HRE2018-0276).

Availability of data and materials

The datasets for the study are available from the corresponding author on request.

Declaration of competing interest

PH co-invented the software used in the study, which the software’s owners made available for the research. AMB was a member of the development group for the clinical guideline used in the research.

Acknowledgements

Thank you to the study participants for their time and expertise.
