Bridging the gap between evidence and what people value from osteoarthritis care in New Zealand using multi-criteria decision analysis (MCDA)

Jason Yu Yeong Chua

A thesis submitted for the degree of

Doctor of Philosophy

At the University of Otago, Dunedin,

New Zealand

November 2020

ABSTRACT

International clinical practice guidelines (CPG) for osteoarthritis (OA) consistently recommend core management strategies of exercise, weight-loss and self-management education. However, these interventions are not routinely delivered or taken up by people experiencing OA, resulting in a tendency to deliver low-value care. Tailoring the delivery of recommended OA care to the preferences of stakeholders in a health system may help support greater implementation of high-value OA care. However, little is known about the preferences of stakeholders for evidence and recommendations for OA care in a CPG.

The aim of this research was to establish a framework to prioritise knee interventions for managing OA and evaluate the relative importance of these interventions across the healthcare sector in New Zealand (NZ) using multi-criteria decision analysis (MCDA).

This research used a mixed-methods approach to develop the MCDA framework. The first stage involved focus group discussions to generate a thematic framework of what OA consumers, health care providers, policy-makers, Māori representatives and OA experts consider in their choice of knee OA interventions. A Delphi survey was used to verify the thematic analysis and rank the most important criteria concerning the characteristics of interventions; these findings informed the criteria selected in the second stage.

In the second stage, interventions were indirectly prioritised by systematically combining preference data with intervention performance data on the criteria. First, a survey involving trade-offs was used to measure stakeholders' preferences, represented as 'weights', within and between the criteria. Evidence for the performance of 15 recommended knee OA interventions were then

extracted from a CPG for hip and knee OA, and rated on the criteria according to their level of achievement (e.g. high, medium, low). To prioritise the interventions, a total score for each intervention was calculated by summing the weights associated with the intervention ratings, after which they were ranked by importance. Associations between the weights and stakeholder groups were explored using regression analysis.

Thematic analysis of data from six focus groups produced a framework comprising three overarching categories, consisting of characteristics of the: (i) intervention, (ii) consumer and (iii) health system. Participants identified and ranked nine characteristics of interventions; the most important eight were included in the MCDA framework. The choice-based survey revealed that stakeholders valued the intervention characteristics, in decreasing order of importance (weight): 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 side-effects (9.4%), Cost (6.6%) and Accessibility (6.3%). Total scores for the 15 guideline-recommended interventions revealed that for first-, second- and third-line OA care respectively, all land-based exercise (total score= 71.7%), NSAIDs (topical) (74.2%) and total joint replacement (74.3%) were ranked first. For first-, second- and third-line OA care, the recommended core interventions of weight management and self-management education ranked between 11th and 15th (48.0% to 56.0%). Regression analysis identified only small differences in weights (≤5.7%; p<0.01) between stakeholder groups. These findings suggest that stakeholders' preferences for the core interventions of weight management and self-management education represent a system-wide barrier to their implementation.

This research addresses an important knowledge gap concerning cross-sectoral stakeholders' preferences for knee OA interventions in a CPG. By inclusively, systematically and transparently incorporating what matters to people with evidence and recommendations in a CPG, the MCDA framework developed in this thesis can help support the design of patient-centred, high-value healthcare for people experiencing OA.

ACKNOWLEDGEMENTS

Thank you to my supervisors Haxby Abbott, Andrew Briggs and Paul Hansen for their tremendous patience, wisdom and encouragement. Also to my family, friends and in particular, my loving partner Dawnelle V. Clyne.

I also give my thanks to the following individuals for their support: Dr Cathy Chapple, Dr Richard Egan, Dr Robyn Fary, Dr Andrew Gray, Professor David Gwynne-Jones, Mr Mark Lipkin, Dr Greg Marcar, Ms Aroha Meilke, Ms Anupa Pathak, Ms Elena Meda Petrici, Associate Professor Aswo Safari, Dr Saurab Sharma, Professor Helen Slater and Dr Ross Wilson.

Included are the respondents who contributed to the snowball sampling, focus groups, Delphi survey, pilot testing and main study sample, and the organisations which helped to identify and invite individuals to participate in this research.

Thank you to the Health Research Council; Department of Surgical Sciences, University of Otago; and the Maurice and Phyllis Paykel Trust for financially supporting this work.

I extend my gratitude to everyone who helped make this thesis what it is.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND AND RATIONALE

1.1.1 Burden

Osteoarthritis (OA) is a disabling, long-term and highly prevalent condition. Globally, OA is the third highest musculoskeletal condition by years lived with disability (YLD) after low back pain and neck pain. Across all non-communicable diseases, OA is ranked the 13th highest contributor to global YLD, increasing almost 10% between 1990-2017.[1] The burden of OA is increasing globally, with the Australasian region characterised by the highest prevalence rates of OA in the world along with North America, North Africa and the Middle East.[1] OA is a major public health concern.

In NZ, about 10% of the population or 404,000 people are affected by OA.[2] In 2018, the total cost of arthritis (of which OA is the most common form) was estimated to be NZ$4.2 billion, with approximately NZ$993 million incurred as health system costs, and another NZ$3.2 billion through indirect costs, such as productivity, carer and efficiency losses.[3] Population growth and rates of population aging, obesity and deprivation are projected to increase in the future, with the subsequent economic burden of knee OA alone rising from NZ$199 million in 2013 to NZ$370 million by 2038.[4, 5] An intersection of similar factors has also raised concern about the projected financial burden of OA in Australia, too.[6] At an individual level, health losses in NZ due to knee OA over people's lifetimes are 3.4 quality adjusted life years (QALYs) per person.[7] This extrapolates to 467,240 QALYs across the adult population, representing an enormous personal and socioeconomic burden of OA in NZ. In the absence of a cure for OA and current growth, aging and obesity projections, OA will be a leading health burden in the 21st century. In NZ, a call to action has been recently voiced,[8] highlighting the urgent need to address this highly prevalent and disabling disease now, more than ever.

Pain is a common symptom of OA and typically restricts an individual's ability to do physical work - their functional capacity. For example, knee OA is associated with increased assistance in walking, stair climbing and other lower-extremity tasks more than any other medical condition among the elderly.[9, 10] These two factors alone may contribute to reduced living standards, retirement wealth and productive life years, compared to those living without OA.[11, 12] They may also have far reaching consequences for an individual's physical, psychological and social wellbeing. Ultimately, the disability caused by OA compromises an individual's quality of life such that managing the disease may require costly and potentially harmful interventions, such as total joint replacement (TJR), to address the pain functional disability caused by the disease. However, earlier management of the disease with conservative interventions may potentially delay or offset the need and downstream costs of performing such interventions.

1.1.2 Management and evidence-practice gaps

Exercise, self-management education programmes and weight loss, if appropriate, consistently feature among high quality clinical practice guidelines (see 2.9.1) for managing musculoskeletal pain conditions, including OA.[13-17] OA interventions are often classified in three stages. First-line (core) interventions - exercise, self-management education and weight loss, if appropriate, are consistently recommended through the disease continuum for long-term management.[14] Second-line optional or adjunctive management and advanced pharmacological attempts can be used to supplement core interventions (e.g. non-steroidal anti-inflammatory drugs and massage therapy). Third-line interventions, namely surgical interventions, should be referred only after all first- or second-line interventions have been trialled and failed.[13, 17]

While recommendations for OA care are consistent, the care delivered or utilised by people with OA is not consistent with guideline recommendations.[18, 19] The persistent failure to translate evidence into practice is a major threat to evidence-based practice. For OA, this manifests as a tendency to deliver lower-value interventions, such as drug interventions; underutilisation of conservative options, such as exercise and self-management education; and greater utilisation of surgical referral.[18-20] This results in lost opportunity for health gains and greater downstream health system costs, namely due to expensive joint replacement surgery.[21-23] Targeting the barriers and facilitators to implementing recommended OA care could help close this evidence-practice gap.

1.1.3 Context and fit

The implementation of recommended OA care is complex and influenced by many stakeholders at different levels in a health system. For example, from patient and provider attitudes and beliefs about OA (micro-level),[24, 25] through organisational and workforce characteristics (meso-level),[26] to the socio-political environment of a health system (macro-level).[27] No single overarching strategy has been identified to maximise implementation success, however strong evidence points towards incorporating context into the decision to adopt or commission interventions for implementation, particularly in primary care settings.[27, 28]

In particular, establishing the 'fit' of an intervention within a particular context or health system could help to maximise implementation success and drive greater delivery of high-value care.[29] Specifically, more closely combining evidence and recommendations about OA interventions with what matters to cross-sectoral stakeholders in a health system – such as the benefits, risk of harms and cost of interventions – could help support the delivery of high-value OA care. Pragmatically, this may help decision-makers, such as healthcare planners and funders, to identify which interventions offer the greatest implementation potential in a health system.

However, formulating recommendations for policy-making is a complex endeavour. Ideally, recommendations should represent the views and opinions of the people most relevant to OA care in a healthcare system (i.e. context), such as people with lived experience of OA, Māori advocates, providers, policy-makers and OA experts. Yet, engagement is often constrained with limited patient involvement, which may challenge the trustworthiness of recommendations, such is the case with clinical practice guidelines.[30-34]

There are currently gaps in the literature with respect to what matters to cross-sectoral stakeholders when choosing or recommending OA interventions in the NZ healthcare system, and if their preferences differ by group. Understanding what is of importance to cross-sectoral stakeholders could help support efforts to deliver high-value OA care.

Stakeholders also typically make complex decisions involving many considerations, or criteria, which more often than not conflict and compete against one another. One might ask:

- How can consensus be achieved?
- What are the most salient factors or criteria that need to be considered in the decision-making process?
- Do different types of stakeholders hold different perspectives?
- How do I determine which interventions are most desirable?
- What happens if new evidence, ideas or alternatives are introduced?
- How can all of this be achieved in a fair, transparent and robust process?

Such queries raise more questions about the transparency, fairness and rigor of the process used to construct recommendations for health policy. Multi-criteria decision analysis is a tool that can be used to address these important considerations.

1.1.4 Multi-criteria decision analysis

Multi-criteria decision analysis (MCDA) can enhance decision-making by incorporating subjective and objective data in a systematic and transparent process that identifies and weighs multiple evaluation criteria in order to prioritise different health care interventions.[35] In this thesis, subjective data refers to stakeholders' preferences for criteria that represent what matters to them when choosing or recommending OA interventions. Objective data refers to the evidence informing interventions' performance on the same criteria that matter to them. Using MCDA, the preferences of stakeholders, for criteria representing characteristics of OA interventions, are measured and combined with interventions' performance data (on the same criteria) to evaluate the relative importance of the interventions themselves (i.e. indirectly). In this way, MCDA can be used to prioritise not only a wide range of real, present interventions, but also hypothetical, potential future interventions. This makes MCDA a particularly useful tool for combining contextual factors such as the preferences of cross-sectoral stakeholders, with evidence about OA interventions (e.g. from a clinical practice guideline).

In recent times MCDA has become widespread in healthcare research.[36, 37] For example, it has been used to explore patients' preferences for physical activity,[38] surgical,[39] and drug[40] interventions. However, an evidence gap exists with respect to using MCDA to prioritise interventions in clinical practice guidelines for OA.

1.2 AIMS AND OBJECTIVES

The overall aim of this thesis is to establish a framework to prioritise knee interventions for managing OA and evaluate the relative importance of these interventions across the healthcare sector in New Zealand using MCDA. The objectives to address this aim are illustrated in Figure 1.

Objective 1

To identify what criteria are important to cross-sectoral stakeholders when choosing or recommending OA interventions in NZ and their relative importance.

Objective 2

2a. To evaluate and prioritise knee OA interventions according to stakeholders' preferences and their performance on criteria identified in Objective 1.

2b. To explore if stakeholders' preferences differ by group, and if stakeholders can be grouped by patterns in their preference weights.

Figure 1. Thesis objectives used to establish a MCDA framework to prioritise knee interventions for managing OA and evaluate the relative importance of these interventions across the healthcare sector in New Zealand.

1.3 THESIS STRUCTURE

The **introduction** (this Chapter 1) is an executive summary which introduces the reader to the topic and rationale for this thesis: why OA is an issue, what is known, what is the problem, and how we provide new knowledge addressing that problem, using multi-criteria decision analysis.

Chapter two, "Background" is a narrative literature review of: the burden of OA, key concepts in healthcare, recommendations for managing OA, evidence-practice gaps for OA care and strategies for closing the evidence-practice gap. It frames previous research around the barriers and facilitators to successful implementation of interventions and argues that establishing what OA interventions people want is an important research agenda.

Chapter three, "Introduction to the design and methods (Part A)," begins by introducing the reader to the methodological framework used to conduct the research in this thesis. It is followed by a broad overview of what is multi-criteria decision analysis and the key steps to 'good practice'. This lends to the subsequent order of the chapters in the thesis and the remainder of the current chapter which explains the key methods (focus group discussion, thematic analysis and Delphi survey) used to inform the first objective: identifying what characteristics of OA interventions are important to stakeholders when choosing or recommending OA interventions in NZ. Hence, the chapter is labelled '…Part A'.

Chapter four describes primary research conducted using the methods described in Chapter 3 to address Objective 1. In this chapter a mixed-methods study involving a qualitative component (focus groups) and a subsequent quantitative component (eDelphi survey) is used to identify what characteristics of OA interventions are considered important by cross-sectoral stakeholders when choosing or recommending OA interventions in NZ. It is followed by a discussion of the key research findings, limitations and conclusion of this research activity.

Chapter five, “Methods (Part B),” reconnects with the introduction to MCDA explained in Chapter 3, and then builds on the outcome of the previous chapter, i.e., the identified characteristics of interventions. It addresses the methodological approach used to achieve Objective 2 (see Figure 1). This chapter, therefore, begins with introducing the key considerations when constructing choice-based surveys – how to choose and structure appropriate criteria, so that interventions’ performances can be measured against them. Next, the method used to capture stakeholders’ preferences for the criteria – their weights – is described. Armed with this knowledge, the chapter then describes how the choice-based survey for this thesis was designed and pilot tested. The chapter then describes: the key sampling and recruitment methods; methods for analysis used to explore preference weights and sociodemographic characteristics; methods used to evaluate intervention performance on the criteria and rate interventions’ performances; and, the weighted-sum model, which combines preference weights against intervention ratings on the criteria (i.e., their performance) to compute intervention total scores.

Chapter six describes primary research conducted using the methods described in Chapter five. In this chapter, stakeholders’ preferences for the criteria are captured in a choice-based survey, and then combined with interventions’ performance ratings on the criteria to reveal the relative importance of OA interventions. The chapter reports the results of this process, and the results of the exploration of stakeholder group against preference weight data, including a sensitivity analysis, interpretation of findings, limitations and conclusions.

Chapter seven begins with a brief summary of the research outcomes achieved in chapters three through six. What follows is a meta-discussion about the implications of the original research reported in this thesis, the limitations of this research, and concluding remarks.

1.4 PUBLICATIONS AND PRESENTATIONS IN THIS THESIS

1.4.1 Published papers and abstracts

The primary research conducted in this thesis has been published in the peer-reviewed journal, Osteoarthritis and Cartilage OPEN:

- *Choosing interventions for hip or knee osteoarthritis – what matters? A mixed methods study.*[41] This publication represents the findings of this thesis which are discussed in Chapters 3 and 4.
- *Stakeholders' preferences for osteoarthritis interventions in a health system: a cross-sectional study using multi-criteria decision analysis study.*[42] This publication represents the findings of this thesis which are discussed in Chapters 5 and 6.

I was the first author on these publications and led the data collection, analysis and writing. I acknowledge the contributions of my supervisors and co-authors who helped me with editorial input. Approval was granted from the publishers to reproduce this work in this thesis.

Work arising from this thesis also appears in:

- *Implementing Models of Care for musculoskeletal conditions in health systems to support value-based care* [29]
- *The cost-effectiveness of recommended adjunctive treatments for knee osteoarthritis: Results from a computer simulation model.*[43]
- *The cost-effectiveness of recommended adjunctive treatments for knee osteoarthritis: Results from a computer simulation model* [44]

The following abstracts were published in the journal, Osteoarthritis and Cartilage:

- *Establishing the core factors considered by stakeholders choosing or recommending treatment options for hip or knee osteoarthritis in New Zealand* [45]

- *What attributes of interventions for osteoarthritis drive preferences? A discrete choice experiment involving cross-sectoral and multi-disciplinary stakeholder groups* [46]
- *Integrating values and preferences with the best available evidence: a multi-criteria decision analysis approach* [47]
- *The cost-effectiveness of recommended adjunctive osteoarthritis management options in New Zealand: results from a computer simulation model.*[43]

1.4.2 Conference presentations

I also acknowledge the editorial input of my supervisors and co-authors for the following accepted conference abstracts, posters and presentations:

- 2017 New Zealand and Australian Rheumatology Association conference; poster presentation, *Establishing the core factors considered by stakeholders choosing or recommending treatment options for hip or knee osteoarthritis in New Zealand*
- OARSI 2019 World Congress; poster presentations:
  - *What drives the choice of osteoarthritis interventions?*
  - *What interventions do people want for osteoarthritis?*
- 2019 Australian Physiotherapy Association; symposium speaker 2, *Which interventions for managing osteoarthritis do people want?*
- 2019 New Zealand Orthopaedic Association; presentation, *Which interventions for managing osteoarthritis do people want?*
- 2019 Health Services Research Australia and New Zealand; presentation, *What attributes of osteoarthritis interventions drive preferences?*
- 2019 Health Services Research Australia and New Zealand; presentation, *What interventions do people want for osteoarthritis?*

1.4.2.1Other presentations:

- Seminar (2016): *Management of osteoarthritis in New Zealand: a multi-criteria decision making approach*, Department of Surgical Sciences
- Webinar (2018): *Management of osteoarthritis in New Zealand: a multi-criteria decision-making approach.* Presented at International Society for Pharmacoeconomics and Outcomes Research (New Zealand chapter)
- Presentation (2018): *Management of osteoarthritis, the impact and burden of rising OA burden*, presented at the New Zealand Ministry of Health
- 3MT® winner (2019), *Which OA interventions do people want?* presented at the Health Sciences division round, University Otago
- 3MT® finalist (2019), *Which OA interventions do people want?* presented at the university-wide round, University Otago

Selected works not included in the thesis chapters are presented in the Appendix at the end of this thesis.

1.5 GLOSSARY

ACL	Anterior cruciate ligament
Alternatives	Also the same as ‘options’, ‘profiles’ or ‘choices’. Refers to the different options under consideration – e.g. interventions, investments, diseases or patients
Attributes	Used to described the characteristics of alternatives (e.g. interventions) in a choice-based decision
Choice-based survey	A survey involving trade-offs between hypothetical alternatives that are characterised by and differ in terms of their performance on criteria, which change with each successive survey question. Trade-offs may involve comparing two hypothetical interventions at-a-time (i.e. pairwise), or more. Choice-based surveys quantify stakeholders’ preferences for criteria as ‘weights’
CPG	Clinical Practice Guideline
Criteria	Decision-making criteria considered in a choice between multiple alternatives. Refers to the objectives or factors relevant to the overarching decision, which characterise real or hypothetical alternatives, in order to rank or select alternatives. E.g. the characteristics of OA interventions such as therapeutic benefit, safety and cost
Decision-makers	The stakeholders relevant to the decision problem such as OA consumers, healthcare providers, health policy-makers, health advocacy representatives and OA experts
FWER	Family Wise Error Rate
GBD	Global Burden of Disease
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
KL	Kellgren Lawrence
Levels	Within-criterion performance categories typically informed by their ‘best’ to ‘worst’ plausible range. E.g. high, medium and low.
Macro-level	Concerns health systems or organisations; health policy and strategy priorities; infrastructure and resource allocation; and socioeconomic factors; governance decisions
MCDA	Multi-criteria decision analysis

Meso-level	Health service design; clinical workforce capacity and capability; service delivery information systems; funding models; and clinical infrastructure models
Micro-level healthcare	The participation of the person in her/his care
MoC	Models of Care are principle-based, macro-level policies which influence the building blocks of health systems, such health policy and strategy priorities, resourcing, workforce configurations, service delivery options and health governance decisions.
MoSD	Models of Service Delivery operationalise aspects of MoC at the service (meso) level of the healthcare system, which concern health service delivery and workforce capacity building.
MSK	Musculoskeletal
NZ	New Zealand
OA	Osteoarthritis
OECD	Organisation for Economic Cooperation and Development
OR	Odds ratio
PAPRIKA	Potentially All Pairwise Rankings of all possible Alternatives
RACGP	Royal Australian College of General Practitioners
SMD	Standardised mean difference
TJR	Total joint replacement
TKR	Total knee replacement
Total score	Represents the overall value of an alternative.
Trade-offs	How much a decision-maker is willing to give up value in one criterion to achieve more on other criteria; by virtue, this means that a gain in the value of one criterion must be compensated by a loss in the value of other criteria. These choices are driven by their preferences, or value judgements
Weights	Also known as preference weights, criteria weights, value weights. Refers to the relative importance of criteria, representing decision-makers' preferences (or values). Weights are derived from choice-based surveys to inform preferences within and between criteria

CHAPTER 2: BACKGROUND

2.1 CHAPTER OVERVIEW

This chapter is a narrative review of the literature, for the purpose of aligning the reader to the current body of evidence about the burden, current management and practice gaps surrounding OA. This chapter will build toward the argument that the preferences of stakeholders across the health system for OA care influence their implementation. It will conclude by highlighting that there is an evidence gap with respect to what stakeholders value when choosing interventions, and which interventions stakeholders want for managing OA. Figure 2 shows a broad overview of the key components and subcomponents of this review. This chapter is a segue into the subsequent chapters of this thesis which describe the research approach in depth.

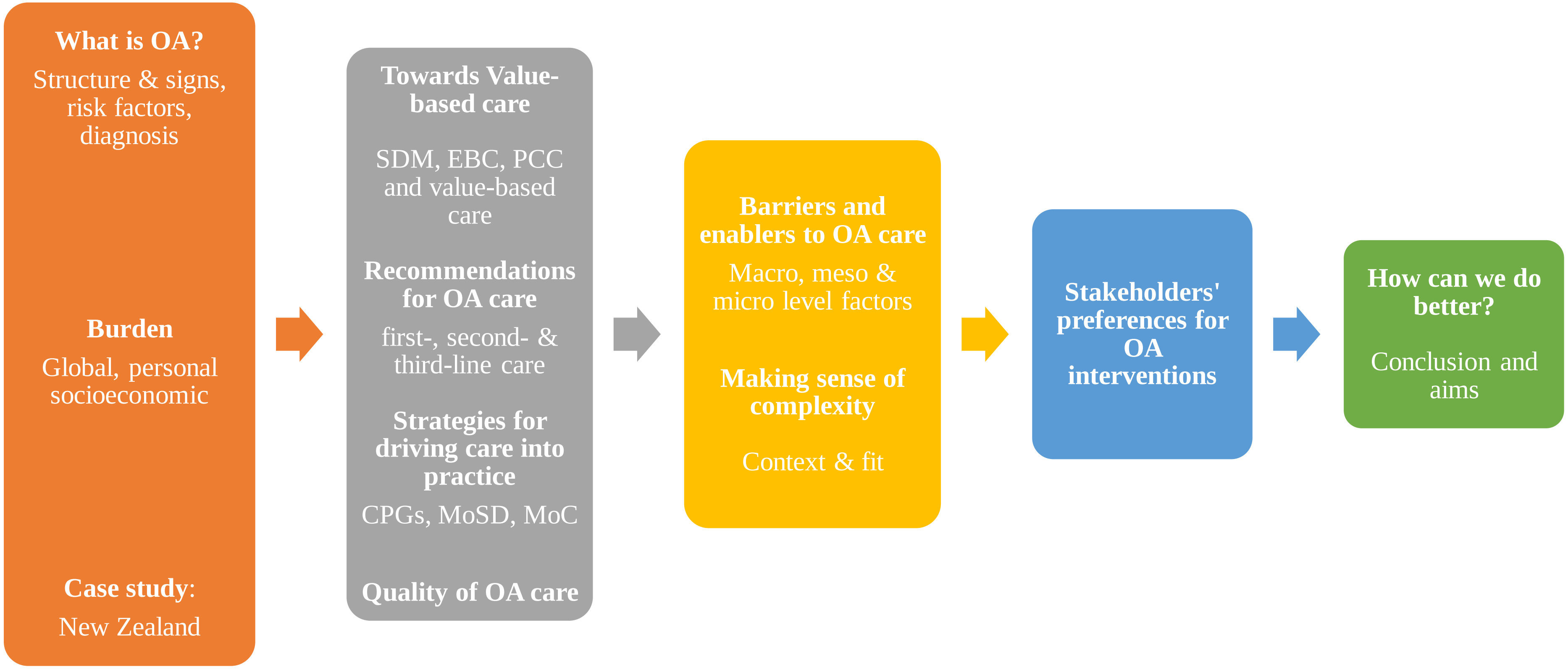


Figure 2. Schematic overview of the key components and subcomponents of the background chapter. CPG, Clinical practice guideline; EBC, Evidence-based care; MoC, Model of Care; MoSD, Model of Service Delivery; SDM, shared decision-making; PCC, Person Centred Care

2.2 WHAT IS OSTEOARTHRITIS?

OA is the most common joint disease worldwide and is a leading cause of pain and disability among adults.[1] It is a chronic, long-term non-communicable disease with no cure that is typically associated with older adults however, it also has concerning health implications for younger adults.[48] Worldwide, it is estimated that 303 million people live with OA.[1] Described as an active and complex disease, it is thought to be influenced by inflammatory, mechanical and metabolic factors. The combination of these factors cause an imbalance between the repair and destruction of articular and peri-articular tissues which ultimately erode the synovial joint. Virtually all the tissues in the synovial joint are involved, including hyaline articular cartilage, subchondral bone, synovium and soft-tissue structures, including ligaments, muscles and menisci.[49] OA commonly develops in the hips and knees and manifests as pain, joint stiffness with sequelae of disability, reduced participation (e.g. work, leisure) and reduced quality of life. However, many people with structural changes consistent with OA are also asymptomatic, suggesting that a diagnosis and measurement of impact should not be based on clinical data alone.[50]

2.2.1 Structural and symptomatic signs of OA

There is poor concordance between the structural signs and symptoms of OA. For example, Javaid et al. (2012)[51] found that structural features of knee OA examined via magnetic resonance imaging (MRI) and radiographic imaging were associated with pain, however, could not discriminate between people with painful and non-painful knees. In the Framingham Osteoarthritis Study,[52] 88% of participants who had structural changes detected by MRI had asymptomatic OA. These studies, however, used a disease-specific instrument (the Western Ontario and McMaster Universities Arthritis[53] (WOMAC) Index) to assess quality of life via three dimensions: pain, stiffness and physical function. Consideration of a broader, non-specific range of dimensions to measure patients' overall health-related quality of life may be more sensitive to radiographic changes. For example, Wilson and Abbott (2018)[54] reported that radiographic knee OA was associated with all dimensions of the SF-6D[55] except for Social Function (the SF-6D is a generic measure of health-related quality of life which estimates the health dimensions of Physical Functioning, Role Limitations, Social Functioning, Pain, Mental Health and Vitality). Nevertheless, over time it may be assumed that poor concordance between the structural signs and symptoms of OA has led to patient reported symptoms as the current standard for diagnosing OA.[56, 57]

2.2.2 Risk factors

Osteoarthritis is associated with age, sex, genotype, weight, bone mineral density, occupation and injury and physical inactivity.[58-61] Of these risk factors, weight, inactivity and injury are modifiable risk factors. While injury might be mostly unavoidable due to its random nature, excessive weight and inactivity is mostly avoidable. Given that rates of obesity are increasing around the globe,[62] we may assume that with concomitant global population growth and ageing, the prevalence of osteoarthritis will expand sharply into the future.[54]

2.2.2.1Age

Age is described as the main risk factor for OA.[63] The incidence of OA increases with age and is the result of increased exposure to various risk factors (e.g. injury) and biological age-related changes in the joint structures (e.g. oxidative damage, thinning of cartilage, muscle weakening and a reduction in proprioception).[63, 64] This makes older adults especially prone to disability caused by OA. For example, the Longitudinal Study on Aging in the USA found that 55% (8.5 million) of community-dwelling people aged 70 years and over had arthritis; 78% of the people in this age group reported a limitation in physical activity and 36% reported a limitation in activities of daily function.[65] Additionally, high quality studies in low to middle income countries is very limited, such that evidence informing the impact of OA in those settings, or on minority groups – such as Māori in NZ – warrants further investigation (the burden of OA in NZ is discussed in section 2.6).

2.2.2.2Sex

The incidence of hip and knee OA is greater among women than men, and for women, the time around menopause seems to increase the risk of OA.[66] However, the role of sex hormones in these sex differences remains unclear – the evidence is conflicting.[67, 68]

2.2.2.3Genetics

Genetics are thought to contribute significantly to OA. Twin studies have shown that genetics contributes to approximately 40%, 65%, 70% and 70% of the risk of developing OA in the hands, knees, hips and spine respectively.[69, 70] More research is required to establish the heritable component of OA.[71]

2.2.2.4Body mass index

A strong risk factor for developing knee OA is overweight ([Body Mass Index] BMI 25-30 kg/m^2) or obese (BMI>30 kg/m^2). In a recent meta-analysis, the odds ratio (OR) of developing knee OA was 1.98 (95% CI 1.57 to 2.20) for overweight and, if obese, an OR of 2.66 (95% CI 2.15 t o3.28), suggesting that increased weight is associated with increased risk of developing knee OA. An examination of the forces passing through the joints and the additional role of metabolically driven inflammation due to increased fat mass may explain this association. For example, a one unit increase in weight increases the forces going through the knee joint up to four times. A 5-unit increase in BMI is associated with an increased risk of knee OA around 35%.[72] Additionally, weight loss is associated with decreased risk of developing knee OA – a two unit change in BMI is associated with an OR of 0.46 (95% CI, 0.24 to 0.86).[73] Coupled with increased body fat composition that is associated with early through advanced OA and independent of obesity, this highly metabolic tissue produces inflammatory molecules that may further accelerate mechanisms leading to OA.[74]

It is worth noting that the relationship with weight and the hip joint, including other non-weight bearing joints such as with hand OA is unclear.[75] Some studies report no association, while others report a weak association.[76-79] These differences seen between knees and hips are thought to be due to the increased forces through the knee joint not experienced through the hip and systemic factors caused by increased BMI, particularly systemic inflammation.[71]

2.2.2.5 Mechanical factors – joint structure and malalignment, trauma, physical activity, muscle strength, occupation

Mechanical factors, including those induced by high BMI (notwithstanding metabolic factors such as systemic inflammation), contribute to the development of OA by altering the biomechanics within the joint, which, in turn, result in disruption to the normal balance of tissue breakdown and renewal. There are several mechanical factors that contribute to OA.

2.2.2.5.1 Joint structure and malalignment

Joint structure abnormalities are associated with OA in the knee[80, 81] and hip[82] and there is strong evidence that varus alignment is a predictor of knee OA progression,[80, 81] however the association between malalignment and knee OA incidence is less clear. For example, varus knee alignment is associated with increased odds of incident radiographic tibiofemoral OA (adjusted OR 1.49; 95% CI 1.06 to 2.10) and also progression of medial tibiofemoral OA (adjusted OR 3.59; 95% CI 2.62 to 4.92).[83]

2.2.2.5.2 *Occupation*

Occupation, in particular repetitive or physical work, such as construction and agriculture work or firefighting, is associated with increased risk of developing OA.[59-61, 84-88] In general, moderate levels of physical activity are unlikely to significantly increase the risk of developing knee or hip OA.[71, 80, 81] High impact sports such as football, handball, hockey, wrestling and weight-lifting are moderate to strongly associated with an increased risk of hip or knee OA.[89, 90] The increased risk associated with sporting is also partly due to injury.

2.2.2.5.3 *Trauma/ injury*

Knees are one of the most frequently injured joints.[63] It is estimated that people who have sustained knee injuries are 4.2 times more likely to develop OA.[91] Compromised anterior cruciate ligament requiring surgical intervention leads to early-onset knee OA in 13% of cases within 15 years and, if other joint structures are also involved, such as the cartilage, subchondral bone and/or menisci – which occurs in 65-75% of anterior cruciate ligament injuries (ACL) – this rate increases 21% to 48%.[92, 93] Computer simulation modelling has estimated that those younger than 25 years of age with ACL injury and meniscal tear are 2.5 times more likely to develop OA and 4 times more likely to undergo total knee replacement surgery than those without injury.[94] This is concerning among countries and younger people where there is evidence of increasing rates of ACL injury – exceeding growth in ACL surgeries – such as in Australia, where an increase in rates of ACL reconstructions has increased 74% among those under the age of 25 years between 2000 – 2015.[95] There is limited information available about the risk of developing hip OA following injury.[96] Risk of joint injury is an important risk factor for developing OA.

2.2.2.5.4 Bone mineral density

Osteoporosis is a disease defined by reduced bone mineral density (BMD) and a common age-related musculoskeletal condition. There is a positive association between increased BMD and hip, knee, hand and spine OA, however the link between BMD and OA progression remains unclear.[58, 71] Emerging evidence from the Multicenter Osteoarthritis Study[97] suggest that the odds ratio of incident knee OA over a 30 month period is 2.3 (95% CI 1.2 to 4.5) for those with high BMD compared to those with low BMD, defined by the highest and lowest quartiles of baseline BMD measured in the study. This may potentially be explained by higher rates mechanical factors (e.g. occupation and injury) that contribute to higher BMD but also higher risk of OA.

2.2.3 Diagnosing OA

A diagnosis of OA can be made based on background risk (population prevalence OA), risk factors for OA (e.g. age, sex, BMI and occupation), patient reported symptoms (persistent pain, morning stiffness and functional limitations) and clinical examination (crepitus, restricted or painful movement, joint tenderness, and bony enlargement).[13] Clinical examination diagnostic criteria for OA are available.[98] Diagnostic criteria are consistent across high quality clinical practice guidelines (e.g. the American College of Rheumatology (ACR),[99] European League Against Rheumatism (EULAR),[100] Osteoarthritis Research Society International (OARSI),[101] National Institute for Health and Care Excellence (NICE)[56] and Royal Australian College of General practitioners (RACGP)[13] consisting of: age 45 years or older, activity related joint pain, and morning stiffness lasting less than 30 minutes. Imaging, such as x-ray or MRI is not recommended but can be considered for atypical presentations or if other diagnoses are suspected.[102] Diagnostic criteria for identification of OA at its early development is currently absent, representing a window of opportunity to begin early management of the disease.

2.2.3.1 The role of imaging in the diagnosis of OA

For most people, imaging represents a low value diagnostic approach in the absence of atypical disease presentations or planning for surgical intervention due to its relatively high cost and poor diagnostic utility for confirming symptomatic OA.[103] The most commonly used scale for radiographically grading OA is the Kellgren and Lawrence (KL) grade, where the severity of OA is graded between 0 to 4, with scores ≥2 indicative of OA. The KL grade has been used to diagnose OA of the hand, hip and knee (only tibiofemoral OA and patellofemoral OA), and does not consider patient reported symptoms for defining OA. Other radiographic systems for grading OA exist such as Croft's grade,[104] minimal joint space[104] and the Tönnis classification[105] exist to detect structural changes to the joint such as cartilage and bone marrow lesions, osteophytes and effusion.[106] Computer tomography and magnetic resonance imaging tools are in development, however they are not typically required for diagnosis and are usually reserved for a secondary causes, atypical presentations or presurgical preparation.[106]

2.3 THE GLOBAL BURDEN OF OSTEOARTHRITIS

According to the Global Burden of Disease (GBD) study data,[107] all musculoskeletal conditions are currently ranked 1st as the cause of global years lived with disability (17.1% of total YLDs; 95%UI 14.9% to 19.8%). OA is a major component of all musculoskeletal disorders and is estimated to be responsible for 2.2% (1.3% to 4.1%) of the global burden of years lived with disability (YLD).[107] All-OA is ranked 16th for all-cause disability worldwide and has seen the fifth highest growth rate (33.8%) from 1990-2017 after diabetes (86.0%), neonatal disorders (56.5%), other musculoskeletal conditions (44.5%) and chronic kidney disease (40.9%).[107] It has been estimated that knee OA alone contributes to approximately 85% of the burden of OA worldwide;[108] the global burden of all-OA is most likely underestimated.

Globally, OA is the second fastest growing musculoskeletal condition by YLD after gout. However, gout only represents a fraction of global burden of YLD compared to OA (0.2% vs 2.2%). Across all non-communicable diseases, OA is ranked 13th and among the highest 20 ranked conditions by YLD, third in growth only to diabetes (86.0%) and other musculoskeletal conditions (44.5%). From 1990 to 2017 YLDs for OA increased 33.8%, or from 165.1 (95%UI 84.0 to 324.5) per 100,000 YLD to 244.9 YLDs (123.7 to 486.7) per 100,000 YLD. Among the other musculoskeletal disorders captured in the GBD study (rheumatoid arthritis, low back pain, neck pain, other musculoskeletal disorders and gout), OA also is the second most prevalent musculoskeletal condition, featuring only after low back pain (6821.5 vs 7346.7 cases per 100,000 respectively). Compared to all non-communicable diseases, the prevalence of OA has now exceeded that of diabetes and chronic obstructive pulmonary disorder (6821.5 (OA) vs 5943.5 and 2744.3 cases per 100,000, respectively). In NZ, OA is the most prevalent musculoskeletal condition which has seen the second highest growth rate after gout. It is the 6th highest ranked non-communicable disease in NZ, outranking low back pain (10th), chronic kidney disease (12th) chronic obstructive pulmonary disorder (16th) and diabetes (17th).

2.3.1 Prevalence and incidence

Differences in how studies define OA has resulted in varying estimates of the prevalence of OA.[109] Most data are from population based radiographic surveys, which are insensitive to detecting early OA and don't differentiate between symptomatic and asymptomatic OA. These studies may therefore underestimate the true prevalence of OA. For example, the 2010 Global Burden of Disease (GBD) study defined OA as >30 years of age and symptomatic OA of the hip or knee radiologically confirmed as Kellgren–Lawrence grade 2–4 which may have underestimated the true prevalence of OA[110] compared to broader definitions, such as self-reported OA. However, recent GBD estimates from 2020[1] have expanded their OA definition to include data using other definitions, such as self-reported OA. Based on this expanded this definition, GBD data estimates suggest that high-income Commonwealth countries (Malta, NZ, Canada, Australia, UK, Singapore) share similar prevalence of OA between 13.0% to 14.5%, with a greater prevalence among females and increasing prevalence from around the age of 40-44 years.

Among some OECD countries, the same GBD data estimates the prevalence of OA (all ages) to be greatest in Japan (18.8%) and the USA (16.7%), followed by Republic of Korea (16.5%), Austria (15.6%), Portugal (15.4%), Spain (15.2%), Iceland (15.4%), Denmark (14.9%), Italy (14.7%), Belgium (14.6%), NZ (14.4%), Canada (14.0), Australia (13.9%), Germany (13.6%), Greece (12.5%), Italy (14.7%), Finland (13.7%), and the UK (13.4%). Similar to high-income Commonwealth countries, the prevalence of OA increases from around the age of 40-44 years and is higher among females. To put these percentages into perspective, in the USA it is estimated that by 2040, approximately 25% of the population (78.4 million people) will have diagnosed OA, of which 34.6 million people (44.1%) will experience activity limitations due to OA.[111] In Australia, by

2030 the prevalence of OA is projected to reach 16%, affecting over 3 million people.[6] Current estimates for the prevalence of OA in NZ are around 10%, or approximately 400,000 people.[2] However, this estimate is likely to be an underestimate of the true prevalence because of the conservative criteria used to define OA and poor sensitivity for identifying early OA – which may explain the difference between this and the greater GBD estimate of 14.4%.

Emerging evidence supports that this pattern of OA among age and sex appears to be consistent among lower and middle income countries. In the study by Brennan-Olsen et al. (2017),[112] which defined hip or knee OA using a less conservative definition (by self-reported healthcare professional diagnosis and a symptom-based algorithm among persons of age 18 years and older), the prevalence of OA among low to middle-income countries was estimated to be 17%, and associated with people of lower education and divorced/separated/widowed women.

2.4 THE PERSONAL BURDEN OF OSTEOARTHRITIS

The personal experience of OA is best described using a biopsychosocial framework that considers the whole person in their own context. This framework proposes that biological, psychological and social factors are interconnected and all play an important role in an individual's experience of OA.[56] A recent systematic review of qualitative studies[113] describing the lived experience of knee OA produced the following seven themes which were consistent irrespective of disease severity:

- Pain and how to manage it dominates the lived experience – pain is 'omni-present' and interrupts or deters one from daily activities and less confident in their bodies.
- The causes of OA are multifactorial and lead to structural damage and deterioration over time
- OA negatively impacts activities of daily living and participation – quality of life is reduced by altered pain perception and functional capacity[114]
- OA leads to life adjustments – seeking out health-related information, taking measures to alleviate symptoms and protect knee joint through lifestyle/work modification, exercise, weight loss and everyday routines such as less frequent stair climbing.
- OA has an emotional impact – negative impact on mood, resulting in feelings of loss, anxiety, inadequacy, frustration, irritability, emotional distress, depression, embarrassment, fear for the future and uncertainty of the outcomes of knee pain
- OA has a social impact –reduced ability to stay connected/increased isolation due to poor access to public transport and decreased mobility
- Interactions with health professionals can be positive or negative – positive experiences were described as being listened to, offering hope for the future and providing recommendations for managing OA, whilst negative experiences were related to dismissive health professional behaviour and limited provision of information about OA.

These themes highlight the multidimensional impact which OA has on an individual, which may negatively contribute to their quality of life and accumulate as the severity of OA intensifies.[115]

2.4.1 Young people

OA is typically associated with older people however there is growing evidence that OA also has a significant impact among younger people. For example, in a study among Australians[48] the overall prevalence of high/very high distress among people with OA aged 20-50 years was 4 times higher compared to normative Australian population data (RR 4.2; 95%CI 3.5 to 5.0) and in the same cohort 67% reported moderate to considerable OA-related work disability. Young people living with arthritis also experience work limitations and higher work disability prevalence rates than those without arthritis[116] that may limit their productivity compared to those without arthritis. For example, an Australian qualitative study of 21 people experiencing arthritis Berkovic et al. (2020)[117] reported that arthritis profoundly impacted their finances, which were associated with distress and anxiety due to factors such as the costs of clinical care and medication, reduced employment wages and impact on family and reduced income due to arthritis. Of note however, is that the sample was at risk of selection bias caused by oversampling of females and underrepresentation of people with lower education and socioeconomic status. Therefore, it may well be that the causes and implications of arthritis-related financial distress are far greater than what was reported in this study.

Sustaining OA earlier in life may be caused by injury due to sport, which is associated with increased risk of developing OA and therefore more years lived with disability. In a cohort study of 141 students, 36 years later the relative risk for subsequent osteoarthritis following injury was 5.2 (95%CI 3.1 to 8.7) and 3.5 (0.8 to 14.7) for the knee and hip.[118] This emerging evidence suggests that the burden of OA – often associated with older people – may also be significant among younger aged people, highlighting the need for prevention, early detection and management of OA.

In the USA, it is estimated more than half the 14 million people aged 25 years and older with symptomatic knee OA are younger than 65 years,[119] highlighting the possibility of huge unmet research need in this area.

2.4.2 Comorbid conditions

People with OA experience comorbidity, but OA is also comorbid with other non-communicable diseases. In the context of the burden of OA, both dimensions are relevant and considered in the following sections, as it relates to anxiety and depression, systemic diseases and other musculoskeletal conditions.

Anxiety and depression are highly prevalent among people living with OA and significantly impair the quality of life of people by altering their pain perception, self-efficacy and functional capacity.[114] Typically, people with anxiety and depression have less ability to cope with pain and higher physical limitations which result in poorer outcomes to both conservative and surgical interventions. The culmination of these effects ultimately results in greater use of drug therapies and other health services.[120] Additionally, anxiety and depression are also associated with poor outcomes for chronic obstructive pulmonary disease (COPD) and coronary artery disease (CAD).[121, 122]

People with OA typically have other comorbid diseases, and the presence of comorbid diseases is associated with worse, or greater deterioration of pain and performance-based physical function.[123] Importantly, modifying health outcomes for people living with OA can have a positive impact on other non-communicable diseases, such as obesity, diabetes, cardiovascular disease and COPD, which are commonly occurring comorbid conditions of OA. OA is a risk factor for developing other non-communicable disease.[124] For people with OA, up to 40% can have cardiovascular disease,[125,126] and those with cardiovascular disease are up to 3 times more likely to have heart failure (RR 2.8; 95%CI 2.3 to 3.5) or ischaemic heart disease (1.8; 1.2 to 2.7).[125] Thus, OA is associated with morbidity and in combination with comorbid diseases, the risk of mortality also increases.

OA is also highly prevalent among people with diabetes mellitus (up to 30% of patients with diabetes mellitus also experience OA, and up to 14% with OA also experience diabetes mellitus).[127] A large systematic review and meta-analysis including over 1 million participants reported an overall odds ratio risk of OA in the diabetes mellitus population of 1.46 (95%CI 1.08 to 1.96) and that of diabetes mellitus in the OA population was 1.41 91.21 to 1.65).[128] However, there is conflicting literature on the association between diabetes mellitus and OA, with other studies (mostly of hand OA) suggesting there is no association between diabetes mellitus and OA.[129-131]

Obesity is strongly associated with onset of knee pain: being overweight increases the odds ratio of knee pain by 1.98 (95%CI 1.6 to 2.2) and the odds ratio of being obese by 2.7 (2.2 to 3.3). It is estimated that almost 25% of new cases of knee pain is related to overweight or obesity.[132]

People living with OA also experience other musculoskeletal conditions, such as back pain. For example in Australia,[133] the most common MSK conditions are back pain problems where up to 31% also experience arthritis. Similarly, almost 80% of people with arthritis have at least one other condition – the most common being CVD (48%) and back pain problems (33%). A systematic review including approximately 2.6 million people[124] also produced compelling evidence that compared to no musculoskeletal condition, having a musculoskeletal condition increases risk of developing a chronic disease by 17% (hazard ratio 1.17, 95%CI 1.13 to 1.22, I^2 52%).

2.4.3 Summary

From 55-98% of people over the age of 65 have comorbid conditions and the prevalence of comorbid conditions is associated with increasing age, sex and lower socioeconomic status.[134] Given that COPD, diabetes, cardiovascular disease and OA tend to reduce physical activity and increase sedentary activity,[135] addressing the burden of OA has the potential to reduce or prevent mortality and morbidity in comorbid diseases and the physical decline associated with these conditions.[136] These converging factors highlight why management of OA is a meaningful research pursuit which may also have a broad impact across the population.

2.5 THE SOCIOECONOMIC BURDEN OF OSTEOARTHRITIS

Cost estimates for the burden of OA vary from country to country due to different methodologies used to estimate cost and sociodemographic characteristics such as obesity and related comorbidities.[137] However, one paper estimates that among some high-income countries – USA, Canada, UK, France and Australia – the cost of OA may be as high as 1-2.5% of the GPD of these countries.[138] The high prevalence of OA due to a long disease course (currently until death) reduces workforce participation which leads to lost revenue due to early retirement, lost taxation for governments and potentially greater disability payments, and greater health system pressure. The socioeconomic burden of OA can be split into two categories – direct costs and indirect costs.

2.5.1 Direct costs

Direct costs relate to costs directly attributable to OA care. Medical costs, for example, include health care provider (including allied health) visits, drug costs, joint replacement costs (including hospital stay), costs incurred due to changes in living environment, assistive devices, non-prescribed alternative medicines and, community services and health administration. Out-of-pocket expenses are those which are paid directly from the consumer/patient. It has been estimated that the annual direct costs of OA per person range from 2015 US$1442 to US$21,335 with increasing cost driven by hospitalisation and outpatient costs.[139]

Having other comorbid disease may increase the direct costs of healthcare substantially. For example, in a large Dutch study of almost 9000 participants,[140] having a musculoskeletal condition with two or three other non-communicable diseases was estimated to increase health care costs by a factor of three to five, respectively. However, the study only recruited 22% of the invited participants, which may have potentially introduced a risk of sampling bias. The risk of bias could have potentially been offset by a randomised study design. Another limitation of this study was that they did not capture out-of-pocket expenses, which may have consequently underestimated the true direct costs of health care.

2.5.2 Indirect costs

Indirect costs generally refer to the value of lost of productivity due to: illness and disability (e.g., decreased productivity due to presenteeism and absenteeism), the value of future earnings lost by people who die prematurely (e.g., lost taxation, superannuation loss), as well as welfare payments, carer costs, and mobility/accessibility costs. For example, the systematic review by Agaliotis et al. (2014)[141] found that for people with OA, rates of absenteeism was between 5-22% over a 12-month period and that 71% of the participants reported reduced work hours due to OA over the same time frame.[142] At an individual level, the per person cost of OA in 2015 was estimated to be between US$238 to US$29,935, with greater OA severity associated with greater cost.[139] In 2012, Australian estimates for the financial value of lost quality of life due to osteoarthritis was $9.1 billion.[143]

2.6 CASE STUDY: THE BURDEN OF OSTEOARTRITIS IN NEW ZEALAND

OA is the most common type of arthritis in NZ affecting about 10% of the total population or about 404,000 people (95%CI 381,000 to 428,000).[2] OA is more prevalent among females and people of older age.[144] NZ Health Survey data[144, 145] suggest that the NZ population is growing and aging, with increasing rates of adult morbidity and decreasing rates of physical activity. Almost one-third (30.7%) of the NZ population is experiencing obesity[62] and more recent projections predict that population mean BMI will exceed 30kg/m^2 (obese) within the next 15 years.[54] This combination of these risk factors for OA suggests that the burden of OA will expand rapidly over time in NZ.

2.6.1 Socioeconomic impact

It is likely that the availability of health resources in the NZ public health system will not keep pace with increasing demand over time. For example, for the most of the past 60 years, the growth of health expenditure has exceeded the growth of national income GDP.[146] This raises concern about the long-term sustainability of the health system when public funding already accounts for approximately 83% of health expenditure[146] and health expenditure consumes approximately 9.4% of gross domestic product (GDP). In 2010, the economic cost of OA made up 70% of all arthritis; total costs of arthritis in 2010 was 1.7% of GDP, or $3.2 billion dollars.[3] The majority of these costs are due to inpatient costs, primarily due to hip and knee total joint replacements.

In 2013, the cost of total knee replacement to the health system was NZ$199 million. Computer modelling projects this cost to increase to $370 million in 2038, with the incidence of annual total knee replacement almost doubling from 5,070 to 9,040 over the same period.[4] In the same study, Wilson and Abbott also reported that population obesity rates (above the obesity prevalence seen in 2013) accounted for 25% and 47% of the projected increase in per-capita healthcare costs and total knee replacement provision rates, respectively. This will place greater pressure on the health system to continue to meet this level of demand for total knee replacement.

Other unavoidable factors such as population aging and end-of life care; income and technology demand; and rising costs of service delivery versus low productivity growth relative to the rest of the economy, will likely cause this trend of increasing health system pressure to continue.[146] Therefore, it is imperative that every dollar spent on health is utilised for maximum health benefit. Small changes to highly prevalent health conditions early in the disease course – such as with OA – could have a large impact on the health budget.

2.6.2 Māori health

Maori people make up about 0.78 million people or 16.5% of the NZ population after NZ Europeans, who are the largest ethnic group in NZ (3.3 million people, 70.2%).[147] In the 2019/20 NZ health survey, the broad ethnic groups in NZ who reported a diagnosis of OA were, in order of decreasing prevalence: NZ European plus 'Other'; 373,000, 12.3%), Māori (39,000, 7.7%), Pacific people (8,000, 4.6%) and Asian (13,000, 2.2%).[147] After adjusting for age and gender, compared to non-Māori, the adjusted prevalence rate ratio for Māori was 1.12 (95%CI 0.14 to 8.08). This is second only to disabled vs non-disabled adults (adjusted prevalence rate ratio 1.3 (95%CI 0.18 to 9.4).[147] Māori also experience approximately seven years reduced life expectancy and are at increased risk of multimorbidity compared to their non-Māori counterparts.[148, 149] Also, unmet access to primary healthcare in 2014/15 is about 1.3 times greater for Māori compared to non-Māori people.[144] These examples highlight the health disparities between Māori and non-Māori people of New Zealand.[150] Conducting research and developing health policy which is inclusive of Māori perspective is therefore important for the social cohesion and health equity of New Zealand.[151] Considering the reach of OA and increased risk of morbidity and mortality among

Māori, the potential to manage OA better in the community (quality of OA care is discussed in Section 2.10), gaining insight into the experiences of Māori people may have a useful impact on closing these health disparities in New Zealand.

2.6.3 Current management of osteoarthritis in New Zealand

2.6.3.1 Surgical management

Total joint replacement surgery is well utilised in NZ. However, Hooper et al. (2014)[152] concluded that demand for knee and hip replacement surgery is unsustainable and requires 'continuing service and training requirements for the delivery of an efficient and effective National Orthopaedic Service.' Notably, no mention of conservative or alternative strategies for managing osteoarthritis were mentioned, although the focus of the report was with regard to elective surgery. Nevertheless, a report published in the previous year similarly recommended additional provision of services for joint surgery. Interestingly, both reports failed to recommend investment in conservative treatment strategies, despite consistent recommendations for high-value, early and conservative intervention such as self-management education, exercise and weight loss, if appropriate.[13, 14, 16, 101] A shift in how the health system plans and delivers OA care is needed,[8] utilising strategies to help deliver value-based care such a national Model Of Care for OA and Model of Service Delivery (discussed in Section 2.9.2)

2.6.3.2 Conservative management

There are few studies in exploring OA care in NZ. Two studies suggest that the quality of care received by patients in primary care is discordant with clinical guideline recommendations. For example, Larmer et al. (2019)[153] reported that patients were less frequently recommended weight management strategies and daily activity aids assessment by their GPs, but more frequently offered paracetamol. The same authors also reported that the patient journey was inconsistent, aligning with an earlier study of OA patients and their care experience in NZ.[154] In a qualitative study of 13 consumers with knee OA, Darlow et al. (2018)[155] also found that that their biomechanical beliefs about OA were discordant with positive health behaviours and self-management. This appeared to be exacerbated by their health care provider's language and explanations of OA. However, their sample comprised of participants >50 years, warranting confirmation in a younger sample. Another mixed-methods study of consumers', health care providers' and arthritis advocates' perceptions of a novel information booklet for OA also found variability in the beliefs about consumers' informational needs between groups.[156]

2.6.3.3Policy landscape

In response to the NZ health system's poor utilisation of early intervention for people with musculoskeletal conditions, in 2015 the NZ government allocated $6 million over three years to trial and evaluate health care programmes in NZ aiming to improve access, education, and management of people with MSK conditions – the Mobility Action Programme (MAP).[157] In contrast, $44 million was allocated for more hip and knee replacements in the same period.[158] It is hoped that the funded MAP programmes demonstrating sustainable and effective outcomes will be rolled out nationally in the future; NZ's Ministry of Health has yet to release information about which programmes will have continued funding in the future.

New Zealand currently does not have any other overarching policies, frameworks or strategies specifically for MSK conditions, arthritis, or integrated non-communicable disease prevention or management policies that have been reported to the World Health Organisation,[159] except for their overarching NZ Health Strategy.[8] In contrast, other countries, such as Australia and Canada have developed strategies and frameworks to facilitate better management of OA across the health system. For example, Australia has developed a National Strategy for OA,[160] National Strategic Action Plan for Arthritis and a National Knee OA Clinical Care Standard[161] including jurisdictional health policies and frameworks to address the evidence-policy and practice gap for MSK conditions (e.g. the Victorian Model of Care (MoC) for Osteoarthritis of the Hip And Knee,[162] NSW OACCP,[163] WA Elective Joint Replacement MoC[164]). Another important point of difference between NZ and Australia, is that the latter considers musculoskeletal health as a national priority area, while NZ does not. This may explain the gap in health policy which address this important condition and more broadly musculoskeletal conditions at large. Other countries have also developed specific service models.[165] An example includes the Good Life with osteoArthritis in Denmark (GLA:D) programme, which is currently at the early stages of being introduced into NZ.[166]

2.7 TOWARDS VALUE-BASED HEALTH CARE

Evidence-based (or evidence-informed) practice and patient-centred care are two paradigms which underpin quality healthcare (whereby quality may be broadly characterised by that which is effective, safe, and implemented a culture of excellence that produces desired outcomes[167]). These paradigms represent the integration of evidence, expert judgement and patient values and preferences in a shared decision-making model. In the past decade 'value-based healthcare' has also gained traction as a philosophy of delivering population-level healthcare in the face of limited healthcare resources and increasing healthcare waste and overuse documented around the world.[23] This is a pertinent issue for people living with OA, where quality of care is discordant with best practice, due to multiple barriers across the healthcare system (discussed in Sections 2.10 & 1.16) The next section will describe how these key healthcare models relate to each other in the pursuit of better health outcomes for the individual and society at large.[168]

2.7.1 Shared decision-making

Shared decision-making can be described as the pinnacle of evidence-based practice and patient-centred care. Without it discussion about alternative treatments, benefits and harms, consideration of patient values, preferences and circumstances cannot take place.[169]

Although this definition implicitly refers to the clinical context, shared decision-making is equally relevant in the context of co-designing clinical practice guidelines (discussed in section 2.9), where recommendations about health issues are made on behalf of the population of interest in a shared decision-making process involving consumer representatives and/or organisations.

The concept of shared decision-making is also important for implementation efforts such as health system reform. Engagement is thought to create buy-in, ownership and align contextual factors that may affect implementation decisions.[170] Models of Care,[171]for example, are co-designed strategies aiming to influence macro level health system factors (e.g. health policy and strategy priorities, resourcing and health governance decisions) which are deliberately evidence-based and involve cross-sectoral stakeholder-engagement for this reason. For people living with OA, engaging in shared decision-making is important at every level of the health system to help shape the design, delivery, uptake and surveillance of care that is more aligned with their wants, needs and best-practice recommendations.[29, 32]

2.7.2 Evidence-based care

Evidence-based practice is currently the dominant best-practice paradigm in modern clinical medicine. It has philosophical roots based in mid-19th century philosophy[172] and can be linked to the birth of clinical epidemiology and Archie Cochrane's seminal work on clinical effectiveness and efficiency.[173] This paradigm relies on current best-practice information regarding a disease course and the anticipated outcome of various treatments. It is characterised by the explicit and rational process of clinical decision-making and is the integration of three key components: (1) scientific evidence, (2) expert judgement and (3) patient preferences.[174] The extension of these components is high-value care, which also considers cost-effectiveness.

Scientific evidence is clinically relevant evidence that concerns the efficacy (i.e. the capacity for an intervention to do more good than harm under ideal or controlled circumstances), effectiveness (i.e. the performance of an intervention under real-world conditions, such as in clinical practice) and the efficiency of an intervention in relation to the resources it consumes (e.g. cost-effectiveness).[175] It has been the focus of evidence-based medicine and involves the process of locating, appraising and synthesising evidence. In recent times, the proliferation of scientific evidence has led to unmanageable amounts of data for the modern health professional to integrate into their evidence-based practice. In light of this, organisations such as the Cochrane Collaboration (www.cochranelibrary.com) have been established primarily to produce high quality evidence syntheses (systematic reviews) that analyse and present vast amount of scientific evidence available. For example, an updated Cochrane systematic review of exercise on knee OA[176] identified an additional 22 papers in a 5-year interval following the last reporting period.

Clinician expertise integrates with scientific evidence and patient preferences to deliver healthcare that is informed by the most up to date information available and tailored towards patients' values and preferences. It can also be described as healthcare providers' experiential expertise which leads to more effective and efficient diagnosis and a more robust consideration of a patient's predicaments, rights, and preferences in making clinical decisions about their care.

Evidence alone does not define high-value care. It must be based on patient preferences, too.

Patients' values and preferences are the perspectives, beliefs, expectations, goals of health care and the processes that individuals use in considering the potential benefits, harms, costs and inconveniences of one alternative compared to another.[177] Van der Weijden et al. (2010)[178] make several points why patient participation is justified. First, patient autonomy – the ability of an individual to make choices about their own health care – is a key underlying value of health care itself. Second, where there is inconclusive evidence available to make clear recommendations, such as when there is insufficient scientific evidence, or two treatment options of equivalent benefit and harm exist (equipoise), patient preferences ought to be the deciding factor. Third, where preferences vary among individuals. Fourth, the psychological impact where patient involvement and control leads to better treatment adherence and therefore, better outcomes (also known as patient-centred care, discussed in section 2.7.3).[179]

Despite the evidence-based practice approach, there is inconclusive evidence to suggest that evidence-based practice is effective.[180-182] One reason may be because the paradigm has been criticised for ignoring patient values and preferences, or de-emphasised due to other dimensions of evidence-based practice, despite patient preferences being a key component of the evidence-based practice model.[172] Perhaps this is because patient values and preferences are the most difficult and poorly understood aspect of evidence-based medicine[169] and because evidence-based practice has historically focussed on scientific evidence.[173] Nevertheless, patient participation is a key concept of evidence-based care, and is discussed in the following section in terms of the patient-centred care paradigm.

2.7.3 Patient-centred care

Patient-centred care focusses on patient values, a patient's healthcare experience and the clinical expertise necessary to engage in dialogue which enables patient values to guide all clinical decisions.[183] Compared to evidence-based practice, patient-centred care is relatively new, dating back to the early 1990s.[184] It is described as the process of involving patients in decision making, characterised by three core elements:[184] (1) patient participation and involvement; (2) relationship between the patient and health professional and; (3) the context where care is delivered. Unlike evidence-based practice, the key message of patient-centred care is that it is patients', not experts' views on patient-centred care that predict health outcomes and efficiency of healthcare.[185] In other words, it facilitates meaningful patient engagement to deliver evidence-based care. Governments, international organisations and advocacy groups are now focussing on patient-centred care as a core guiding principle of more effective healthcare delivery,[32, 184, 186] highlighting a growing interest in this relatively new paradigm.

People living with OA do not always receive patient-centred non-operative care. A recent Dutch study into the quality of care received by 235 OA patients in primary care, from the patients' perspective, found that there was generally low adherence (38%) to process-related quality indicators, with considerable variability 23-97%.[187] In particular, the quality indicators for exercise, self-management education and weight loss were only achieved in 43%, 40% and 23% of patients, respectively. A potential critique of the findings, however, is that patients were asked to recall their health care experience which, may have resulted in underreporting. Further care gaps are highlighted in the narrative systematic review by Paskins et al. (2014)[188] where they reported patients often feeling like their symptoms were not legitimised following their GP consultation, and that OA was perceived as a low priority by their health care provider. Moreover, patients reported feeling dissatisfied with the information they received about OA and the management of OA. For example, patients felt that they did not receive enough information regarding disease progression and the benefits and harms of drug treatments, leaving them to be responsible for their own choices, and a breakdown in the shared decision-making process – despite expressing their desire for a good doctor-patient relationship.

With respect to operative care, there is evidence that patients may remain dissatisfied following knee replacement demonstrating that their needs have not been met. For example, in a study of 27,372 patients in Sweden following total knee replacement, up to 17% of unrevised patients reported feeling dissatisfied or uncertain about their treatment outcomes.[189] Similarly in the UK, only 22% of 22,278 patients were very satisfied with their TKR, despite 71% reporting improvement with their knee symptoms.[190] This highlights that typical surgical outcome measures following total knee replacement – such as range of movement, joint stability and post-operative alignment – may not necessarily capture the full breadth of patents' needs. For example, a recent systematic review of 181 papers[191] suggest that patient satisfaction could be improved if surgeons performing total knee replacement considered factors important to patients such as no history of mental health problems, back pain, and pre-operative expectations being met.

Patient engagement is an important aspect of defining high-value care. Patients need to be involved in defining what health outcomes matter.

2.7.4 Value-based healthcare

> In a well-functioning health care system, the creation of value for patients will determine rewards for all system actors. Value is measured by outputs, not inputs. Hence value in health care depends on the actual patient health outcomes, not the volume of services delivered. Porter (2010)[192]

Value-based healthcare is becoming increasingly relevant as a model for healthcare reform for the 21st century given the growing evidence of overuse and underuse of healthcare services globally, in both low- and high-income countries.[22, 23] Value-based care may be defined as "the creation and operation of a health system that explicitly prioritises health outcomes which matter to patients relative to the cost of achieving this outcome".[193] Building on evidence-based and patient-centred care, the benefits of value-based care extend to patients, providers, payers, suppliers and society as a whole. The seminal article by Berwick et al. (2008)[194] introduced the 'triple aim' of healthcare which has since been updated[195] to include a fourth aim. The quadruple aim of healthcare describes the pillars of modern value-based healthcare: improving the individual experience of care; improving the work life of health care clinicians and staff; improving the health of populations and reducing the per capita costs of care for populations (Figure 3).

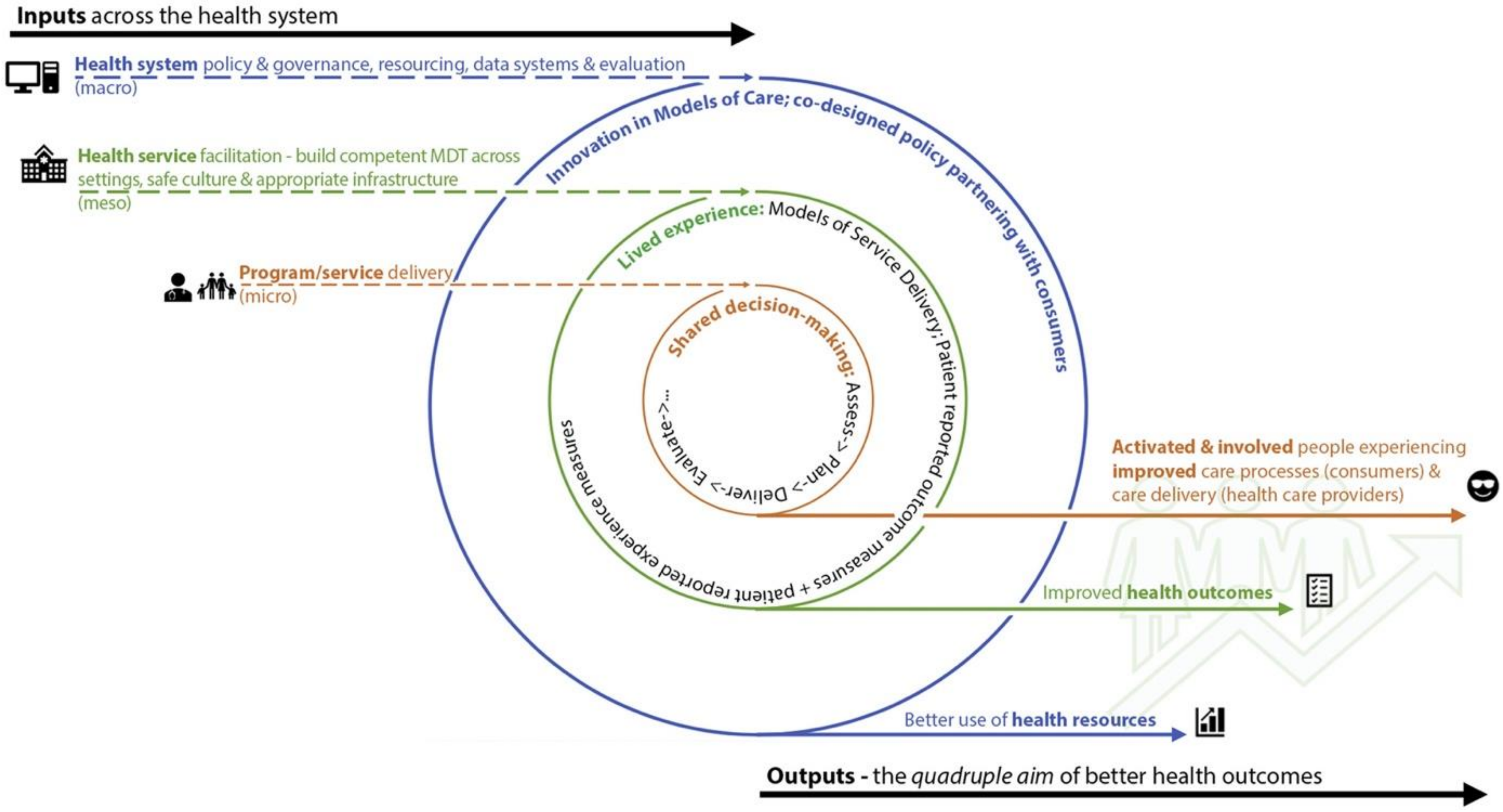


Figure 3. The inputs and outputs of value-based healthcare across the macro, meso and micro levels of a health system. Replicated from Speerin et al. (2020)[29] with permission (Elsevier).

Value-based care tries to increase health outcomes in an efficient way (i.e. relative to the total costs of healthcare). Importantly, value is derived from what matters to patients, rather than other stakeholders such as policy-makers or healthcare providers. It is believed that using patient reported outcome measures (PROMs) and patient reported experience measures (PREMs) can help identify patient needs and goals, improve communication and decision making with their providers, and support greater patient-centred care. Preliminary evidence suggests that PROMs and PREMs may bring us closer to this goal;[196] in New South Wales, Australia, for example, PROMs and PREMs are being piloted across multiple jurisdictions to evaluate the feasibility of systemwide implementation in the NSW healthcare system,[197] bringing the theory into practice. An example of a PROM is the Patient Reported Outcome Measures Information System[198] (PROMIS), which provide a global assessment of health status via the three domains of health – physical, mental and social wellbeing. A specific set of patient reported outcome measures for hip or knee OA has also been developed by the International Consortium for Health Outcomes Measurement Hip and Knee OA Working Group.[199] The main outcomes assessed include: joint pain, physical functioning, HRQOL (health-related quality of life), work status, mortality, reoperations, readmissions, and overall satisfaction with treatment results.

In value-based healthcare outcomes are considered across the continuum of care rather than at an individual service. This forces healthcare providers to shift the focus of cost from episodic or volume-based care – which are typically based on what healthcare providers believe are important – to the total cost of healthcare across the continuum of care and the health system as a result. This creates joint accountability for outcomes and costs in a health system where patients receive what matters to them, what health care providers want to practice, and what the health system can fund. Figure 3 shows how health system policy and governance (macro), health service capability and capacity (meso) and program/service delivery factors (micro) influence the quadruple aim of healthcare. Addressing these areas of healthcare can help to achieve, for example lower costs and better outcomes for patients; greater provider satisfaction – realised through better patient experience and care efficiencies; greater alignment of costs to patient outcomes, and overall greater societal health and reduced healthcare spending.[200]

There are several examples[165] of successful value-based OA programs such as the Osteoarthritis Chronic Care Program (OACCP, Australia), Better management of patients with osteoarthritis (BOA, Sweden) and Good life with osteoarthritis in Denmark (GLA:D) programs. An example of success is the OACCP; a comprehensive and multidisciplinary 52 week management program which aims to slow disease progression. Outcome measures include the pain Visual Analogue Scale (VAS), knee injury and OA outcome score (KOOS), Patient Global Assessment, EQ-5D HRQOL; and other measures, such as willingness for surgery. The program has been active since 2012, and a 2014 evaluation found that the program resulted in improved clinical outcomes (e.g. pain, mobility and function), with up to 11% of patients on waiting lists for knee replacement removed, as they were no longer willing to elect for surgery by the end of the program.[201]

2.7.4.1Summary

Evidence-based practice and patient-centred care are major paradigms that underpin modern decision-making in healthcare. Value-based healthcare aims to reorient health systems closer to what matters to patients in the face of increasing global healthcare waste in a fiscally constrained environment. Designing Models of Service Delivery and Models of Care to help drive value-based care across a health system may lead to better patient outcomes and health efficiencies.

2.8 RECOMMENDATIONS FOR MANAGING OSTEOARTHRITIS

In the absence of a cure, OA care consists primarily of long-term disease management until surgical intervention is required. There are many interventions to choose for OA care. For example, the 2019 OARSI guidelines for hip and knee OA[17] identified 67 interventions for knee OA (31 non-pharmacological, 24 pharmacological and 12 nutraceuticals); the 2018 Royal Australian College of General Practitioners (RACGP) clinical guideline for hip and knee OA[13] considered evidence for 37 non-pharmacological, 35 pharmacological, 3 surgical and 4 combination therapies for hip and knee OA; and the 2019 American College of Rheumatology (ACR) Foundation guideline for the management of the hand, hip and knee reviewed some 20 and 23 non-pharmacological and pharmacological interventions for knee OA only (totalling 48 interventions for hand, hip and knee OA). There are many clinical practice guidelines that provide recommendations for managing OA. For example, the systematic review of OA CPGs conducted in 2014 by Nelson et al. (2014)[202] identified 16 guidelines for managing OA.

In general, the non-surgical management of OA (excluding clinical diagnosis) consists of three 'lines' of OA care that also broadly align with clinical practice guidelines for musculoskeletal conditions:[15] (1) first-line (core) interventions (self-management education, exercise and weight loss, if appropriate; (2) second-line optional adjunctive and advanced pharmacological (e.g. NSAIDs, massage and intra-articular corticosteroids) and (3) third-line surgical interventions (namely referral for surgical intervention when all conservative options have failed). Discrete interventions for OA care can also be categorised into three 'types' that align with the principles of OA care:[203] (1) conservative or non-pharmacological interventions, (2) pharmacological (drug) interventions and, (3) surgical interventions.

Notably, among first-line (core) interventions, there is consistency among high quality clinical practice guideline recommendations. In particular, self-management education, exercise and weight loss, if appropriate.[14, 202] Table 1 shows that these management principles continue to feature strongly among recently published CPGs for managing OA, for example the 2019 ACR,[16] 2018 RACGP[13] and 2019 OARSI[17] (Osteoarthritis Research Society International) guidelines.

Core interventions for managing OA offer effective and relatively low-risk (high-value care) options compared to second- third-line interventions. Exercise and weight loss are both associated with long term positive effects ≥12 months and low risks compared to more commonly prescribed drug interventions such as NSAIDs, which increase the risk of potentially serious medical complications (e.g. gastrointestinal side-effects).[204] Education to self-manage (versus education about self-management) articulates with the patient-centred care paradigm and empowers consumers with knowledge to change their attitude about OA (which may be counter-productive to intervention adherence[205, 206]) and skills to manage their disease, which includes their ability to navigate and access the health system.[207] This may help consumers to develop better relationships with healthcare providers and also improve intervention adherence.[205, 208]

Table 1. Summary of the key recommendations and approaches used to build consensus across seven clinical practice guidelines for osteoarthritis.

	Guideline						
	AAOS	**ACR**	**EULAR**	**ESCEO**	**NICE**	**OARSI**	**RACGP**
Recommendation							
Management principles							
Patient-centred care	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Assessment							
Diagnosis: exclude serious pathology		✓✓			✓✓		✓✓
Assess psychosocial factors		✓✓	✓✓		✓✓		✓✓
Routine use of radiological imaging					0		0
Undertake physical examination			✓✓		✓✓		✓✓
Evaluation/ re-evaluation and measurement			✓✓		✓✓		✓✓
Management							
Provide education/information			✓✓	✓✓	✓✓	✓✓	✓✓
Prescribe physical activity/exercise	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Offer self-management programmes	✓✓	✓✓			✓✓	✓✓	?
Weight loss if overweight/obese	✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Pharmacological: paracetamol alone	?	✓		✓ (short-term only)	✓	✓✓	?

	Guideline						
	AAOS	**ACR**	**EULAR**	**ESCEO**	**NICE**	**OARSI**	**RACGP**
NSAIDS	✓✓	✓✓		✓✓	✓	✓	✓
NSAIDS plus paracetamol		✓✓			✓		
Opioids - acute pain	✓				✓	?	0
Opioid - chronic pain	✓	✓ (tramadol only)		✓	✓	?	0
Duloxetine		✓		✓		✓	✓
Glucosamine and chondroitin	0	0		✓✓	0	0	0
Capsaicin		✓			✓	✓✓	0
Passive interventions – non-invasive: manual therapy	?	0			✓		✓
Use manual therapy with other modalities		0			✓✓		✓✓
Electrotherapy		0			✓	0	0
Braces/orthoses	?	✓✓ (brace only)	✓		✓	✓	?
Psychological therapy		✓					✓
Passive interventions – invasive: surgery: Trial conservative care first before surgery					✓✓		✓✓
Knee arthroscopic lavage and debridement for OA	0				0		0

	Guideline						
	AAOS	**ACR**	**EULAR**	**ESCEO**	**NICE**	**OARSI**	**RACGP**
Passive interventions – invasive injections; Intra-articular corticosteroid injection for knee OA	?	✓✓		✓	✓	✓	✓
Complementary medicine: acupuncture	0	✓			0	?	✓✓
Facilitate continuation/resumption of work			✓✓				
Guideline decision-making considerations							
Scope	Knee OA	Knee, hip, hand OA	Non-pharmacological core management of hip and knee OA	Knee OA	OA	Non-surgical management of knee, hip and polyarticular OA	Non-surgical management of knee and hip OA
Evidence summaries (method)	GRADE	GRADE	EULAR††	GRADE	GRADE	GRADE	GRADE
Experts' opinion	Y	Y	Y	Y	Y	Y	Y
Number in decision-making panel	21	15	21	18	25	13	13
Multidisciplinary panel	Y	Y	Y	Y	Y	Y	Y
Patient representatives (n)	N	Y (NR)	Y (2)	Y (NR)	Y (2)	Y (4)*	Y (NR)
Country represented by panel	US	NR	10 EU countries	International	UK	US & Europe	Australia

	Guideline						
	AAOS	**ACR**	**EULAR**	**ESCEO**	**NICE**	**OARSI**	**RACGP**
Panellists consider preferences beyond typical clinical outcomes (e.g., patient preferences for care)	Y	N	N	Y	Y	Y	Y
Explicit explanation the values and preferences	N	N	N	N	Y**	N	N
Consensus via voting (method)	Y (NGT)	Y (NR)	Y (Delphi)	Y (NR)	N (NR)†	Y (NR)	Y (Delphi)
Rating scale	NR	NR	11 point	5 point	-	2 point	11 point
Rounds	3	NR	5	NR	-	4	2
Threshold for consensus	NR	70% agreement	80% agreement	75% agreement	-	75% agreement	70% agreement
Anonymous voting	Y	NR	Y	Y	-	Y	Y

Table and data extraction adapted from Lin et al. (2019)[15]

✓✓Should do; ✓ Could do; O Do not do; ?Uncertain; Y, Yes; N, No; NR, Not Reported

* Patients not involved in voting process, only reviewing and commenting on the final draft report; ** Social Value Judgements guidance

† Guidance suggests 'formal consensus techniques' including Delphi & NGT; however, neither are specified; †† Dougados et al. (2004)[209]

EU, European Union; NGT, Nominal Group Technique; NSAIDS, Non-steroidal anti-inflammatory drugs; OA, Osteoarthritis
AAOS,[210] American Academy of Orthopaedic Surgeons; ACR,[16] American College of Rheumatology; EULAR,[211] European League Against Rheumatism; ESCEO,[212] European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases; NICE,[56] National Institute for Health and Care Excellence; OARSI,[17] Osteoarthritis Research Society International; RACGP,[213] Royal Australian College of General Practitioners

2.9 STRATEGIES FOR DRIVING EVIDENCE INTO PRACTICE

Clinical practice guidelines and models of care for OA are strategies which aim to drive recommendations for managing OA into practice. Collectively, these strategies help to drive evidence into practice for OA care at every level of the health system: from the level of patient and their health care provider (micro), through service delivery (meso) to the level of health policy-making (macro). They achieve this by providing guidance about appropriate healthcare: what, when, where, to whom and how should it be delivered?

2.9.1 Clinical practice guidelines

Clinical practice guidelines are considered one of the foundations for the improvement of healthcare and feature as key enablers for evidence translation in healthcare policy, planning, delivery, evaluation and quality improvement.[214] They are vehicles *intended* for ensuring best clinical practice for people with specific conditions across any aspect of their condition, reducing unwarranted variation in care and regulating services cost.[215]

The Institute of Medicine[183] defines clinical practice guidelines as “statements that include recommendations intended to optimise patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options”. In this context, recommendations refer to a shortlist of interventions, or treatment strategies, that have met a threshold of evidence according to some criteria. A key assumption of clinical practice guidelines is that providing best-evidence about interventions will lead to optimal decision making and care.[216]

Among guidelines for OA there is variation in terms of the methods used to make decisions and recommendations about interventions. For example, Table 1 shows that some guideline developers use consensus methods such as the Nominal Group Technique, while others use the Delphi method to make decisions and recommendations. Nevertheless, a common feature of developing recommendations is that they all involve an assessment of scientific evidence and experts' opinion about the evidence, based on their values and preferences.

However, recommendations alone cannot determine the outcome of a clinical decision in the absence of expert judgement and patient values and preferences. Rather, recommendations help decision-makers such as patients and healthcare providers make informed decisions by synthesising evidence in conjunction with their values and preferences.

2.9.1.1 Development

Contemporary clinical practice guidelines, for example using the GRADE Evidence to Decision framework,[217] are developed through a process of consensus building which combines the subjective values and preferences of decision-makers/stakeholders with scientific evidence. In this context, stakeholders often make complex decisions that involve a trade-off between alternatives or interventions, based on their performance across a number of criteria or objectives, such as their clinical benefits, harms and cost. Outputs are typically expressed as evidence statements or recommendations about the alternatives (for example, a drug, therapy, or surgical intervention). Recommendations are formulated via consensus building, which may or may not be open to the lay public for a period of consultation. To improve the legitimacy of clinical practice guidelines among those effected by the recommendations, methods for grading the quality of evidence and strength of recommendations have been developed.

As an example, the National Institute of Health and Care Excellence (NICE) is an independent government body (in the UK) which utilises working groups to make decisions about recommendations for various health conditions using formal (for example, voting or Delphi technique) or informal techniques (for example, verbal agreement).[218] The working groups consist of unbiased and independent experts (or advocates) which contain at least two patients (or consumers) with experience using healthcare services. The performance of interventions are measured against a set of values which the working groups discuss until consensus is reached and recommendations are formulated. The process concludes with public engagement and endorsement from the government agency.

Implicit in this description is that the values of the working group (for example patients, clinicians, health policy professionals) must be incorporated into the decision making process, and ought to represent the preferences of all those affected by the decision. NICE state that all value judgements (preferences) made in their decision-making process reflect the values of society.[219]

However, not all organisations are so explicit in this regard. In fact, the NICE guidelines (Table 1) are the only organisation which outline their social values in healthcare decision-making. Furthermore, engagement with stakeholders is often limited – the seven guidelines included in Table 1, for example, engaged between 13-25 in their working groups, and within those groups, there were only 0-4 patient/consumer advocacy representatives. Another observation is that the guidelines were all developed in high-income countries or representatives typically located in high-income countries, which may limit their applicability in low- and middle-income countries.

The research in this thesis can address these limitations in guideline development by introducing a multi-criteria decision analysis framework that fosters inclusive, broad, systematic and transparent decision-making, which could also potentially be translated for use in low- and middle-income countries.

Nevertheless, integrating preferences and doing it transparently in health policy decision-making is difficult, however decisions cannot be made without eliciting individual values and preferences. Hence the development of clinical practice guidelines and more broadly any decision will always include subjective preference information.

2.9.1.2 Criticisms of clinical practice guidelines

A key criticism of clinical practice guidelines (CPGs) is that they do not translate well into practice. In a systematic meta-review, Correa et al. (2020)[220] for example, concluded that implementation success in clinical practice was influenced by a number of barriers and enablers which the authors described in five contexts (Table 2): political and social, health organisational, clinical practice guideline, health professional and patient.

A key observation is that the barriers and enablers to CPG implementations are influenced by many factors across the health system. For example, at the political and social level, the lack of leadership, coordination, prioritisation and funding to implement new interventions may influence the ability for the health services to provide appropriate equipment and infrastructure necessary to implement CPG recommendations. In turn, health care providers' and patients may not be able to access the necessary resources needed to deliver recommended OA care. This complex implementation environment is further complicated by a deficit in dissemination and implementation guidance, specifically in OA CPGs.[202]

Similarly, the characteristics of the CPGs – the clarity, trustworthiness of evidence and recommendations, and adaptability of CPGs to different contexts – influences the extent to which CPG recommendations are taken up by the different stakeholders in a health system (e.g. patients, health care providers and policy-makers). More effective engagement is needed with stakeholders in different contexts (e.g. from the patient through to the system manager) to help address these barriers, implement CPG recommendations, and deliver value-based OA care.

Table 2. Clinical practice guideline barriers and enablers according to the 5 contexts described by Correa et al. (2020)[220]

Context	Barriers	Enablers
Clinical practice guideline context Utility, access, adaptability, trustworthiness, quality of evidence	Lack of guideline clarity; distrust in evidence, or the belief that the evidence is insufficient to properly inform; a belief that CPGs are too rigid, challenging professional autonomy, and difficult to implement into usual practice	Interventions that were supported by clear and consistent evidence of benefit or applicability, relevant to setting; clear and simple guideline recommendations
Health organisational system context Provision of services, generation of resources, finances and administration, organisational processes, leadership, workforce characteristics	Lack of time; clinician and patient access to appropriate health services infrastructure (e.g., equipment) and services; availability of interventions; workforce health literacy (knowledge), capability (confidence) and capacity (workload); lack of access to information and information systems to store and retrieve data	Existence of multidisciplinary teams; use of technology to aid practice (e.g., automated alarms or reminders); efficient organisational processes; good communication and positive organisational culture
Health professional context Health literacy, skills, knowledge attitudes and behaviours	Poor knowledge/ignorance of CPGs or recommendations; low confidence implementing guideline recommendations; distrust in the outcomes of treatment and fear of causing harm; greater confidence in clinical experience versus guideline recommendations; lack of skills to effectively communicate, research and learn new skills	Good communication and behaviour change skills of healthcare professionals; positive attitudes towards change and evidence
Patient context Patient-Dr relationship, health literacy, motivation, beliefs and attitudes	Low disease knowledge; poor awareness of guidelines; poor compliance, motivation and family support; presence of comorbidities (e.g., depression & anxiety), polypharmacy and self-empowerment capacity; low socioeconomic status	Structured management plans for patients

Political and social context Level of support for CPG, opinion of colleagues, collaborations or networks, trust, normative frameworks	The absence of a leader who establishes priorities and manages implementation process; coordination and agreement between colleagues and in teams; absence of clear roles and responsibilities and financial support to adopt new interventions and lack of information systems to support access and storage of information.	Leadership, clear objectives and coordination across multidisciplinary teams; financial incentives to achieve positive goals of implementation; communication between care staff; appropriate use of technologies and integrated information systems

2.9.1.3Patient preferences

The past decade has seen a call to improve clinical practice guidelines by incorporating patient preferences.[216, 221, 222] Patient participation represents three important perspectives. Firstly, the ethical perspective of patient autonomy which refers to patient choice. Secondly, the psychological perspective that greater patient involvement and control in patient healthcare leads to better treatment adherence and therefore, better outcomes (patient-centred care). Thirdly, the epidemiological perspective where patients are viewed as informed stakeholders who, in conjunction with reviewing the benefits and harms of alternative treatment along with their healthcare provider, can contribute to the decision-making process along with their health provider.[216]

Given these important reflections, it is surprising that among 13 clinical practice guidelines for managing OA reviewed in 2014, only four reported patient participation in their guideline development groups.[206] Although more recent guidelines have featured consumer/patient representatives in their development panels, they still feature poorly, and in some cases given the scope of the guideline, their involvement may be quite limited. For example the 2019 OARSI[17] guidelines were developed for an international audience, yet only two patient representatives were involved, and they did not feature in the development of recommendations (Table 1).

There are many potential reasons for the low level of engagement, such as time, funding and access limitations, however one reason may be that clinical practice guideline developers don't believe that patients can make meaningful contributions to the complex decision-making involved in formulating recommendations. Fraenkel et al. (2016)[223] however, found that recommendations for rheumatoid arthritis care derived from an informed patient panel were similar to experts', unless the decision relied heavily on clinical judgement in the absence of adequate scientific evidence. The value of the patient perspective should not be underestimated.

Ultimately, patient participation in the development of clinical practice guidelines is important because it improves health service delivery across a range of settings[224] which may help support greater implementation of recommended OA care.

2.9.1.3.1 *Utility*

Another explanation for the limited uptake of clinical practice guidelines may be their poor utility in terms of what is important to end users; for example patients, healthcare providers and policy-makers. The Royal Australian College of General Practitioners guideline for the management of knee and hip OA provide a working example.[13] Included are over 70 interventions which were assigned five levels of recommendation – strong for, conditional for, conditional (neutral), conditional against and strong against. Fourteen interventions were awarded strong or conditional for recommendations, with another 17 conditional (neutral) recommendations. Whilst the guideline provides guidance about core management (long-term management), optional adjunctive management and advanced pharmacological attempts (short-term) and referral for surgical management, no guidance is provided with respect to guideline developers' preferences for the interventions within each category of recommendation (for example, 12 interventions were assigned a 'conditional for' recommendation, but are they equally valued?). Determining the overall value of each intervention, based on the explicit preferences of guideline developers, might help end-users such as patients and healthcare providers navigate clinical practice guidelines by readily presenting those interventions that are most appropriate for their circumstances in a shared decision-making paradigm (e.g. as a shortlist). For policy-makers, knowledge of which interventions align most closely with end-users could help inform, for example, health service planners, funders and coordinators make decisions about which interventions might be appropriate to offer in a service or at the population level.

2.9.1.3.2 Trustworthiness

Clinical practice guidelines over the past two decades have failed to adhere to methodological standards for developing guidelines. For example, a review of 279 guidelines published between 1985 to 1997 revealed that only 50% of the guidelines met accepted principles of guideline development at that time.[225] A similar study published in 2012[226] using the Institute of Medicine standards for developing clinical practice guidelines produced similar results. More recently, a lack of clarity and a distrust of the evidence provided in clinical practice guidelines remains a strong barrier to the uptake of clinical practice guidelines.[220]

For knee and hip OA clinical practice guidelines only, one systematic review found that 50% of the guidelines the research team identified were based on lower quality evidence and almost half of the guidelines did not disclose any conflict of interest (or when disclosed, presented multiple potential conflicts of interest).[206] This suggests that the decisions made to reach recommendations in clinical practice guidelines may be biased, or that these important considerations are not thoroughly considered or reported.[227]

Tools for evaluating the process of guideline development, such as the Appraisal of Guidelines Research and Evaluation (AGREE) tool[228] and standards, such as those produced by the Guidelines International Network[229] may help standardise the quality of clinical practice guidelines. Common to these tools is the need for transparent decision-making. A tool which can be used to guide the trustworthiness of decision-making in healthcare is the accountability for reasonableness framework (A4R). According to the framework, fairness is achieved if these four conditions are met:

1. rationales for priority setting decisions must be publicly accessible (publicity condition)
2. these rationales must be considered by fair-minded people to be relevant to priority setting in that context (relevance condition)

3. there must be an avenue for appealing these decisions and their rationales (appeals condition)
4. there must be some means, either voluntary or regulatory, of ensuring that the first three conditions are met (enforcement condition).

It is clear from this framework that conducting decision-making in a rigorous, transparent and structured approach is key, and must involve 'fair minded people', or stakeholders relevant to the decision-making process in a particular context[1]. In the context of OA clinical practice guidelines, it may be argued that items 1 and 2 are constrained or limited, as evidenced by the limited number of patients who actively contribute to guideline development groups, as well as the consideration of contextually-sensitive information that matter to people in a health system (Table 1).

2.9.2 Models of care

One important factor to consider is that any discrete effort to implement[1] recommendations for OA care into practice may be stymied by multiple barriers and enablers across different levels of a health system. One strategy which has been proposed to address this problem is system-strengthening health strategies or frameworks.

Models of Care (MoC) are a system-strengthening approach that can promote quality of care by aligning socio-political, organisational, workforce and other health system characteristics to support implementation of best-practice (i.e. clinical practice guideline recommendations). Briggs et al. (2016)[171] describe three levels of a health system which MoC can help strengthen to facilitate change (represented as inputs in Figure 3):

---

[1] Implementation is a social process that is intertwined within the context in which it takes place. Therefore, context is the set of circumstances or unique factors that surround a particular implementation effort (Damschroder et al., 2009).

1. The macro level: the functionality and scope of health systems or organisations; health policy; infrastructure and resource allocation; and socioeconomic factors.
2. The meso level: health services; clinical workforce capacity and capability; health professional and student/trainee education; service delivery systems; funding models; and clinical infrastructure.
3. The micro level: the participation of the person in his/her care.

Compared to clinical practice guidelines, MoC (which focus on *what* care is delivered) focuses on who, when, where and how care is to be delivered and re-evaluated for a specific health condition.[230] In other words, MoC aim to ensure people get the *right care*, at the *right time*, by the *right team*, in the *right place*, with the *right resources*[171] by focussing less on episodic care and more to care delivery across the continuum of disease for an individual. Thus, MoC support the benefits of value-based healthcare[29] and are considered a key driver for the practice of evidence-based health policy and practice.[231] Examples of MoC include the jurisdictional Victorian Model of Care for osteoarthritis of the hip and knee[162] and, the Australian National Osteoarthritis Strategy.[160]

The next step in the implementation of MoC are referred to as Models of Service Delivery (MoSD). Briggs et al. (2016)[171] define MoSD as:

> A model of Service Delivery is not the same as a MoC. A Model of Service Delivery operationalises the MoC and describes in detail how a given MoC is to be implemented in a local setting or health service at the operational level. A MoSD is therefore the next step in the implementation continuum.

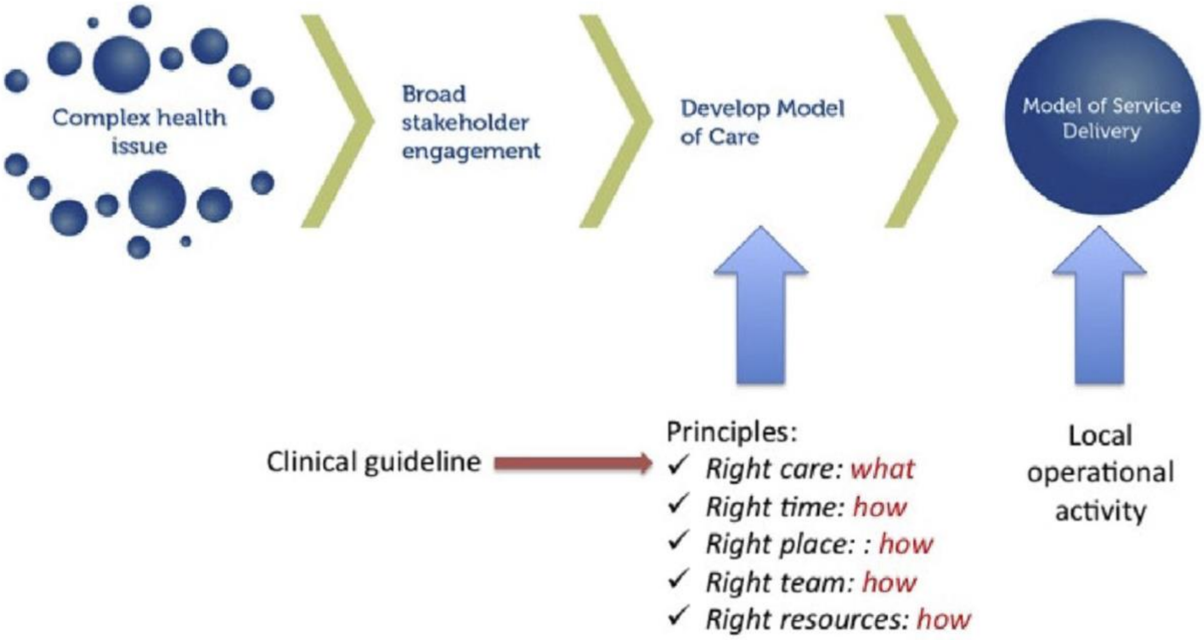


Figure 4. Schematic of the continuum between identifying a complex health problem and implementing best practice care within a local setting. The MoC provides principle-level guidance on what care and how to implement it, while the Model of Service Delivery operationalises these principles into local operational activity, informed by an implementation plan. Replicated with permission from Briggs et al. (2016)[171] (Elsevier license No. 4946840333439).

There are many different types of MoSD for arthritis that may differ in terms of their scale and mode of delivery (e.g. some may be reliant on remote health services, ‘telehealth’, or other specialised arthritis programmes).[232] Allen et al. (2016)[165] provide a review of the evidence for OA programmes across primary prevention, non-surgical management, surgical prioritisation, and management of persistent pain (for example, the Better Management of Patients With osteoarthritis (BOA) and Good Life with Osteoarthritis in Denmark (GLA;D)) and emerging PARTNER service delivery model.[233]

Importantly, MoC and MoSD strengths lie in their structured, system-wide and system-strengthening approach which are tailored for local health system context. Though unproven, it is believed that tailoring implementation strategies to local context can improve success;[234] the cross-sectoral, multi-stakeholder informed approaches described here provide strong working examples that can address multiple barriers in a health system 'simultaneously' using a coordinated approach. The important concept of 'context' is discussed in Section 2.12.1.

New Zealand currently does not have a MoSD for OA. The research in this thesis could help to support the design of a national MoSD by linking evidence with contextually appropriate information about what matters to people in the NZ health system for OA care. This could help support a nationally consistent approach to OA management, which is currently absent in NZ.[8]

2.10 QUALITY OF OSTEOARTHRITIS CARE

The World Health Organisation defines the quality of care as care that is safe, effective, timely, efficient, equitable and people-centred.[235] Up to half of people with knee OA in the health system do not access quality health services for OA.[236, 237] Over the past two decades studies continue to report that core interventions tend to be underutilised, while second-line drug therapies and referral for surgical intervention tend to be overutilised.[19, 238-244] This is, perhaps, reflected in the global increase in prevalence and disability of OA (Section 1.8). For example, Basedow and Esterman (2015)[19] systematically reviewed and meta-analysed quality indicators for OA care in 14 high quality guidelines. Quality indicators were compiled into four domains of quality (pain and functional status assessment, non-drug treatment, drug treatment and surgical referral) seen in Figure 5, which show low pass rates with respect to practicing recommended OA care. Similar results were also produced by Hagen et al. (2016)[18] reported six quality indicators covering approximately the same domains by Basedow and Esterman (2015).[19] A notable consistent result between both studies is the mismatch between the utilisation of high-value, first-line care versus second-line care and referral for surgical intervention.

More recently in 2019,[26] a large multi-country study of healthcare providers (GPs, GP registrars, primary care nurses and physiotherapists) and final year medical and physiotherapy students across Australia, Canada and NZ highlighted that confidence in OA knowledge and skills varied across healthcare professions, suggesting continued variation in the quality of OA care in the current and emerging workforce among high-income western countries. Collectively, these studies suggest that evidence-practice gaps for OA care remain an issue for delivering high-value OA care.

One important finding in the study by Briggs et al. (2019)[26] was that clinicians felt that health system factors were the biggest barrier to delivery of OA care (Figure 6). In particular, clinicians cited within-service organisational (meso-level) barriers, including: poor integration of allied health clinicals for timely/efficient referral and support; inappropriate funding models to support OA services (e.g. appropriate care facilities, equipment, limited funding to support allied health services, extended consultations); poor patient accessibly to healthcare or other facilities necessary for care (e.g. exercise venues); and poor access to readily available resources to support clinical delivery of high-value care (e.g. self-management materials for nutrition, diet, weight loss, exercise, pain management and clinical course of OA). These factors highlight opportunities for health services reform to strengthen the delivery of recommended OA care.

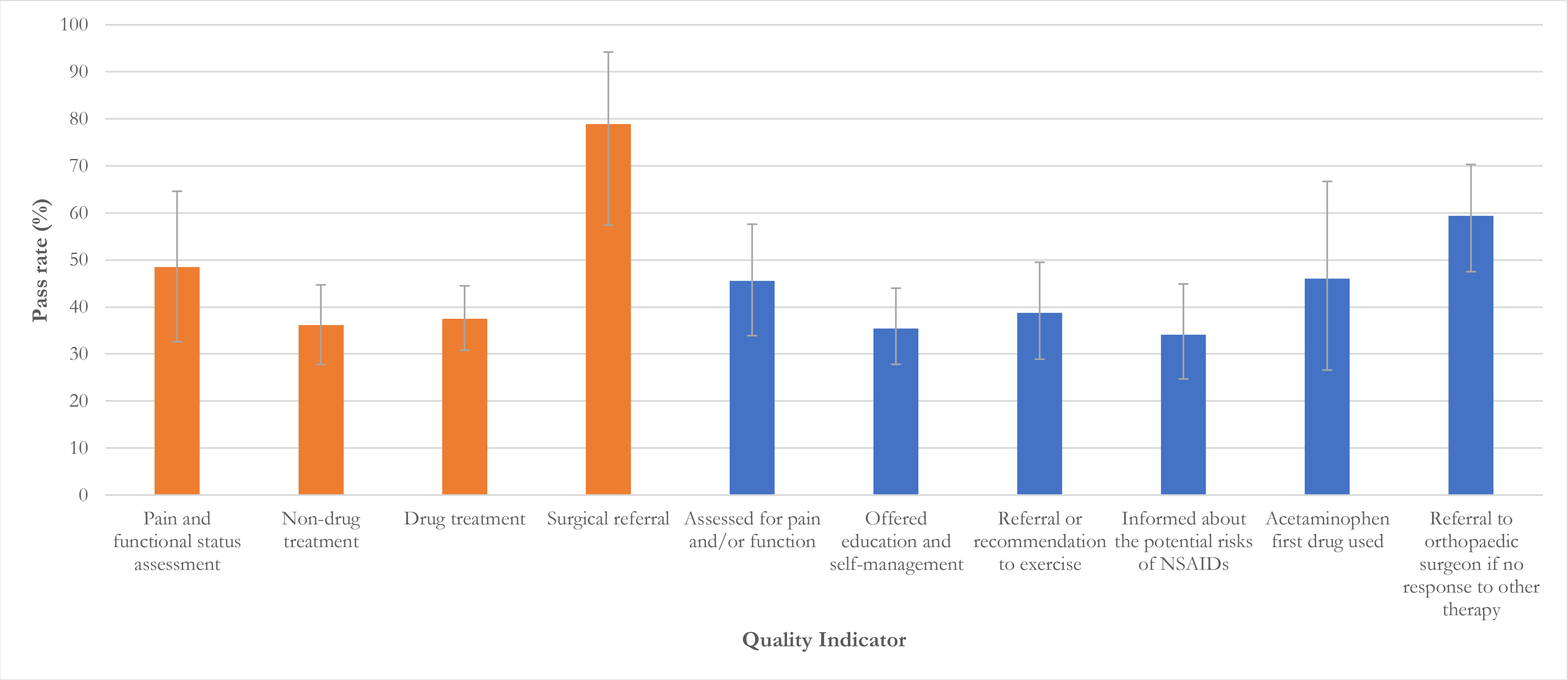


Figure 5. Pass rates and 95%CI (error bars) for osteoarthritis quality indicators extracted from two systematic reviews. The first four indicators (in orange) were extracted from Basedow and Esterman (2015),[19] and the last six (in blue) from Hagen et al. (2016).[18] Overall, the low pass rates suggest a wide gap between recommended OA care and practice.

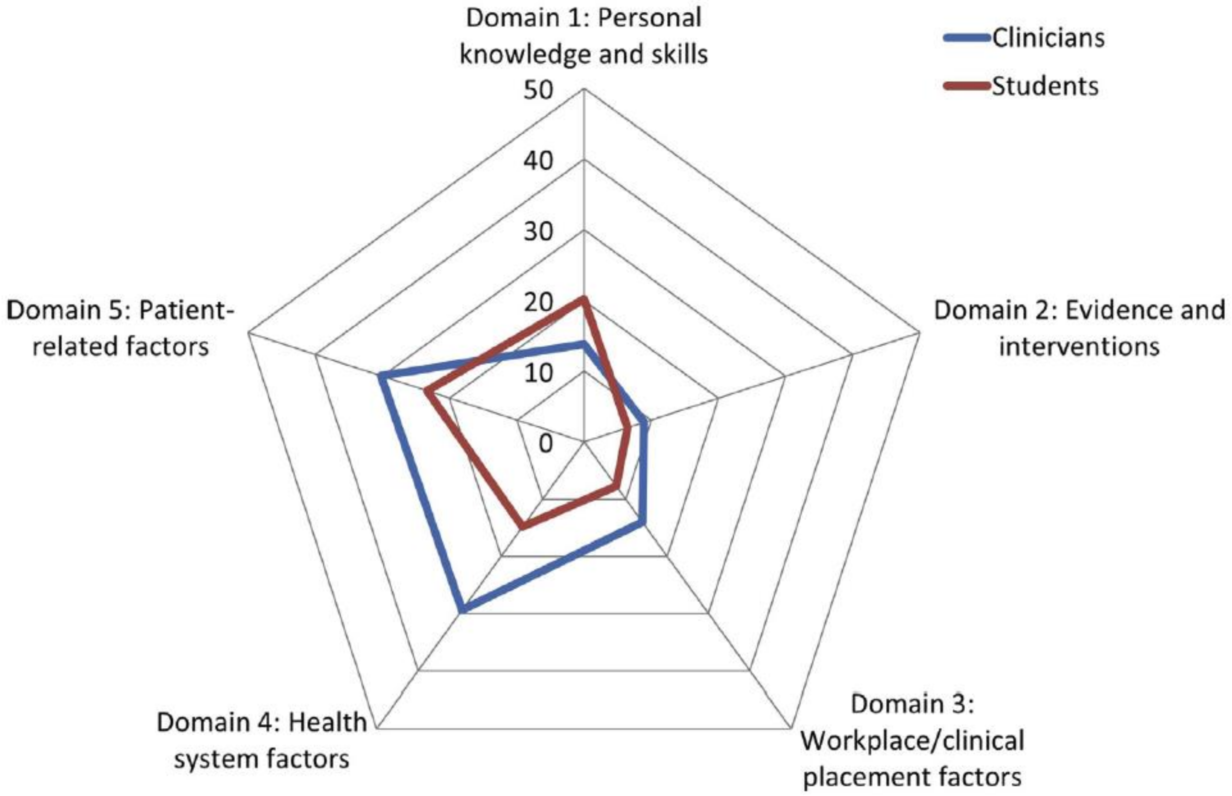


Figure 6. Radar plot of the mean number of clinicians and students (%) who reported barriers to delivery of OA care as applicable or highly applicable across five domains. Replicated with permission from Briggs et al. (2019)[26] (Elsevier license No. 4947790728037).

2.10.1 Total joint replacement

Although total joint replacement is a cost-effective option for appropriate populations it is characterised by optimal health service delivery and high initial cost. NZ for example, has a higher rate of joint replacement surgery than Denmark and the USA,[152] and the private healthcare cost for joint replacement is approximately NZ$20,600 to NZ$30,600 per procedure.[245] Coupled with rising demand for total joint replacements in NZ,[54, 152] which is also expected to quadruple from 2015 to 2030 among OECD countries,[21] total joint replacement will place greater pressure on health systems to deliver and fund these specialised, costly interventions. In Australia for example, which has similar total joint replacement costs to NZ, it is projected that the incidence of total knee and hip replacements is expected to rise 276% and 208%, respectively, at a cost to the healthcare system of A$5.3 billion, by year 2030.[246]

One important consideration is the potential to offset downstream costs by offsetting or delaying the need for total joint replacement. In Australia for example, delaying knee joint replacement with conservative interventions in 2015 could have saved up toA$170 million alone in hospital costs and yielded cost savings of up to A$233 million by 2030.[247] More recent projections, based on a conservative 50% uptake of a nationally adopted conservative OA care service, project cost savings of almost half a billion dollars saved from 2019-2029 due to avoiding total knee replacement.[248] These findings highlight that a nationally coordinated approach to management of OA can meaningfully influence the uptake of high-value OA care and offset the potential costs incurred by the need for total joint replacement.

To improve the quality of OA care Australia has embarked on a societal change approach that includes a national health strategy for OA care,[160] and jurisdictional Models of Care.(e.g. OA Chronic Care Program Model of Care[163]) In the context of NZ however, there is currently no national Model of Care for OA,[8] nor a national Model of Service Delivery, such as the comprehensive conservative care service described in the paragraph above, to help guide, support and standardise the delivery of OA care. The work in this thesis could potentially address part of these gaps by informing the development of a national Model Of Service Delivery for OA.

2.10.2 Summary

The care received by people with OA is typically suboptimal in quality despite consistent clinical practice guideline recommendations for conservative management of OA. Conservative OA care needs to be promoted to offset the physical, psychological and economic impact of and achieve the quadruple aim of value-based healthcare. The management of OA is unbalanced and a paradigm shift is required to improve the delivery of higher-value, sustainable, core interventions for managing OA.[249, 250] This is particularly relevant in New Zealand where a national approach to care is absent and there are no system-level strategies to minimise unwarranted care variation.

2.11 WHAT INFLUENCES OA CARE?

The barriers and facilitators to implementation success – particularly in primary care settings – are numerous, complex and involve stakeholders across different levels of a health system.[27, 28, 251] For example, one systematic review of reviews investigating patient adherence found that it was influenced by socio-economic-, healthcare team-, system-, condition-, intervention-, and patient-related. factors[252]

2.11.1 Patients' perspective

Patients' decision to seek medical care is mainly driven by the pain and functional limitations caused by OA and their own beliefs and attitudes about OA and what they can do about it.[24] Patients value a strong relationship with their healthcare provider who is competent and characterised by good interpersonal and communication skills that are conducive to building trust.[253]

Patients have information needs that are not being met by the health service.[254, 255] For example, patients express that they want more information from their health care provider about the disease and management options (including their risks and benefits, surgical options and complementary alternative medicines). In particular, how to engage in self-management regarding health maintenance, exercise, weight and symptom relief is deemed important to patients.[256]

Physical activity, for example, is a challenge for many people with OA.[257] Kanavaki et al. (2017)[258] report that high adherence is characterised by positive exercise experiences and beliefs, knowledge, a 'keep going' attitude, adjusting and prioritising physical activity, and having the support of healthcare providers and social groups. Conversely, low adherence to physical activity was characterised by negative physical activity experiences, beliefs and information, OA-related distress, a resigned attitude, lack of motivation, behavioural regulation and professional support and negative social comparison with co-exercisers. Unmet patient needs with respect to their attitudes may not counter unproductive attitudes to self-management (e.g. a resigned attitude towards OA).[24]

The systematic review by Dobson et al. (2016),[255] which synthesised the barriers to exercise using behavioural change theory, also found that many barriers and enablers were related to environmental context and resources and beliefs about the consequences of OA and exercise. This again highlights the role of the clinician in effectively communicating with patients to ensure that their needs are being met.

Resource factors include access to health care information; access to health care services (e.g., transport, facilities, services, wait time, appointment time, access to appropriate health care providers, time with health care providers) and cost (insurance coverage, out of pocket expenses).[24, 255]

2.11.2 Healthcare providers' perspective

Egerton et al. (2017)[25] identified four key themes explaining the barriers encountered by healthcare providers delivering OA care: (1) beliefs that OA is not that serious, (2) confidence managing people with OA, and (3) personal beliefs about OA care discordant with recommended practice and (4) dissonant patient expectations.

Clinicians may downplay the seriousness of OA due to a belief that it is inevitable part of ageing and that it is less serious than other comorbidities.[25] Furthermore, clinicians may have beliefs about OA that are counterproductive to its management. For example, negative attitudes about the disease progress, doubts about treatment effectiveness (e.g. for exercise[259] and weight loss), and negative views about patient adherence to treatment. This may lead to incomplete care and exacerbate a resigned attitude from patients, which may decrease the patients' motivation to self-manage, or to seek care because of the reinforcing belief that OA is something that cannot be managed.

Egerton et al.'s (2017)[25] review also reported that clinicians felt that they did not have adequate knowledge and confidence managing OA, such as being unfamiliar with OA treatments and CPG recommendations, and issues with CPGs in terms of their clarity and implementation. Clinicians also felt that they were time-poor to digest the information in CPGs, but felt that modifications to CPGs to improve their clarity, simplicity and applicability to more challenging patients would be beneficial. These findings suggest that healthcare providers are therefore reticent to implement high-value care. However, GPs in particular felt 'push-back' from patients who held beliefs about OA and its management that were counterproductive to building trust and delivering recommended OA care. For example, the belief that alternative interventions are effective, or that diagnostic imaging was necessary for a diagnosis of OA. A more recently a survey of healthcare clinicians and students revealed that workforce knowledge, beliefs and attitudes (including undergraduate health professional students) in some disciplines (e.g. nursing) continue to be at odds with best-practice.[260]

These healthcare provider behaviours may explain why there is a tendency to deliver OA care that is low-value and care not patient centred, as evidenced by unmet patient needs with respect to information needs about OA and its management, and the interpersonal relationship with their health care provider.

2.11.3 Health system managers' perspective

Engaging policy-makers with research evidence is important for health system reform.[261] However, policy-makers may struggle to adopt clinical practice guideline/systematic review evidence for several reasons. For example, policy-makers may disagree with the results of systematic reviews because they believe the review outcomes are not relevant to their real-world considerations.

Moreover, policy-makers may be challenged in interpreting and assessing the evidence contained in systematic reviews, leading to distrust in the evidence and recommendations. Policy-makers may also be influenced by external barriers such as clinician and patient beliefs about care that are counterintuitive to recommendations.[262] These barriers however may be overcome by collaborating with policy-makers early and throughout the development of guideline evidence and/or implementation process,[261] and developing evidence summaries that are more aligned with end-user needs (e.g. one page, plain language summary).[262]

Third sector organisations (also referred to as non-government organisations, charities, non-profits, community-based organisations, voluntary organisations or advocacy groups) may also struggle to adopt and implement evidence-based interventions if there is limited: socio-political buy-in,[27] infrastructure (e.g. finance/staff resource limitations), alignment between organisation culture (i.e. the mission-statement of the organisation or policy) and evidence-based intervention, and guidance and expertise about how to implement evidence-based interventions (e.g. adapting an intervention to a specific population without compromising effectiveness).[27, 28] Take for example physical activity interventions, where Waugh et al. (2019)[237] identified that healthcare providers' encountered system barriers included lack of time, patient compliance, resources, health workforce training, clinical networks and reimbursement.

Briggs et al. (2019)[26] provide the most up to date account of barriers related to the system which impact OA care directly, including macro-level health system factors such as financing models (which tend to support low-value care); long wait lists, poor-follow up, conflicting industry interests, and inadequate financial support for patients to access healthcare. Service-related barriers included poor networks of healthcare professionals (e.g. allied health professionals), inadequate access to facilities/venues (e.g. due to transport, cost and patient distance from venue); lack of time to provide care; and lack of funding to services to provide care (e.g. equipment, consultations and home visits).

2.12 MAKING SENSE OF COMPLEXITY

It is clear from the examples in the previous section that the implementation of interventions is influenced by many different factors (e.g. patient characteristics), stakeholders (e.g. OA consumers, healthcare providers and policy-makers) and contexts (e.g. personal, clinical, organisational and socio-political). Adopting a theoretical approach to implementation success may help to navigate this complexity. In Nilsen's (2015)[263] proposed taxonomy of approaches to explain or explore implementation efforts, he reported over 30 theories, models and frameworks, and identified one category of approaches called 'determinant frameworks'. For example, the Consolidated Framework for Implementation Research[170] and the Ecological Framework by Durlak and DuPre (2008).[264] Determinant frameworks are particularly useful for understanding and/or explaining what influences implementation outcomes (e.g. health care providers' attitudes and beliefs about OA). Importantly, determinant frameworks recognise that implementation is multidimensional in nature, with many interdependent factors; hence, they are typically multilevel encompassing various stakeholders from the micro through to the macro level of implementation.[263] Determinant frameworks are therefore particularly suited to understanding and explaining what influences the quality of OA care.

In a systematic review of reviews, Lau et al. (2016)[27] aimed to identify the causes of the evidence-practice gap in primary care. Lau et al. (2016)[27] described a conceptual (determinant) framework of factors influencing change in health systems across four dimensions. Specifically (1) external context in which implementation was taking place; (2) organisational features; (3) characteristics of health professionals involved, and (4) the characteristics of interventions, which cut across all the other dimensions. The authors concluded that no single strategy maximised implementation and that the **context** in which interventions took place – often overlooked and not acknowledged in research – likely mediated the successfulness of any one implementation effort (Figure 7).

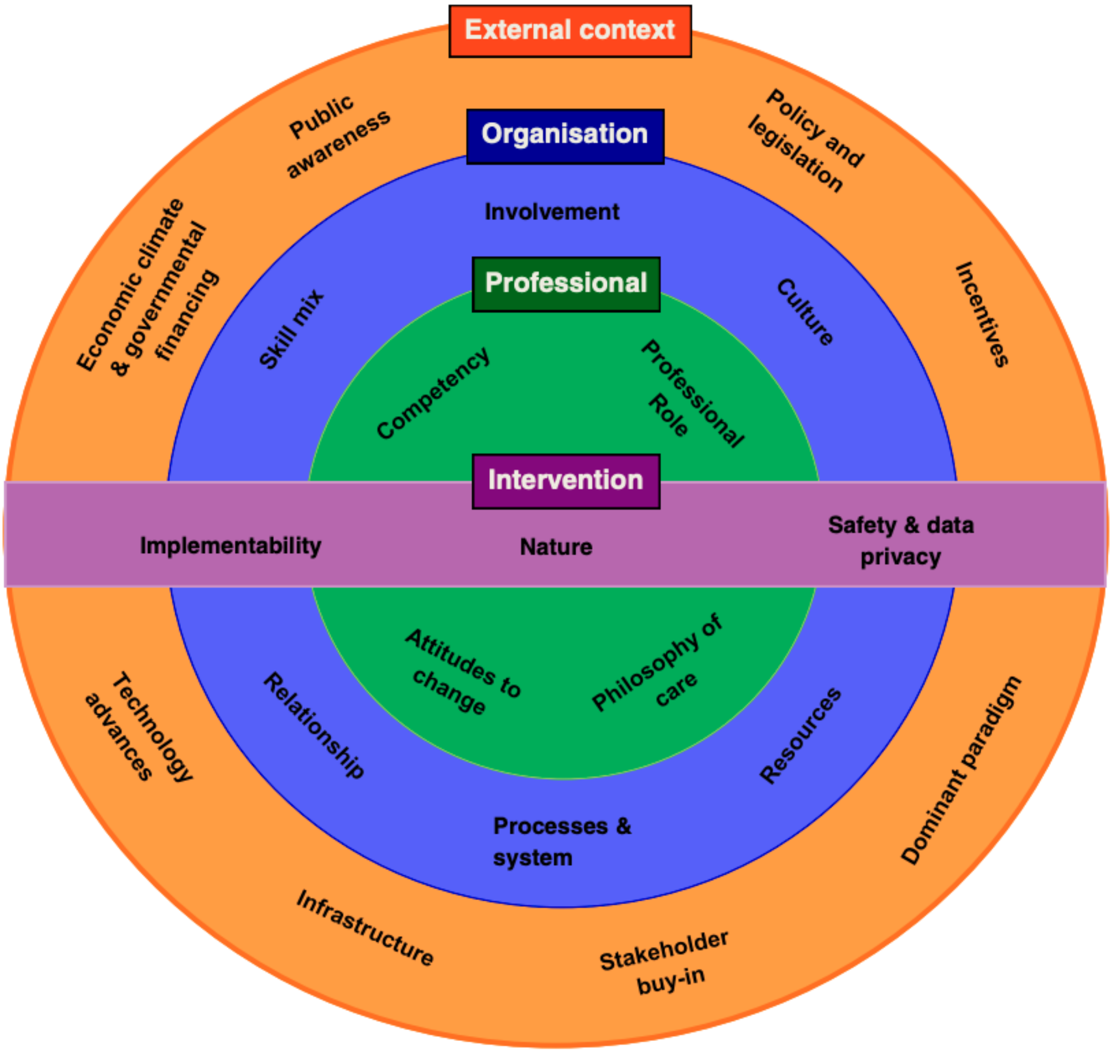


Figure 7. Conceptual framework describing key elements that influence implementation of change in primary care. Replicated from Lau et al. (2016)[27] with permission from Springer Nature (Creative Commons CC BY license).

2.12.1 Context is key

Context is defined as the environmental conditions which influence the barriers and enablers of implementation.[170] Implementation success is influenced by the interrelationships across the macro, meso and micro levels in a health system. Lau et al. (2016) [170] describe this as the connections between the external context (e.g. current policy, national or local agenda, existing infrastructure) and the organisation's existing work practices (e.g. culture, readiness, relationships and leadership) including their beliefs and values (professional attributes) and the attributes or characteristics of interventions which affect their preference, implementability or desirability across the different levels. For example, the implementation of an exercise intervention of OA is influenced by the socio-political environment in which it is being implemented, the health system's workforce readiness to adopt the intervention, the healthcare providers' beliefs and attitudes about exercise for OA, and the 'fit' of the exercise intervention across the different levels of the health system.

2.12.2 'Fit' for implementation

The 'fit' of an intervention refers to stakeholders' preferences for the attributes, characteristics or features of interventions within a health system. The more an intervention aligns with stakeholders' preferences for their characteristics, the less likely there will be resistance to uptake, and more likely implementation success (Figure 8).

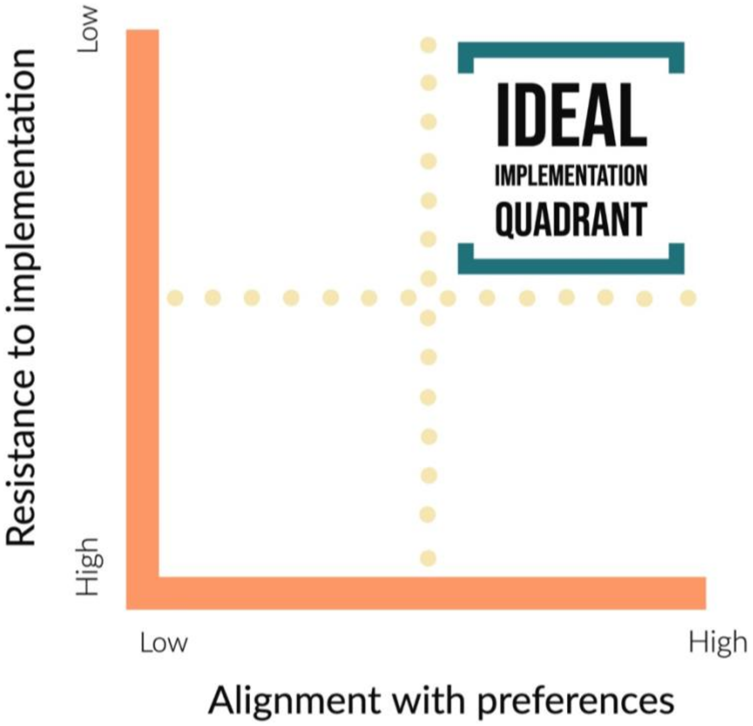


Figure 8. The likelihood of implementation success is enhanced when interventions are aligned with the preferences of stakeholders.

Given that Lau et al.'s (2016)[27] theoretical framework aligned with other well established theoretical frameworks for implementation, such as the Consolidated Framework For Implementation Research,[170] it is unsurprising then that the characteristics of interventions identified in Figure 9 are not dissimilar and highly generalisable.

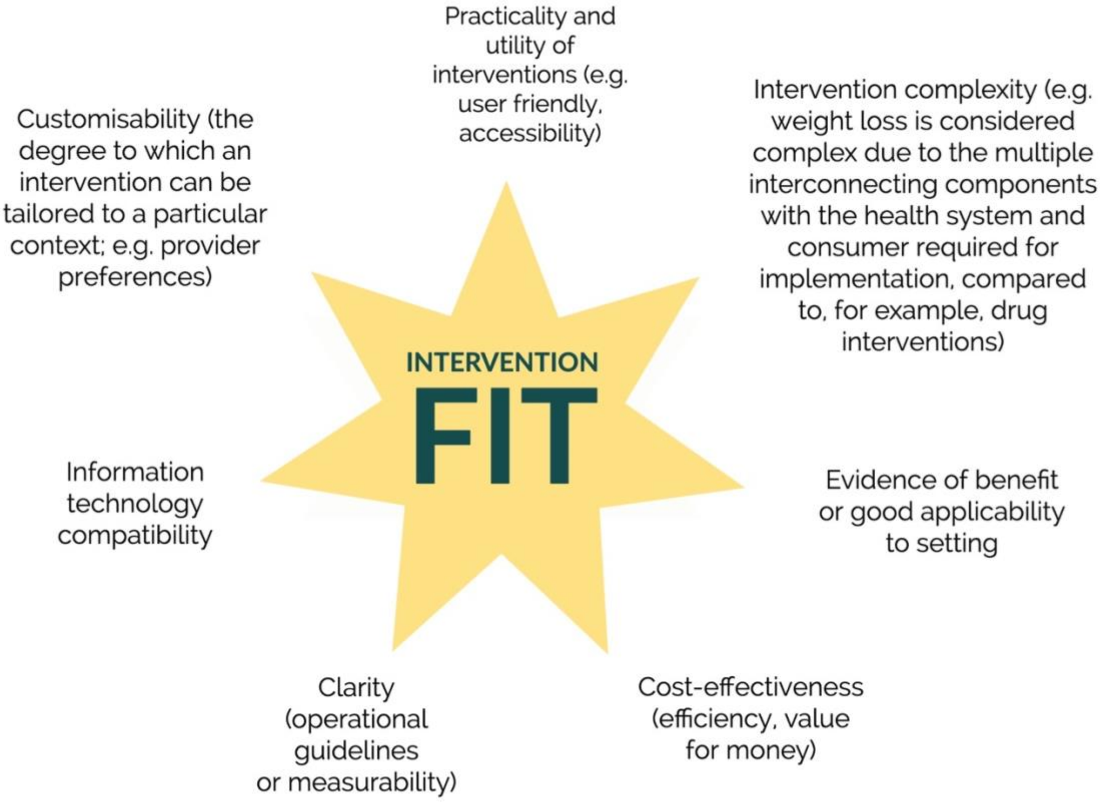


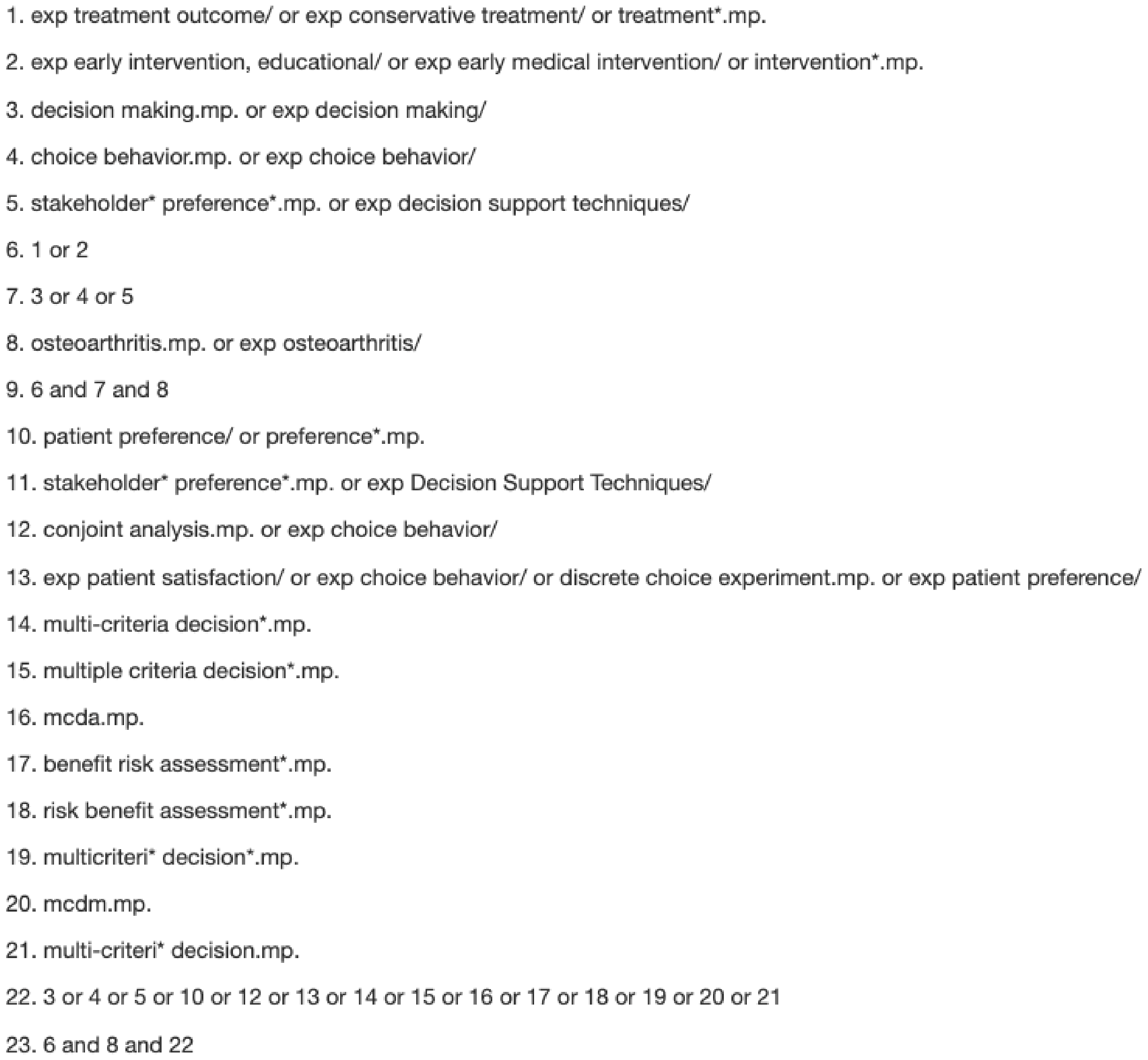
Figure 9. The characteristics of interventions proposed by Lau et al. (2016)[27] which influence implementation success

The next section will specifically address what characteristics of interventions for OA interventions have been studied in the literature to assess the 'fit' between stakeholders' preferences and the characteristics of interventions.

2.13 STAKEHOLDERS' PREFERENCES

The introductory chapter introduced multi-criteria decision analysis (MCDA) as a key methodology for making transparent, systematic and inclusive decisions.[265, 266] Indeed, MCDA is becoming increasingly prevalent in health care research, supporting decision-making for: (i) prioritisation of interventions for coverage or reimbursement (investment); (ii) selection of intervention (prescription); (iii) assessment for licensing (authorisation); and allocation of research funds.[267, 268] However, there are currently no studies which use MCDA to evaluate OA interventions – a gap that the research in this thesis can address. However, studies have explored the preferences of stakeholders for the characteristics of OA interventions using methods appropriate for MCDA. For example, discrete choice experiments and conjoint analysis. These methods are characterised by surveys which involve respondents making choices between criteria that typically represent the characteristics of OA care. Therefore, the focus of this section will be to discuss the literature concerning the preferences of stakeholders for the characteristics of knee OA interventions using methods relevant to MCDA. Methods for eliciting stakeholders' preferences are discussed in Chapter 5, section 5.3.

To identify studies investigating preferences for OA care the following search strategy was undertaken in July 2020. Search strategies and keywords were extracted and adapted from a systematic review of discrete choice experiments[37] and MCDA[268] studies in healthcare research. Using the keywords identified, Ovid Medline was searched with no date restriction and limited to full-text peer reviewed papers in English. The keyword search strategy is shown in Figure 10. Editorials, commentaries, protocols, conference abstracts and dissertations were excluded, including papers which did not report methods and/or analyses. To be eligible, studies must have included a choice-based activity for interventions for OA care. Google scholar was also searched to check for additional papers retrieved in the first three pages. Titles were reviewed, followed by their abstracts. The search produced 2153 articles and after screening for eligibility 23 articles remained which are discussed in paragraphs below. It should be noted that as I was the only person who completed the search and review, which did not include grey literature, there is a possibility that this narrative review is not exhaustive.



1. exp treatment outcome/ or exp conservative treatment/ or treatment*.mp.
2. exp early intervention, educational/ or exp early medical intervention/ or intervention*.mp.
3. decision making.mp. or exp decision making/
4. choice behavior.mp. or exp choice behavior/
5. stakeholder* preference*.mp. or exp decision support techniques/
6. 1 or 2
7. 3 or 4 or 5
8. osteoarthritis.mp. or exp osteoarthritis/
9. 6 and 7 and 8
10. patient preference/ or preference*.mp.
11. stakeholder* preference*.mp. or exp Decision Support Techniques/
12. conjoint analysis.mp. or exp choice behavior/
13. exp patient satisfaction/ or exp choice behavior/ or discrete choice experiment.mp. or exp patient preference/
14. multi-criteria decision*.mp.
15. multiple criteria decision*.mp.
16. mcda.mp.
17. benefit risk assessment*.mp.
18. risk benefit assessment*.mp.
19. multicriteri* decision*.mp.
20. mcdm.mp.
21. multi-criteri* decision.mp.
22. 3 or 4 or 5 or 10 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21
23. 6 and 8 and 22

Figure 10. The Ovid Medline search strategy used to identify studies exploring preferences for OA care.

With respect to whose stakeholders' preferences have been studied, the most widely studied groups are of people experiencing OA followed by health care providers from high-income countries (Table 3). Currently, there is only one study which has explored the preferences of other stakeholders for the characteristics of OA interventions, namely insurance company employees.[269] There are also only a few studies which have explored patient and healthcare provider preferences conjointly.[269-271] This highlights an important gap in the literature with respect to understanding the preferences of cross-sectoral stakeholders in a health system.

In general, most studies have focussed on the characteristics of drug treatments[270-283] that drive treatment choices. For example, Berchi et al. (2016)[281] explored health care providers' preferences for seven characteristics, and found, in decreasing order of importance: pain relief, improvement in function, retardation of joint degradation, annual cost to the patient, risk of moderate side effects, risk of serious side effects, degree of patient acceptance of treatment. Thus, the characteristics of interventions chosen across the studies in Table 3 mainly concern health outcomes related to the benefits and harms of drug treatment (e.g., pain and functional improvement vs risk of bleeding ulcer and stroke). The most common non-health related outcomes were related to the out-of-pocket costs of treatment and the route of administration (e.g., oral vs injection). Other types of intervention characteristics considered in the literature also include all treatment types,[269, 284, 285] physical activity,[38] exercise programs,[286] exercise and drug treatments[287] and surgical treatments.[39, 288, 289]

Due to methodological differences such as question framing and the characteristics of interventions chosen, it is difficult to make direct comparisons between the studies. For example, two studies chose to examine patients' preferences for the risk of harm of treatment, but described the risks of harm differently. Other contextual differences, such as government co-payments for treatment may also influence preferences for certain characteristics of interventions. Notwithstanding these limitations, there are generally mixed outcomes with respect to some studies showing that patients value the benefit over the harms of treatment,[277, 281, 284, 285, 288] and other studies suggesting that the opposite is true.[270, 274, 276, 278, 280, 282] In the studies of preferences for physical activity/exercise,[38, 286] patients tended to value intervention characteristics related to convenience and benefit, though each study differed in terms of the value which patients placed on the cost of intervention.

The only three studies of patients and other stakeholders (healthcare providers and insurance employees) preferences suggest that their preferences may differ by group. For example the multi-country study (France, Germany, Italy, Spain, Sweden, and the United Kingdom) exploring preferences for characteristics of opioid treatment reported that patients were more concerned about the risk of nausea compared to clinicians, who placed more importance on pain control.[271] Similarly the study by Byun et al. (2016)[270] reported that patient and health care providers' preferences for characteristics of cyclooxygenase-2 inhibitors differed in terms of their preferences for the benefits and risks of treatment. Patients valued the treatment effect on function more relative to the risk of cardiovascular disease, whereas the opposite was true for clinicians. However, the study only included 158 participants and did not report pilot testing the survey instrument prior to data collection; factors which may have introduced sampling bias and systematic error into the results. The study by Arslan et al. (2020),[269] which compared patients', health care providers' and insurance employees' preferences for characteristics of OA care – which only considered characteristics of OA care that were a non-health outcome or process-related – reported that patients and health care providers place greater importance on out of pocket expenses than insurance company employees. Health care providers and insurance company employees also placed more importance on the duration of consultation than patients.

Another limitation in the literature identified in Table 3 is that the sample population is generally about the age of 50 years and from high income countries. The study findings may therefore miss an important and potentially growing population of younger people experiencing OA, whose preferences may differ from elder population groups. Some studies have reported associations between demographic characteristics and preferences. For example Ratcliffe et al. (2004)[284] found that preferences for risk of serious harm decreased as age increased.

There are also some methodological concerns of the studies included in Table 3 that warrant consideration. For example, some of the studies did not report how the characteristics of interventions were identified,[272, 273, 277, 278, 287] while others reported only referring to the literature to identify the intervention characteristics.[39, 270, 275, 276, 280] Other studies did not report if the survey instrument was pilot tested prior to data collection.[270-273, 275-278, 282, 287, 288] These factors may lead to spurious results due to important characteristics of interventions not being explicitly considered in the survey instrument choices, or the unintended interpretation of how the characteristics are worded. Ideally, studies should use qualitative methods to inform the selection of important characteristics of interventions to consider in conjunction with other methods, such as a literature review.[266, 290] For example, qualitative interviews were used to identify the characteristics of opioid treatments for OA[271] and the characteristics of physical activity important to people experiencing OA.[38, 291] Qualitative methods used in this thesis are introduced in Chapter 3.

Table 3. Summary of studies investigating the characteristics of OA interventions. DCE, Discrete Choice Experiment; NR, Not Reported

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Al-Omari et al. (2017)[283] UK	11 patients, NR	To explore how conjoint analysis can be used to facilitate shared decision making. Adaptive conjoint analysis	Systematic search', feasibility study, recommendation from an earlier study that more side-effect attributes should be included than in previous choice experiments relating to NSAIDS for OA, recommendations from a reference user group. Pilot survey	Drug treatments. 8. Risk of kidney and liver side effects, risk of heart attacks and strokes, risk of stomach side effects, availability, expected benefit, risk of addiction, frequency of taking, way of taking the medication.	There was a high degree of heterogeneity in the study sample. The authors concluded that preferences highly variable at the individual level. Choices were driven by the risks of harm.
Arden et al. (2012)[282] UK	475 health care providers (291 GPs, 89 orthopaedics, 72 rheumatologists, 25 internists), 43±10.1	To quantify the relative importance that physicians attach to the benefits and risks of drugs treatments. DCE	Recommended clinician attributes, clinical outcomes from WOMAC chosen by the authors which were validated among 10 physicians and 10 patients using face to face interviews. NR	Drug treatments. 7. Benefits - easing of ambulatory pain, resting pain, stiffness, daily activities); Risks - (bleeding ulcer, stroke, MI).	Physicians placed greater importance on risk than pain reduction; ambulatory pain trumped resting pain; little importance was placed on moderate to mild pain; physicians weighted benefits and risks equally, regardless of patient characteristics, when analysed by speciality. GP and specialist preferences were the same.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Arslan et al. (2020)[269] Netherlands	874; 648 patients, 61.7±NR; 76 healthcare providers, age NR; 150 insurance company employees, age NR	To determine patients', healthcare providers', and insurance company employees' preferences for knee and hip OA (KHOA) care. DCE	Literature review, interviews with key stakeholders (experts), Criteria levels informed by experts and nationally relevant published data. Pilot survey	Knee & Hip OA Care. 6. (for consumer group) out of pocket costs, healthcare providers during consultation, waiting time per week, travel distance, access to specialist equipment, duration of consultation.	Patients and healthcare providers valued low out of pocket costs the most, while insurance company employees valued the involved healthcare providers during consultation as most important (followed by out-of-pocket costs).
Berchi et al. (2016)[281] France	188 health care providers (106 GPs, 82 Rheumatologists), age NR	To determine whether out of pocket costs influence physicians' treatment choices. DCE	Literature review & interviews with OA experts. Pilot survey	Drug treatments. 7. Pain relief, improvement in function, retardation of joint degradation, annual cost to the patient, risk of moderate side effects, risk of serious side effects, degree of patient acceptance of treatment.	Physicians take into account out-of-pocket costs. Changes in co-payments for some OA treatments may lead to changes in prescribing practices.

Reference, country	Sample population, mean age ±SD	Aim Methods	How were the criteria Identified? Was the study piloted?	Criteria scope Number of criteria in decreasing order of importance	Main finding
Byrne et al. (2006)[288] US	391; 193 no OA, 198 consumers with OA, 55±14.3	To determine whether there are ethnic differences in preferences for surgery vs. medical treatment of knee OA. Conjoint analysis	NR. NR	Surgical treatments. 6. Outcomes/attributes of total knee replacement: pain, walk, cost, death, complications, failure.	African Americans are: less likely than whites or Hispanics to choose surgical intervention for knee OA; and valued walking ability more than Whites and Hispanics.
Byun et al. (2016)[270] Korea	98 patients with OA or RA, 55±7.8; 60 orthopaedic surgeons, 40±4.9	To elucidate and compare benefit–risk preferences among Korean patients and physicians concerning cyclooxygenase-2 inhibitor treatments for arthritis. DCE	Literature review. NR	Drug treatments. 4. Patients - pain improvement, functional improvement, cardiovascular (CV) risk, gastrointestinal (GI) risk. Physicians - CV risk, pain improvement, GI risk, functional improvement.	Patients placed more value on benefit attributes than risk attributes; the reverse was true for the physicians.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Chancellor et al. (2012)[271] Europe (France, Germany, Italy, Spain, Sweden, and the United Kingdom)	242 pain (cancer and OA/chronic pain sufferers); 270 physicians; mean age NR	To identify the attributes of greatest interest to physicians and pain sufferers when they consider prescribing or taking strong opioids for chronic pain. DCE	Focus group & structured interview plus ranking exercise. NR	Drug treatments. 5. Patients ranked nausea, pain impact, energy, alertness, and constipation; physicians ranked pain response, central nervous system effects, nausea, dose form, and constipation in descending order of importance.	Participants were unwilling trade severe side effects to decrease pain, whereas physicians were willing to trade between the criteria.
Copsey et al. (2019)[280] UK	300 consumers, 60±13.3	To investigate if duration of treatment effect should be considered in a benefit-risk assessment using a case study of OA medications. DCE	Literature review, and a ranking exercise completed by 10 patient representatives. Pilot survey	Drug treatments. 6. Stomach ulcer bleeding risk, Duration of treatment effect, Pain, Heart Attack Risk, Function, Stiffness.	Patients are willing to accept drugs that are less effective if they are longer lasting.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Dennis et al. (2020)[279] US	602 patients; 70 with OA only, 66 with OA and chronic low back pain (CLBP), 110 with CLBP only; 63.7±10.8	To quantify preferences for attributes of drug treatments. DCE, Best worst scale (BWS)	Focus groups; clinical experts in rheumatology and chronic pain, 'soft launched' to participants following internal pilot testing with research team. Pilot survey	Drug treatments. 6 (DCE): symptom control; reducing risk of physical dependence; heart attack, mode/frequency of administration and cost. 10 (BWS): stroke, physical dependence on pain medicine, risk of heart attack due to medicine, increased risk of severe joint problems because of medicine, and a risk of bleeding ulcer when first starting a medicine.	A pharmaceutical treatment with a risk of severe joint problems was viewed as an acceptable alternative to other treatments with comparable efficacy but risks associated with NSAIDs or opioids.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Fraenkel et al. (2004)[278] US	100 patients, 69±6.0	To assess the influence of providing an additional treatment alternative on patient preferences for nonselective nonsteroidal anti-inflammatory drugs (NSAIDs) and cyclooxygenase-2 (COX-2) inhibitors.	NR; Face and content validity of the attribute descriptions were assessed by rheumatologists, conjoint experts and lay persons. NR	Drug treatments. 7. Importance NR: route of administration, onset of action, chance of benefit, common adverse events, risk of annual ulcer, cost.	Patients are unaware of drug risks; their preferences are influenced by risk of adverse events.
Fraenkel et al. (2004)[272] US	100 patients, 70±7	To test if risk of serious adverse events is related to merely risk reduction between COX-2 inhibitors and NSAID choice. Adaptive Conjoint Analysis	NR. NR	Drug treatments. 7. Importance NR: route of administration, onset of action, chance of benefit, common adverse events, risk of annual ulcer, cost.	Patients' preferences for COX-2 inhibitors may be mediated in part by a perception that these drugs eliminate, as opposed to reduce, the risk of toxicity.
Fraenkel et al. (2004)[273] US	100 patients, 70±7	To examine patient treatment preferences for knee OA. Adaptive Conjoint Analysis	NR; face and content validity reviewed by rheumatologists, conjoint analysis experts and lay persons. NR	Drug treatments. 7. Gastro ulcer, common adverse events, chance of benefit, time to benefit, cost, route of administration, label.	No associations between demographic characteristics, clinical characteristics and treatment preference. Weights for the most important 3 criteria were very similar, suggesting consideration of multiple criteria to arrive at a decision.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Fraenkel et al. (2008)[287] US	90 patients, 68±9	To compare patient preferences for exercise in comparison to other options. DCE	NR. NR	Exercise and drug treatments. 5. Route of administration, risky of dyspepsia, risk of bleeding ulcer, decrease in pain, improved strength.	Patients prefer exercise when presented with the risks of drug treatment.
Fraenkel et al. (2014)[277] US	304 patients, 57 (range 34-89)	To quantify patient preferences for potential disease modifying osteoarthritis drugs Conjoint analysis	NR. NR	Drug treatments. 4. Benefit, Risk, Cost, Administration.	A total of 180 participants were willing to try a parenteral medication which benefits 40% of patients and is associated with a serious risk of infection requiring prolonged hospitalisation. The findings suggest patients are willing to accept greater risk to preserve joint health.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Franco et al. (2015)[286] Australia	115 patients, 68±6	To explore older people's preferences in relation to the characteristics of exercise programs, and to examine the relative value placed on these particular attributes. DCE	Qualitative systematic review, consultation with experts in the field of aging and DCE experts. Pilot survey	Exercise programs. 9. Exercise type, transport type, out of pocket expenses, improvement in the ability to undertake daily activity tasks at home, chance of falling, improvement in the ability to leave the house or undertake tasks or socialise, travel time, frequency & time spent on exercise.	Older people place higher values on exercise characteristics than on their actual benefits; access to treatment is more importance than benefit.
Harris et al. (2018)[289] US	200 OA patients, 64.3±9.32	To examine which aspects of proximal interphalangeal joint surgery matter most to respondents. Conjoint analysis	Literature review & survey of surgeons to identify most important attributes. Pilot survey	Surgical treatments. 5. Joint stiffness, grip strength, need for future surgery, cost, recovery time.	Benefits of treatment outweighed the need for future surgery, cost and recovery time.

Reference, country	Sample population, mean age ±SD	Aim Methods	How were the criteria Identified? Was the study piloted?	Criteria scope Number of criteria in decreasing order of importance	Main finding
Hauber et al. (2013)[276] UK	289 patients, 59±8.4	To estimate OA patients' risk tolerance for serious adverse events including bleeding ulcer, myocardial infarction and stroke. DCE	Package inserts, clinical trial literature and clinical experts. NR	Drug treatments. 7. Myocardial infarction, Stroke, difficulty doing daily activities, bleeding ulcer risk, resting pain and stiffness.	Patients attached greater importance to eliminating the risks of adverse events than in reducing pain. Risk tolerance varied according to the baseline level and type of symptom relief.
Hilligsmann et al. (2020)[285] Europe (Belgium, France, Italy, Netherlands, Portugal, Spain and UK)	253 patients, 71.3± NR	To evaluate the preferences of patients with OA for treatment. DCE	Two scoping reviews, interviews with patients, experts, patient survey to rank most important criteria, expert consensus meeting. Pilot survey	All treatments. 7. Disease progression (joint structure), walking improvement, pain improvement, inability to manage domestic abilities, improvement in overall energy and well-being, manage social activities, risk of severe side-effects.	The most important outcomes were impact on disease progression and improvement. All criteria were considered important.

Reference, country	**Sample population, mean age ±SD**	**Aim** **Methods**	**How were the criteria Identified?** **Was the study piloted?**	**Criteria scope** **Number of criteria in decreasing order of importance**	**Main finding**
Laba et al. (2013)[274] Australia	188 patients, 62±8	This study aimed to estimate the relative influence of medication-related factors and respondent characteristics on decisions to continue medications among people with symptomatic OA. DCE	Literature review & context specific treatment information plus survey pilot testing. Pilot survey	Drug treatments. 7. Risk of high blood pressure, heart/liver/kidney problems as side effects, out-of-pocket costs, the possibility of heartburn/ reflux, or stomach ulcers as side effects, treatment schedule (i.e.: daily versus when required), mode of action (slowing OA versus symptomatic pain relief) and the possibility of drowsiness or constipation as a side effect.	Treatment factors, as opposed to respondent characteristics including self-reported pain levels and physical functioning, were driving adherence decisions. Treatment efficacy did not drive preferences.

Reference, country	Sample population, mean age ±SD	Aim Methods	How were the criteria Identified? Was the study piloted?	Criteria scope Number of criteria in decreasing order of importance	Main finding
Moorman et al. (2017)[39] US	323 consumers, NR	To determine the relative importance of different attributes of surgical treatments for knee OA. Conjoint analysis	Research group & consultation with an orthopaedic surgeon. Pilot survey	Surgical treatments. 9. Amount of cutting and removal of existing bone required, chance of additional surgery, amount of pain relief, limits or complicates any future treatments needs on the knee, length of hospital stay required.	Risks of surgery drives preferences; ~50% preferred no surgery, and 32.5% preferred KineSpring System, 7.1% unilateral knee arthroplasty, 5.9 high tibial osteotomy, 5.3% for total knee arthroplasty.
Pinto et al. (2019)[38] US	146 consumers, 65±NR	To investigate individual preferences for physical activity (PA) attributes in adults with chronic knee pain, to identify clusters of individuals with similar preferences, and to identify whether individuals in these clusters differ by their demographic and health characteristics Adaptive conjoint analysis	Focus group discussion. Pilot survey	Physical activity. 6. Health benefit, enjoyment, convenience, PA effort of PA, PA Cost, time needed to engage in PA	Patient preferences for OA interventions are not homogenous.

Reference, country	Sample population, mean age ±SD	Aim Methods	How were the criteria Identified? Was the study piloted?	Criteria scope Number of criteria in decreasing order of importance	Main finding
Posnett et al. (2015)[275] Europe (France, Germany, Italy, Spain, and the UK)	2073 patients, 75±NR	Secondary aim: identify preferred attributes of OA knee treatments. Conjoint analysis	Package inserts, proprietary data held by Sanofi. NR	Drug treatments. 3. Co-pay, duration of pain relief, effect on pain, injection, time to pain relief, impact on joint progression.	The treatments most commonly received by patients with OA knee are not generally the same as the treatments that score highest on measures of perceived effectiveness or of patient satisfaction.
Ratcliffe et al. (2004)[284] UK	412 patients, 72% of the sample was aged 61-80yrs	To investigate patient preferences for attributes associated with the efficacy and side-effects of treatment for OA. Conjoint analysis	Literature review & interview with rheumatologists and patients. Pilot survey	All treatments. 5. physical mobility, joint aches, risk of serious side-effects, joint pains, risk of mild to moderate side-effects.	Participants were willing to trade improvements in physical mobility for risk of side-effects.

2.13.1 Summary

The purpose of this section was to describe the studies exploring stakeholders' preferences for OA interventions using MCDA. The key findings are:

- MCDA has not been used to inform choices between OA interventions.
- There is an evidence gap with respect to the preferences of cross-sectoral stakeholders; studies mainly focussed on the preferences of patients and health care providers.
- There is an evidence gap with respect to identifying generalisable characteristics of interventions; only three studies considered generalisable characteristics of OA interventions; the majority considered drug treatments.
- The preferences of stakeholders for OA interventions in low- to middle-income countries is an area for future research.
- Exploration of younger people's preferences for OA interventions is an area for future research in general.
- How the criteria were identified and selected in studies was variably reported, and few studies were grounded in qualitative research.

2.14 WHERE TO FROM HERE?

This chapter established that OA is a major cause of pain and morbidity, resulting in huge social and economic costs worldwide, including New Zealand. The previous sections also established that clinical practice guidelines are important tools for translating evidence into practice; however, they are not without their short-comings in terms of accessibility, broad stakeholder engagement, clinical usefulness and trustworthiness. International CPG for OA care consistently recommend interventions of exercise, self-management education and weight-loss for OA care (Table 1). However, these high-value interventions are not routinely delivered to or taken up by people living with OA, resulting in a tendency to deliver low-value care and unwanted care variation. This suggests that a problem with knowledge translation into policy and practice, leading to the observed evidence-practice gaps for OA care.

An examination of the factors influencing implementation reveals a complex mix of barriers and enablers, involving many stakeholders across different levels of a health system, with no clear dominant strategy implementation for success. However, a common thread that cuts across this complexity is the concept of context, whereby successfully implementing an intervention in one context may not necessarily be the same in another context. Evaluating the 'fit' of an intervention within a particular context is therefore desirable to support implementation efforts. Indeed, context is a key feature in the development, implementation and evaluation of musculoskeletal Models of Care.[292] For example, it is recommended that MoCs should target a local health issue, defining and describing the problem using local data, including the preferences of local health service providers.[292, 293]

Closing the evidence-practice gap for OA should involve broad consultation with stakeholders across the health system about what matters to them – what their preferences are for recommended OA care – and redesign of the health system based on their needs and the best-evidence about OA interventions. This philosophy is reflected in the seminal publication by Speerin et al. (2020).[29] Indeed, primary care has signalled a need for broader engagement with central stakeholders, such as funders, researchers, policy-makers, and healthcare providers to help inform how their needs could be met to implement evidence-based interventions into their particular context.[28]

Given that CPGs are a foundational reference point for what OA care to deliver in a health system, but are typically not developed with local end-users in mind, this begs the question: what matters to cross-sectoral stakeholders when choosing OA interventions, and which interventions do they want in a CPG based on what's deemed important?...which interventions are good 'fit'? Combining the preferences of cross-sectoral stakeholders, with evidence and recommendations in a CPG to identify interventions with good 'fit' could answer these questions. Greater alignment between cross-sectoral stakeholders' preferences and recommendations for OA care could potentially help guide or support system-strengthening strategies such as an national MoSD in NZ (Figure 11).

The preferences of cross-sectoral stakeholders for OA interventions are poorly studied in the literature, focussing primarily on patient and health care provider preferences for drug interventions. An evidence gap exists with respect to cross-sectoral stakeholders' preferences for OA interventions with good 'fit' in a health system. MCDA can address this evidence-gap by meeting the demands of the problem: combining cross-sectoral preferences with CPG evidence and recommendations to prioritise OA interventions in an accessible, inclusive, systematic and transparent process. Thus, the overall aim of this thesis is to establish a framework to prioritise knee interventions for managing knee OA and evaluate the relative importance of these interventions across the healthcare sector in New Zealand using MCDA.

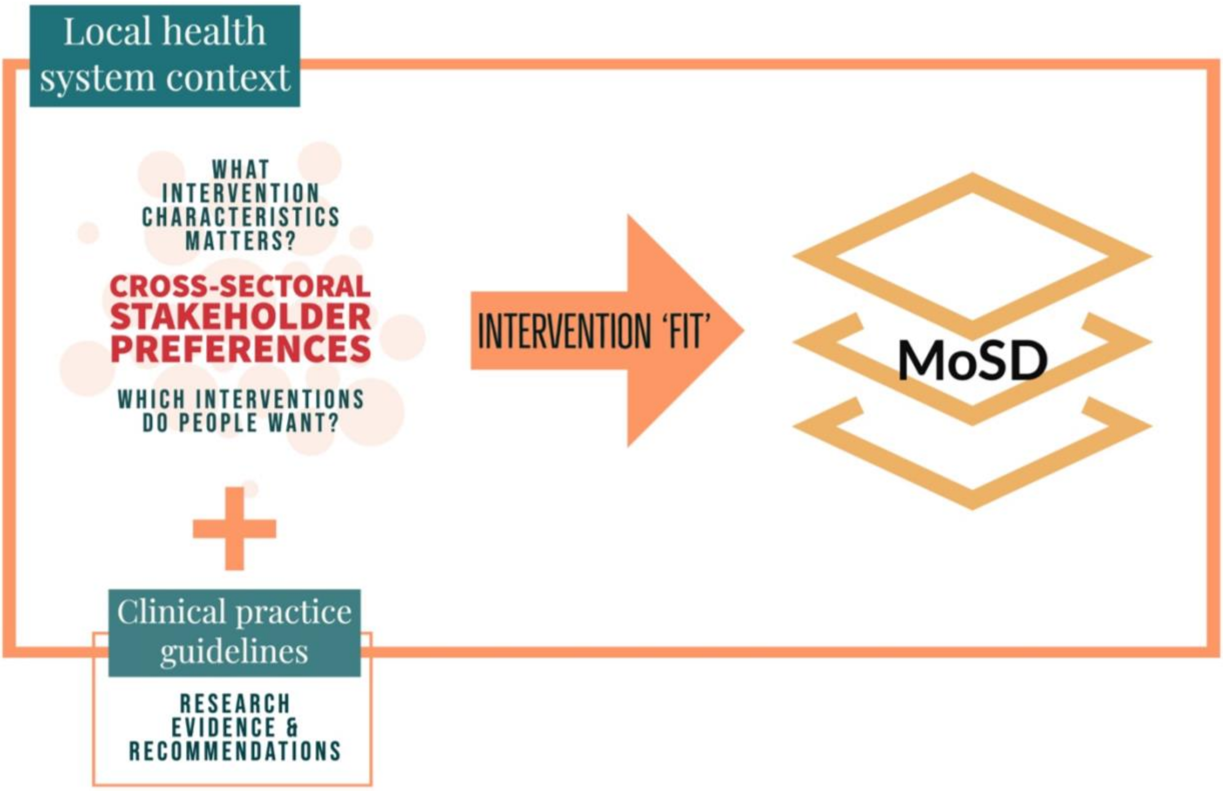


Figure 11. Conceptual framework of the connection between clinical practice guidelines (CPGs) and the preferences of cross-sectoral stakeholders in a local health system context. The combination of CPG evidence and recommendations with cross-sectoral preferences can inform the 'fit' of interventions in a local health system, and potentially guide the co-design of a value-based Model of Service Delivery (MoSD).

In this thesis, MCDA involves making trade-offs between criteria that matter to stakeholders in a health system when choosing OA interventions – i.e. the characteristics of interventions. However, what matters to cross-sectoral stakeholders when choosing or recommending OA interventions in the New Zealand health system has not been addressed in the literature. It follows then, that the first objective of this thesis is to identify what characteristics of interventions matter to cross-sectoral stakeholders when choosing or recommending OA interventions in New Zealand, and rank their relative importance. The remainder of this thesis will describe how this gap in the literature is answered empirically using an overarching MCDA approach.

CHAPTER 3: INTRODUCTION TO THE DESIGN AND METHODS (PART A)

3.1 CHAPTER OVERVIEW

The purpose of this chapter is to:

1. Provide a description of the overarching methodological framework used in this thesis
2. Describe the key methods used to answer Objective 1: What characteristics of interventions matter when choosing or recommending OA interventions?

The methods used to answer Objective 2, 'Which interventions for managing OA do people want in a health system?' are discussed in Chapter 5: METHODS (PART B).

3.2 METHODOLOGICAL FRAMEWORK

Multi-criteria decision analysis (MCDA) is the overarching methodology used in this thesis to evaluate and prioritise knee OA interventions. MCDA structures decision-making problems so that a decision-maker's trade-offs between alternatives (i.e. interventions) is explicit, systematic and transparent. This is particularly important when decisions are complex involving many stakeholders, comprise multiple and often conflicting objectives (i.e. criteria) and demand transparency – as is commonly encountered in healthcare practice and policy. Belton and Stewart (2002; p.2)[294] describe MCDA as "*an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter*."

In recent times, MCDA has become increasingly widespread in health research and practice and policy,[295, 296] supporting decisions for investment, resource allocation and prescription.[268] Importantly, MCDA can optimise decision-making by formally incorporating subjective data with objective data in a systematic and transparent process that identifies and weighs multiple evaluation criteria to solve decision problems.[35] This methodology is therefore ideal for combining the preferences of stakeholders for OA interventions with recommendations and evidence about OA interventions in a health system.

Commonly used terms for applying MCDA are described in Table 4. At the heart of MCDA is the premise that a set of alternatives can be ranked with respect to their performance on a number of criteria, through the production of an overall numerical score (i.e. indirectly). In this thesis, this broadly involves: modelling decision-makers' preferences for criteria that represent the characteristics of alternatives (e.g. intervention cost, therapeutic benefit and safety), rating the performance of alternatives on the criteria, and then combining the preferences and ratings to create a total score for each alternative (through which they are ranked).

A trade-off represents how much a decision-maker is willing to give up value in one criterion to achieve more on other criteria; by virtue, this means that a gain in the value of one criterion must be compensated by a loss in the value of other criteria. Value judgements are required to make trade-offs: they are in the eye of the beholder and represent their preferences.

For example, choosing which breakfast to eat at a café might involve ranking three alternatives: a fruit salad, cinnamon scroll, or 'full breakfast'. Fundamental to the decision objective – that is to rank the breakfast options from 'best' to 'worst' in order to select the option with best 'fit' – is the need to make trade-offs between the alternatives. Intuitively, the alternative which aligns the greatest with a decision-maker's preference is the 'best' alternative, and the others, 'worse'. However, this begs the question: what drives a decision-maker's trade-off; or more explicitly, which criteria, and what are their preferences for the criteria?

Table 4. Common elements of multi-criteria decision analysis

MCDA element	Description
Alternatives	The different options under consideration – e.g., interventions, investments, diseases or patients
Choice-based survey	A survey involving trade-offs between hypothetical alternatives that are characterised by and differ in terms of their performance on criteria, which change with each successive survey question. Trade-offs may involve comparing two hypothetical interventions at-a-time (i.e., pairwise), or more. Choice-based surveys ultimately quantify stakeholders' preferences for criteria as 'weights'
Criteria	The objectives or factors relevant to the overarching decision, which characterise real or hypothetical alternatives, in order to rank or select alternatives. E.g., the characteristics of OA interventions such as therapeutic benefit, safety and cost
Levels	Within-criterion performance categories typically informed by their 'best' to 'worst' plausible range. E.g., high, medium and low. Can be quantitative or descriptive.
Weights	The relative importance of criteria, representing decision-makers' preferences (or values). Weights are derived from choice-based surveys to inform preferences within- and between criteria
Decision-makers	The stakeholders relevant to the decision problem such as OA consumers, healthcare providers, health policy-makers, health advocacy representatives and OA experts
Trade-offs	How much a decision-maker is willing to give up value in one criterion to achieve more on other criteria; by virtue, this means that a gain in the value of one criterion must be compensated by a loss in the value of other criteria. These choices are driven by their preferences, or value judgements
Total score	Represents the overall value of an alternative, which, in this thesis is the sum of weights associated with the performance ratings for a given alternative on the same criteria

Following this example, consider that a decision-maker's choices are driven by three criteria, each described by two levels of 'performance': (i) fat content (high/low), (ii) protein content (high/low) and (iii) carbohydrate content (high/low). The decision-maker's preferences for the criteria are:

- low fat content (relative to high fat)
- low protein content (relative to high protein) and
- high carbohydrate content (relative to low carbohydrate).

Seen in Table 5, the performance of the breakfast alternatives are rated on the criteria (e.g., fruit salad is rated low on fat and protein, but high on carbohydrate content). Based on the decision-maker's preferences for the criteria, it is clear that 'fruit salad' is the 'best' overall option (ranked 1st) because of the alignment between its performance ratings and the decision-maker's preferences for the criteria, i.e., it is low fat, low protein and high carbohydrate content, relative to the other alternatives – cinnamon scroll and full breakfast – which are characterised by worse performance on the fat and protein criteria.

Table 5. Example 'performance matrix' showing the performance of three hypothetical breakfast alternatives on the fat, protein and carbohydrate criteria

Breakfast alternative	**Criteria ratings**			**Rank**
	Fat	Protein	Carbohydrate	
Fruit salad	low	low	high	1st
Cinnamon scroll	high	low	high	2nd
Full breakfast	high	high	high	3rd

In essence, MCDA codifies the above example by structuring the decision objective into parts: alternatives, criteria, weights (representing decision-makers preferences for the criteria levels), alternative performance ratings (as seen in the performance matrix). These components are quantified, so that numerical scores are calculated for each alternative, allowing them to be ranked by importance.

Of course, this example can be complicated by introducing more criteria and levels, and unequal preferences within- and between- the criteria. For example, in the example above the decision-maker's preferences for low fat, low protein and high carbohydrate are treated equally. However, some decision-makers might place twice as much weight (or importance) on low fat compared to the low protein and low carbohydrate (or some other combination of performance levels). In practice, these weight differences can be estimated using choice-based surveys (e.g., conjoint analysis, discrete choice experiment), which typically involve making trade-offs between the criteria (the choice-based method used in this thesis is described in Chapter 5, section 5.3.1).

The process of using MCDA to structure decision-making can be described using the six 'good-practice' six steps shown in Table 6.[266, 290, 297] The first chapter established that prioritising OA interventions based on the preferences of stakeholders in a health system was the overarching decision objective (Step 1). The remainder of this chapter will focus on specifying the criteria (Step 2; the other steps are discussed in CHAPTER 5:). However, first the philosophical standpoint of the author will be briefly explained, as this has implications on the methods chosen for scientific enquiry and the interpretation of research findings.

Table 6. Steps in the MCDA process[266, 290, 297]

Step	Description
1. Structure the decision problem	What is the objective of the decision? What are the alternatives? Who are the relevant decision-makers? What is the output?
2. Specify the criteria and their levels	a. Identify which criteria are relevant to the decision-makers b. Structure the criteria into mutually exclusive and exhaustive levels (i.e., categories) for differentiating between alternatives in terms of their characterisation on each criterion
3. Weight criteria and their levels	Determine weights for the criteria and levels, representing their relative importance to decision-makers
4. Measure and rate alternatives' performance	Measure alternatives' performance on the criteria Rate alternatives on the criteria, according to their level of performance
5. Apply scores and weights to rank alternatives	Calculate total scores for the alternatives by summing the weights corresponding to the alternative's ratings on the criteria
6. Support decision-making	Use the MCDA results, and sensitivity analysis to support decision-making

3.2.1 Philosophical perspective

> Genuine intellectual integrity is found in experimental knowing. Until this lesson is fully learned, it is not safe to dissociate knowledge from experiment, nor experiment from experience. John Dewey

Pragmatism emerged in the late 1800s by philosophers Charles Pierce, William James, and John Dewey. Arising from the Greek word “action,” pragmatism focusses on achieving the practical and achievable over the theoretical or ideal. Pragmatism is the belief that knowledge is only meaningful when coupled with action. Therefore, nothing is true or false and we believe in truths, but only if they work (e.g. the will to believe and the question of God).[298]

Rather than focusing on truth, pragmatism focusses on the outcomes of action;[299] thus, pragmatists believe in multiple realities and reject the objective or dualist perspectives (objective vs subjective). In other words, pragmatists have no problem with believing the existence of one ‘real world’ and also multiple individual realities. To addresses the issues of incommensurability with these two world views, pragmatists introduce the concept of intersubjectivity as a key element of social life. This translates to knowledge creation through action that people or groups achieve together, rather than the individual experience.[300] Epistemologically, pragmatists believe that knowledge is both context-specific and constructed through the reality we live in: experience.

This thesis is underpinned by pragmatism, which leaves the researcher methodologically open to using whatever methodology is appropriate to generate knowledge. The ability for the researcher to mix and match research methods is a fundamental principle of mixed-methods research,[301] which is discussed in the next section.

3.2.2 Methodological perspective

There are three predominant types of methods used to inform methodological approaches: quantitative, qualitative and mixed methodologies (also known as mixed-methods).[302] All three methods are used in this thesis and are described in the following sections.

3.2.2.1 Quantitative and qualitative methods

A key distinguishing factor between qualitative and quantitative methods is that the former seeks to explore phenomena, whilst the latter aims to confirm hypotheses about phenomena.[303] The analytical objectives of qualitative methods describe variation, whilst quantitative methods quantify variation.[303] For example, questions relating to the effectiveness, benefits or harms of a health intervention can be addressed using quantitative methods, such as randomised controlled trials. From a theoretical stand-point, quantitative methodology stems from positivistic ontology relying on the assumptions of realism, objectivism and determinism, which is the expectation of mechanistic laws and variables (i.e. a phenomenon is explained by empirical observation, through which it can be predicted and controlled). In contrast, qualitative methodology is useful for exploring individuals' experience of a health service or treatment. This is because rather than organising beliefs, experiences and perspectives of individuals into predetermined categories (for example, level of satisfaction or effect size), qualitative methods seek to develop new frameworks or theories by systematically interrogating qualitative data, making qualitative methods useful for establishing and clarifying unknown research variables.[302] Because of this, qualitative research is suited to providing culturally specific information about the values, opinions, relationships, behaviours and/or the experience of particular populations. Qualitative data are typically textual (e.g. audio transcriptions, field notes, or video tapes), rather than numerical (e.g. pain rating scale), which is characteristic of the quantitative approach. Focus groups are a prominent qualitative method used in health research.[304]

3.2.2.2Mixed-methods research

Johnson et al. (2007)[305] describes mixed-methods research as an approach that combines qualitative and quantitative methods to achieve breadth and depth of understanding and corroboration. Aligning with pragmatism, mixed-methods researchers are therefore open to many different types of qualitative and quantitative methods of enquiry, allowing the researcher to overcome the problem of the incompatibility thesis (the proposition whereby qualitative and quantitative methodologies are incompatible or cannot be mixed in the same research).[306]

Different types of mixed-methods design are possible and various taxonomies have been described to differentiate the classifications. For example, Creswell and Plano Clark (2011),[307] Morse and Niehause (2009)[308] and Johnson and Christensen (2014)[306] have developed various mixed-methods typologies. The Johnson and Christensen (2014; p. 658)[306] typology builds on the work by Morse and Niehause (2009).[308] Their framework features two dimensions: (1) time orientation, whether the research is conducted sequentially or concurrently and (2) paradigm/research-approach emphasis, which refers to whether the qualitative and quantitative component of the research design have equal weight or emphasis in answering the research question and interpreting the results, or whether one component clearly outweighs the other.

Using Morse and Niehause's typology, this thesis follows a 3 stage sequential equal-status mixed method design (QUAL → QUAN) → QUAN. The sequential design is characterised by a qualitative research component followed by a quantitative component which focusses on testing or generalising the initial qualitative results.[307] The typology indicates that the each stage of the study was conducted sequentially (as indicated by the arrows; the first stage in parentheses), and with equal weighting (as indicated by capitalised letters; also known as "interactive mixed methods research"; p.113[309]). The overall research framework and methodologies are summarised in Table 7.

Table 7. Outline of the philosophical and methodological framework used in this thesis

Research paradigm	Pragmatism		
Research Framework	Multi-criteria decision analysis (MCDA)		
Stage	Objective 1: What characteristics of interventions matters to stakeholders' when choosing or recommending OA interventions?	Objective 2: Which interventions for managing OA do people want in a health system?	
Research design	(QUAL → QUAN) →	QUAN	
Methods	1. Focus groups 2. eDelphi	1. PAPRIKA†	2. Additive value model

† The PAPRIKA method is explained in Chapter 5 (methods Part B)

Schoonenboom and Johnson (2017)[309] assert that interactive mixed methods research is characterised by equal-status mixed methods research design whereby mixed methods research is integrated at the levels of methods, methodology and paradigm. Furthermore, the authors continue that equal-status mixed methods research designs are achievable when qualitative and quantitative approaches are of equal value, feature in the research in alternation, are equally considered, and have outcomes that are integrated during and at the end of the research process. This design is suited to the pragmatist position and mixed or multidisciplinary researcher teams striving to address one superordinate goal.[309]

3.2.3 Summary

- In this thesis, qualitative and quantitative methods are used within a MCDA framework to explore and quantify stakeholder values and preferences.
- This process is conducted sequentially in two stages and focussed on one superordinate goal; prioritising knee OA interventions.
- This two stage approach is needed to identify which criteria are relevant to choosing or recommending knee OA interventions specifically for the NZ context (Objective 1); these criteria also form the backbone for answering which interventions for managing knee OA stakeholders want in a health system (Objective 2).
- The research framework is suited to mixed-methods approaches and the pragmatism research paradigm.

3.3 SPECIFYING CRITERIA FOR AN MCDA FRAMEWORK

The background chapter (Chapter 2) established that implementation of OA interventions is influenced by the preferences of cross-sectoral stakeholders in OA health care: people with lived experience, health equity advocates (in Aotearoa New Zealand, specifically Māori health advocates), healthcare providers, policy-makers and OA experts. Identifying the right criteria that these stakeholders consider in their choice of OA interventions is essential and is a key step in developing a MCDA model.[266] Inappropriate or missing criteria may invalidate the results of the MCDA process by triggering response bias if, for example, important criteria are missing from the MCDA model and are considered implicitly alongside the other decision-making criteria.[37, 310] Therefore, it is essential to ask: *what criteria do stakeholders consider when choosing or recommending OA interventions*?

The first step in the MCDA process is to define the overarching objective in order to completely and unambiguously determine the decision objective.[35] In this thesis prioritising knee OA *interventions* is the overarching decision objective. Therefore, the criteria will be informed by the *characteristics of OA interventions*, rather than, for example, the characteristics of individuals choosing OA interventions.

Identifying criteria can be achieved using several methods. Coast et al. (2012)[311] note that literature reviews, theoretical arguments from the literature, existing health outcomes measures, professional recommendations, focus groups, interviews, patient surveys, expert review, statistical significance in randomised trials and even policy questions have been used to develop criteria. Of these different approaches, qualitative methods are highly recommended[311, 312] to guide subsequent steps of the MCDA process, and understand how criteria are evaluated and interpreted by decision-makers. Therefore, in this thesis qualitative methods will be used to guide the identification and selection of criteria, and subsequent design of a survey instrument designed to capture the criteria weights. To optimise the efficiency and validity of the data collected with respect to our objective, we must first consider the sampling approach.

3.3.1 Sampling and Recruitment

Convenience, purposive and snowball sampling were used in this thesis. The premise of sampling is that it should enhance efficiency and validity. In qualitative research purposive sampling is widely used.[313] The aim of this type of sampling is to identify and select in-depth sources (i.e. participants) to gain deep insight into an issue. In contrast, random (or probabilistic) sampling aims to ensure generalisability by minimising the potential for bias by controlling for known and unknown confounders.[313]

When participants in the same group share similar characteristics such as age, gender or background, and do not know each other, this is known as homogenous sampling.[314] It is believed that this approach facilitates group discussion by encouraging participants to be more honest and open to a wider range of responses, and preventing set behaviours established through previous relationships or leadership power dynamics.[315] It is also often used to select focus group participants in order to describe a particular subgroup in depth, reduce variation and simplify analysis.[316] Maximum variation sampling (or maximum heterogeneity sampling), another type of purposive sampling, was also used in this thesis to encourage variation of responses within groups.[315]

Using both homogenous and heterogeneous sampling can optimise the depth and breadth of data collected across the multi-level stakeholders involved in OA care (i.e. consumers, healthcare providers, health equity advocates, policy-makers and OA experts).

In Objective 1 of this thesis (Chapter 4), convenience and snowball sampling (discussed below) was used to generate a list of potential focus group participants. From this list, homogenous groups of consumers, healthcare providers, health equity advocates, policy-makers and OA experts were created. Within each group, participants were then purposively sampled to maximise group heterogeneity with respect to their demographic characteristics such as their age and years experience in their primary role. In Objective 2 of this thesis (Chapter 6), convenience and snowball sampling was used to identify and invite participants to take part in the survey (discussed in Section 6.3.2).

Bias can be introduced to a study through poor response and follow-up. To enhance study responses, three strategies are available. First, snowball sampling can be used to identify individuals of interest from participants already deemed appropriate for a study (who therefore have similar characteristics). Second, convenience sampling can be used to seek eligible participants who were already easily accessible to the researcher.[313] Lastly, community-based advertisements and flyers can also be used target stakeholders for recruitment. Poor follow-up can be mitigated by incentivising study participation.

A notable limitation of these sampling methods, compared to, for example randomised sampling, is that the sample is subject to risk of sampling bias. For example, recruiting consumers via newspaper advertising may only access a subset of people who read newspapers. Similarly, social media campaigns (e.g. Facebook, Twitter) may not access people who do not use social media or have access to the internet. Sampling bias has implications for the generalisability of the study findings because the information generated from the study may not be representative of the population of interest.

3.3.2 Focus groups

Focus groups can be described as a method to gain in-depth information from a group of individuals whom are purposefully selected to give in-depth information about a particular topic that is otherwise not readily accessible through surveys or other quantitative methods.[173, 317] It is this group interaction that is thought to facilitate deeper and richer understanding of phenomena than what could be gained from, for example, one-on-one interviews.[315] Cleary and Horsfall (2014[318]) consider focus groups as the appropriate method for exploring clinical and professional issues; however, where personal or sensitive issues are explored, or participants are unable or unwilling to attend focus groups, one-on-one interviews are preferred. Between 6-12 participants are recommended for focus groups.[319] Too few participants may limit contrasting views whilst too large may not encourage participants to share their opinions and may consume so much time that it becomes unethical, or no longer cost-effective, as focus group interviews typically last approximately 1-2 hours duration.[315]

Limitations of focus group interviews include the inability to generalise from small groups of purposefully sampled participants, the potential for dominant participants to dominate and influence less outspoken participants, biased results caused by facilitators influencing participant responses, and the development of group norms that may limit open discussion of the topic.[319] These limitations can be minimised by adopting a structured focus group and interview approach to create a ‘safe environment’ conducive to open discussion.

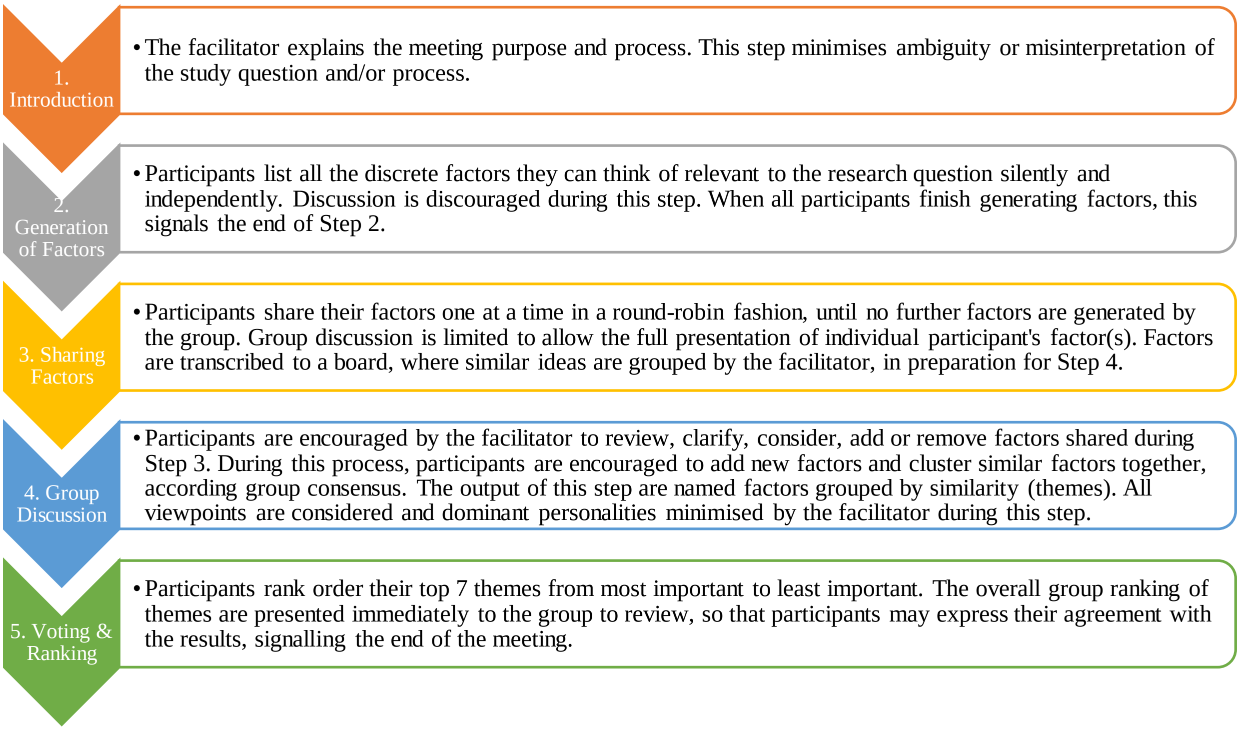
3.3.2.1Nominal Group Technique

The Nominal Group Technique (NGT) allows for focus group discussion with a clear pathway to identifying issues, establishing consensus or priorities.[320] However, it can also be used to inform other methods. For example, NGT has been used to establish national health priorities in Australia,[321] develop clinical standards in South Africa,[322] and inform criteria selection for modelling discrete choice experiments.[323] It is considered by some as a popular approach for developing consensus in health care research.[324]

Considered a mixed-method approach,[325] the benefits of the NGT are that it is time and cost efficient, allows for immediate feedback of results to the group, and creates an environment that gives equal representation to all group members, effectively controlling the potential for unhelpful group dynamics such as overpowering participants or less outspoken participants.[320] The NGT is highly structured, which allows all participants opportunity to have their voices and opinion considered by other participants. This is essential as failing to control for these considerations can inhibit the full potential of focus groups to gain rich and meaningful insight into phenomena.[303] However, it may also be a disadvantage, as it inhibits the organic discussion that would normally flow from typical focus group discussion.

In this thesis, the NGT was deemed appropriate due to its focus on generating discrete ideas or factors, which are akin to criteria used in MCDA, and previous use informing criteria selection in a similar research application.[323]

Potter et al. (2004)[325] describe 5 key steps to performing the NGT, shown in Figure 12. In brief, the NGT is a staged approach that goes from individual idea generation and sharing about a topic to group discussion, consolidation of ideas into higher-level themes, and aggregation of individually importance rated themes into group-level importance ratings. The key outputs of the NGT method are field notes, audio recordings of focus group discussions, ideas, themes (thematically grouped ideas) and ranked themes. To harness knowledge from more than one focus group, the data must be aggregated to aid analysis and interpretation.



1. Introduction
- The facilitator explains the meeting purpose and process. This step minimises ambiguity or misinterpretation of the study question and/or process.

2. Generation of Factors
- Participants list all the discrete factors they can think of relevant to the research question silently and independently. Discussion is discouraged during this step. When all participants finish generating factors, this signals the end of Step 2.

3. Sharing Factors
- Participants share their factors one at a time in a round-robin fashion, until no further factors are generated by the group. Group discussion is limited to allow the full presentation of individual participant's factor(s). Factors are transcribed to a board, where similar ideas are grouped by the facilitator, in preparation for Step 4.

4. Group Discussion
- Participants are encouraged by the facilitator to review, clarify, consider, add or remove factors shared during Step 3. During this process, participants are encouraged to add new factors and cluster similar factors together, according group consensus. The output of this step are named factors grouped by similarity (themes). All viewpoints are considered and dominant personalities minimised by the facilitator during this step.

5. Voting & Ranking
- Participants rank order their top 7 themes from most important to least important. The overall group ranking of themes are presented immediately to the group to review, so that participants may express their agreement with the results, signalling the end of the meeting.

Figure 12. Steps of the Nominal Group Technique from Potter et al. (2004)[325]

3.3.2.2Limitations of the NGT

Combining data from more than one NGT focus group is problematic. For example, can the relative importance of themes be compared across groups if there is variation in group size and demographics, diversity of themes and number of themes generated by the focus group discussion? How is consensus defined? The literature is scant with clear recommendations about how to manage this problem, although McMillan et al. (2014)[326] suggest that the method proposed by van Breda (2005)[327] is appropriate, if analysis of combined data is the objective.

The van Breda method[327] describes seven steps for aggregating NGT data from multiple groups using qualitative and quantitative methods. In essence, van Breda recommends an inductive thematic analysis of the NGT-generated themes into an overarching framework of categories and sub-categories representing all the NGT-generated themes, followed by quantitatively calculating the aggregate relative importance of the ranked NGT-generated themes. However, this method still does not resolve the issues described in the previous paragraph. Additionally, the thematic analysis of NGT-generated themes cannot be verified by the focus group participants.

In order to overcome these limitations of the NGT, a mixed-methods approach utilising the Delphi survey method is used in this thesis to verify the results of the thematic analysis and rank the relative importance of the categories and/or subcategories by including all of the focus group participants as one group. Qualitative analysis of focus group data is discussed in the following section after which the Delphi method will be explained in Section 3.3.4.

3.3.3 Thematic analysis

Analysis of focus group data should be systematic, sequential, verifiable and continuous to reduce bias.[314] Thematic analysis is a method of systematically analysing and reducing data for identification of themes in rich detail.[328] A well-known method described by Braun and Clarke (2006)[328] involves a six-step approach to thematic analysis which can be applied to focus group data: (1) close-reading of transcripts, (2) generation of code segments from the raw data, (3) development of key categories, (4) development of sub-categories within key categories, (5) defining and naming categories, (6) reporting of exemplar quotes and scholarly reporting. There are other methods available, such as the general inductive approach described by Thomas.[329] Importantly, thematic analysis results in a framework of categories and sub-categories summarising in rich detail the key ideas or themes emerging from the raw data. Another aspect of the approach is to consider how the raw data are interpreted to generate meaning.

3.3.3.1Abductive approach

The abductive approach to qualitative analysis described by Morgan (2007)[300] was adopted in this thesis. Abduction (also known as theoretical redescription[330]) can be defined as a process of "*inference or thought operation, implying that a particular phenomenon or event is interpreted from a set of general ideas or concepts*" (p.205).[330] In other words, abduction argues that research cannot operate in a purely theory- or data-driven process. It accepts that the research analyst must constantly move between these two contrasts (i.e. deductive and inductive) to generate meaning from data. Pragmatism is particularly suited to this analytical approach. Importantly, abduction raises the level of theoretical engagement beyond rich descriptions of empirically-derived themes by acknowledging that the chosen theoretical concepts used by researchers to re-describe initial codes into higher-level themes or concepts may be incorrect.[331]

3.3.3.2 Semantic versus latent level analysis

In thematic analysis a decision needs to be made about how themes are generated from the interpretation of the data. Using a semantic approach themes are created from the surface meaning of the data (i.e. explicitly), such that the research analyst does not look beyond what was said or written in the raw data to create themes.[332] A characteristic of semantic analysis is that after describing and summarising the themes, the resultant themes are ideally interpreted for significance and meaning through intersections with the existing literature. In contrast, the latent (or interpretive) analysis approach involves interpreting the underlying meaning (ideas, assumptions, conceptualisations and ideologies) in the data, that inform the semantic meaning of the data. This approach is therefore suitable for theorising the underlying reasons underpinning what is explicitly captured in the data,[332] which was not the aim of the research question for this thesis. For these reasons, a semantic approach was adopted in this thesis on the basis that MCDA requires that stakeholders' make direct, unambiguous value-judgements about the criteria, rather than making assumptions about their underlying meaning or inferences about other criteria at the same time.

In Chapter 4, NGT data will be thematically synthesised, using a semantic-level abductive approach. The application of these methods is described in detail in Section 4.5.2.

3.3.4 Delphi method

Two commonly used applications of the Delphi method are framework development and forecasting and issue identification or prioritising.[333] In this thesis the Delphi method is used to expand the NGT by: (1) verifying the thematic analysis of aggregated NGT-data and (2) rank the importance of the characteristics of OA interventions emerging from the thematic analysis.

The Delphi method is highly utilised in health research[324, 334] and is considered more rigorous than the NGT for gaining consensus across a large numbers of participants, making it particularly suited to guideline development.[335] The key premise of the Delphi method is that group opinion is greater than individual opinion. It can be described as a series of surveys interspersed with controlled feedback that seeks to seeks to obtain the most reliable consensus of feedback from a group of experts.[336] von der Gracht (2012)[337] identified four fundamental characteristics of the Delphi method.

The first characteristic, anonymity, refers to the participants remaining blinded to each other because questionnaires are completed individually and coordinated centrally by the facilitator. Hence, respondents in Delphi surveys are called ‘panellists’ or grouped into ‘panels’. This approach removes any effects group dynamics may have on influencing participant opinion (e.g. dominant individuals), which is a notable limitation of focus groups.

The second characteristic of the Delphi method is that it is conducted in a series of survey rounds. In a classical Delphi method the first round comprises open ended questions which are then qualitatively analysed by the facilitator to feed back into the subsequent rounds. Content analysis is typically used to summarise qualitative data.[338] In subsequent rounds, the research analyst can remove unnecessary information to allow respondents to focus on items which may have not reached consensus and adjust prior judgements based on group information, including content analysed summaries of free-text comments left by respondents. Iteration helps to converge opinion towards consensus, although this is not always required. In order to maintain rigor, a response rate of at least 70% between survey rounds is recommended.[339]

Controlled feedback refers to the deliberate selection of information shared with the participants from one round to the next by the facilitator. It is considered independent of variation (in response to the survey question(s)) because the facilitator decides on the type of feedback.

Finally, 'group response' refers to the use of aggregated descriptive or quantitative statistics for communicating the level of agreement achieved by the group at the end of each round. Level of agreement may be questioned, for example, on the resulting level of agreement or importance of the themes and/or subcategories arising from the analysis of qualitative data from the previous round. The psychometric Likert scale is commonly used in Delphi and is the most widely used scale in survey research in all disciplines.[338] For example, a typical 5-point Likert scale ranges from 1 through 5 in order of very important, important, neither important or important, not important and unimportant. A number of other scales are also used, depending on the construct being evaluated. For example, the RAND/UCLA method recommends a nine-point numeric rating scale.[340]

3.3.4.1Defining consensus

Defining consensus is subjective which leads to many different approaches and poor reporting.[334] Consensus can be defined in Delphi studies using qualitative analysis and descriptive measures. For example, stopping criteria can be determined, based on a certain number of rounds; certain level of agreement defined by central tendency (mean, median and mode), dispersion (standard deviation and interquartile range) or mean/median ratings and rankings; or even according to stability of responses between rounds.[341] The systematic review by Diamond et al. (2014)[334] found that using percent agreement with the same rating is the most common approach and 75% agreement was the median threshold used to define consensus among the 42 studies they reviewed.

To achieve the first objective of this thesis, consensus was defined as ≥80% complete or partial agreement with sub-categories of the thematic framework derived from the analysis of the aggregated (raw) NGT-data using a 3-point Likert scale (completely agree, partially agree, disagree) and ≥80% strongly agree or agree for the overall thematic framework (5-point Likert: strongly agree, agree, neither agree or disagree, disagree, strongly disagree). This conservative threshold for consensus was informed by the desire to capture only the essential characteristics of interventions in anticipation of a large number being generated, and aligned with the same threshold used in an Australian CPG for hip and knee OA.[13]

To generate the relative importance of the characteristics of interventions identified in the overall thematic framework, the mean rank of the characteristics of interventions was calculated from most- to least-important, within and across the groups. The strength of agreement in rankings across and within groups was assessed using Kendall's *W* coefficient of concordance. Kendall's *W* is a non-parametric test, an appropriate test for ordinal data, which produces a result between zero and one. Coefficients of 0.1, 0.3, 0.5, 0.7 are very weak, weak, moderate and strong levels of agreement respectively.[342]

In studies which seek to establish consensus about a particular topic or issue, it is imperative that Delphi studies are robustly conducted and reported, since the credibility of the resulting recommendations are dependent on the rigorous use of the Delphi technique. Junger et al. (2017)[343] provide guidance on Conducting and rEporting DElphi Studies (CREDES), derived from a methodological systematic review of 35 studies, which will be followed in this thesis.

3.3.5 Rigour in mixed-method studies

Rigor is the process by which we demonstrate integrity and competence as a way of establishing the legitimacy of the research process.[344] Mixed-methods research is intended to combine the strengths of individual qualitative and quantitative approaches whilst minimising the weaknesses of either approaches in order to achieve greater insight. Therefore the quality of the design of mixed-methods studies should be greater than the parts to achieve what is called "*multiple validities legitimation*" by Onwuegbuzie and Johnson (p.59).[345] There is no consensus about what are the essential criteria to meet multiple validities legitimation, however, Curry & Marcella Nunez-Smith[346] identify six core domains of quality: (1) justification for mixed methods, (2) design quality, (3) adherence to respective standards for qualitative and quantitative research, (4) adherence to standards for data analysis in mixed methods, (5) quality of analytic integration, and (6) quality of interpretation and inference. Essentially, these criteria draw on the concepts of rigor in qualitative and quantitative methodologies, which, in turn, contribute to the quality of mixed methods study.

3.3.6 Rigor in quantitative studies

Quantitative studies focus on replication, prediction, and causal relationships between variables. Curry & Marcella Nunez-Smith[346] describe four hallmarks of appraising the quality of quantitative enquiry. Firstly, internal validity, which refers to the degree to which the results accurately and precisely represent the phenomenon under study. This concerns whether or not the study measured what it was intended to measure, and if sources of bias and confounding were addressed within the study. Internal validity can be enhanced by randomising study conditions, identifying and controlling for extraneous or confounding variables, comparing control versus intervention groups, and developing instruments through systematic processes such as cognitive interviews and factor analysis.[346]

Second, reliability refers to consistency, stability, and repeatability of observations or measures thin a study. Reliability can be enhanced by multiple measures of the same construct, cognitive testing and piloting of survey instruments, training of data collectors to ensure high inter-rater reliability, data cleaning, and using statistical procedures to adjust for measurement error.[346]

Third, generalisability (also known as external validity) concerns the degree to which the findings of a study can be transferred to the same or another population: it is a critical requirement for knowledge to progress forward.[346] Generalisability of findings can be enhanced by using random selection, specifying inclusion and exclusion criteria, using of validated instruments, assessing potential for non-respondent bias, and providing descriptions of statistical procedures including treatment of missing data and confidence intervals.[346]

Fifth, objectivity refers to the degree in which researchers' bias can remain separated from the research process. This is usually controlled by randomising selection, applying explicit protocols, and performing statistical computations.[346]

Detailed reporting is a prerequisite to demonstrating rigor in these aspects of scientific enquiry. Guidelines exist for different appraising the quality of evidence such as the CONSORT statement[347] for randomised controlled trials and AGREE checklist for clinical practice guidelines.[228]

3.3.7 Rigour in qualitative studies

In qualitative studies validity, reliability and generalisability are referred to as 'trustworthiness'. Lincoln and Guba [348] developed the four quality criteria to assess trustworthiness: credibility, transferability, auditability and confirmability.

Credibility is comparable with internal validity, i.e. establishing "truth", and so it establishes the fit between respondents' views versus the researchers' interpretation of them is credible.[344]

Credibility can be established in various ways including using a variety of methods to gather data, prolonged engagement and from a range of participants, member checking, audit trails and triangulation. Triangulation, in particular, uses multiple sources and perspectives to reduce systematic bias. Various types of triangulation include data, investigator, theory, method and analysis triangulation.[349]

Transferability (comparable to external validity) refers to the generalisability of enquiry.[344] Adequate reporting of sampling and setting addresses this criterion.

Dependability (comparable to reliability) is achieved through an audit trail where others can examine the process and key decisions undertaken to produce data and results of enquiry. Reflexivity is an important aspect of dependability, whereby the research keeps a self-critical account of the research process. Auditing is also used to authenticate confirmability.[344]

Confirmability (comparable with objectivity or neutrality) concerns the strategies used to establish that data and interpretations of the findings are clearly derived from the data and not the researchers imagination.[344] Strategies for limiting bias include the researcher being reflective and keeping a journal , peer review such as asking a colleague to audit critical decisions or interpretation of data; asking participants to verify the reasonableness of interpreted data and having a team of researchers.

In the following chapter, a mixed-methods study employing focus group discussion and Delphi survey will be used to elicit from participants what factors they consider when choosing or recommending hip or knee OA interventions. Multiple validities legitimisation will be reported against the Consolidated Criteria for Reporting Qualitative research (COREQ-32)[350] and Conducting and Reporting Delphi Studies (CREDES)[343] guidelines to report the focus group and Delphi study to best practice recommendations, respectively. Briefly, the trustworthiness of the study will be established through member checking, triangulation (investigator and theory), and an external and independent Delphi panel to achieve the qualities of credibility, transferability, confirmability and dependability.

3.4 CHAPTER SUMMARY

This chapter established that pragmatism underpins this research and that it follows an overarching MCDA methodology to achieve the thesis objectives. A mixed-methods research approach, utilising focus group discussion using an NGT approach and the Delphi survey method was chosen to answer what criteria related to the characteristics of interventions cross-sectoral stakeholders consider when choosing or recommending OA interventions, based on the qualities of these research methods. This was followed by a description of how rigor is demonstrated in qualitative, quantitative and mixed-methods research.

CHAPTER 4: CHOOSING INTERVENTIONS FOR HIP OR KNEE OSTEOARTHRITIS – WHAT MATTERS TO STAKEHOLDERS? A MIXED METHODS STUDY

The original research in this chapter is the application of methods described in the previous chapter and has been published in Osteoarthritis and Cartilage Open[41] (included as SUPPLEMENT 1 with permission from Elsevier for thesis work). As the primary author for this article I am responsible the study design, acquisition of data, analysis, interpretation, manuscript drafting, response to reviewers and approved the submitted version of the manuscript. Supervisor and co-author contributions are acknowledged in the initial study conceptualisation, planning and overall editorial aspects of manuscript preparation prior to the final article submission. Elsevier's copyright policy allows authors reproduce their own articles for theses.

4.1 ABSTRACT

OBJECTIVE: To identify what factors drive choices among interventions for hip or knee osteoarthritis (OA), and to rank the characteristics of interventions.

METHODS: In this phased, mixed-methods design involving cross-sectoral and multi-disciplinary stakeholders – healthcare consumers, providers, policy-makers, Māori health advocates and OA experts – we used the Nominal Group Technique in focus groups to generate data. We conducted thematic analysis of the focus group data to inform a framework of categories and sub-categories describing factors and characteristics influencing the choice of OA interventions. We then used a dual-panel, two-round e- Delphi survey to verify the framework and rank the characteristics of interventions.

RESULTS: From six focus groups (n=38 participants), 364 factors were identified and clustered into 56 themes (mean 9 themes per focus group; range 5-15). Thematic analysis revealed a framework of 3 core categories: characteristics of interventions (10 sub-categories), characteristics of consumers (10 sub-categories) and characteristics of the health system (7 sub-categories). In Delphi round 1, the framework was verified by each of two panels (n=65, ≥80% acceptability). In round 2, two characteristics of interventions were combined, resulting in 9 characteristics (in decreasing order of importance): effectiveness, appropriateness, quality of the evidence, accessibility, harm, cost, duration, passivity, and immediacy of intervention effect.

CONCLUSION: Stakeholders make choices among interventions for hip or knee OA within a framework of characteristics of interventions, of consumers, and of the health system. We identified and ranked 9 key characteristics of interventions that stakeholders consider when choosing or recommending interventions for hip or knee OA.

4.2 SIGNIFICANCE AND INNOVATIONS

- Core interventions for managing hip and knee osteoarthritis (OA) can reduce the individual health and socioeconomic burden of OA, but are underutilised. Successful implementation of interventions for managing OA is influenced by stakeholders' preferences for the characteristics of interventions.
- This mixed-methods study empirically derived a framework describing the factors influencing stakeholders choice of OA interventions, and then, using the Delphi method, verified and drew consensus on nine essential characteristics of interventions considered by cross-sectoral and multi-disciplinary stakeholders in the choice of OA interventions. In decreasing order of importance, they are: effectiveness, appropriateness, quality of the evidence, accessibility, harm, cost, duration, passivity and immediacy of the intervention effect.
- Decision-makers and developers of cross-sectoral strategies for OA, such as Models of Care, should be aware of these characteristics of interventions so that intervention choice can be more closely aligned with stakeholders' preferences.

4.3 INTRODUCTION

Clinical practice guidelines (CPGs) for the management of hip and knee osteoarthritis (OA) consistently recommend core management strategies of exercise, weight loss, education and self-management.[14, 202] Yet the care received by people with OA is often at odds with these core strategies.[19, 236] Successful implementation of interventions and health system strategies for OA are influenced by numerous factors, often involving multiple stakeholders.[27] For example external context (e.g. health policy and infrastructure); organizational culture (e.g. culture and leadership); professional attributes (e.g. beliefs and attitudes to change);[27] and consumers' and providers' attitudes about OA and OA interventions.[25, 208]

Fundamental to an intervention's successful implementation is its 'fit' within a local health system (context), which depends on the compatibility of the characteristics of interventions with stakeholders' preferences for those characteristics.[27] The preferences and expectations[351] of health care consumers and providers about characteristics such as treatment effectiveness, cost and accessibility may impact the intervention's overall suitability within a particular context. For these reasons, characteristics of interventions must be considered when developing or implementing local health system strategies, such as Models of Care (MoC).[352] Preferences for these characteristics are often studied at one level of the health system and are less focussed. For example 'healthcare decision-making' among clinical decision-makers and policy-makers,[353] or separately among patients.[354] An important knowledge gap remains concerning the characteristics multi-level stakeholders consider, and their preferences for these characteristics.

The aims of this study were to: 1) identify the factors considered by relevant stakeholders, across a range of health care settings, when choosing or recommending OA interventions and express these in a conceptual framework; and 2) define and rank the main characteristics of interventions according to the relative importance placed on them by stakeholders.

4.4 METHODS

4.4.1 Design

In this two-phase, mixed-methods study (Figure 13), phase 1 consisted of focus group discussions conducted using the Nominal Group Technique (NGT).[325] The NGT data were pooled and thematically analysed to identify priority themes.[325, 326] Phase 2 was a dual-panel, two-round, eDelphi study. Round 1 aimed to verify the framework and the focus of round 2 was to rank the importance of the characteristics of interventions. Our design included these validation processes: (i) independent parallel coding of NGT data with reduction into a composite framework to minimise researcher bias (ii) a dual-panel eDelphi which included an external and independent panel to verify the content validity of the Framework.

The study was conducted in New Zealand (NZ) between November 2016 and August 2017. The methods and results for phase 1 are presented according to COREQ-32[350] criteria for reporting qualitative studies (SUPPLEMENT 2). Study phase 2 follows guidance on reporting and conducting Delphi studies (CREDES)[343] and the recommendations of Diamond and colleagues (2014).[334] Ethical approval was obtained from the University of Otago Human Ethics Committee (D16-329).

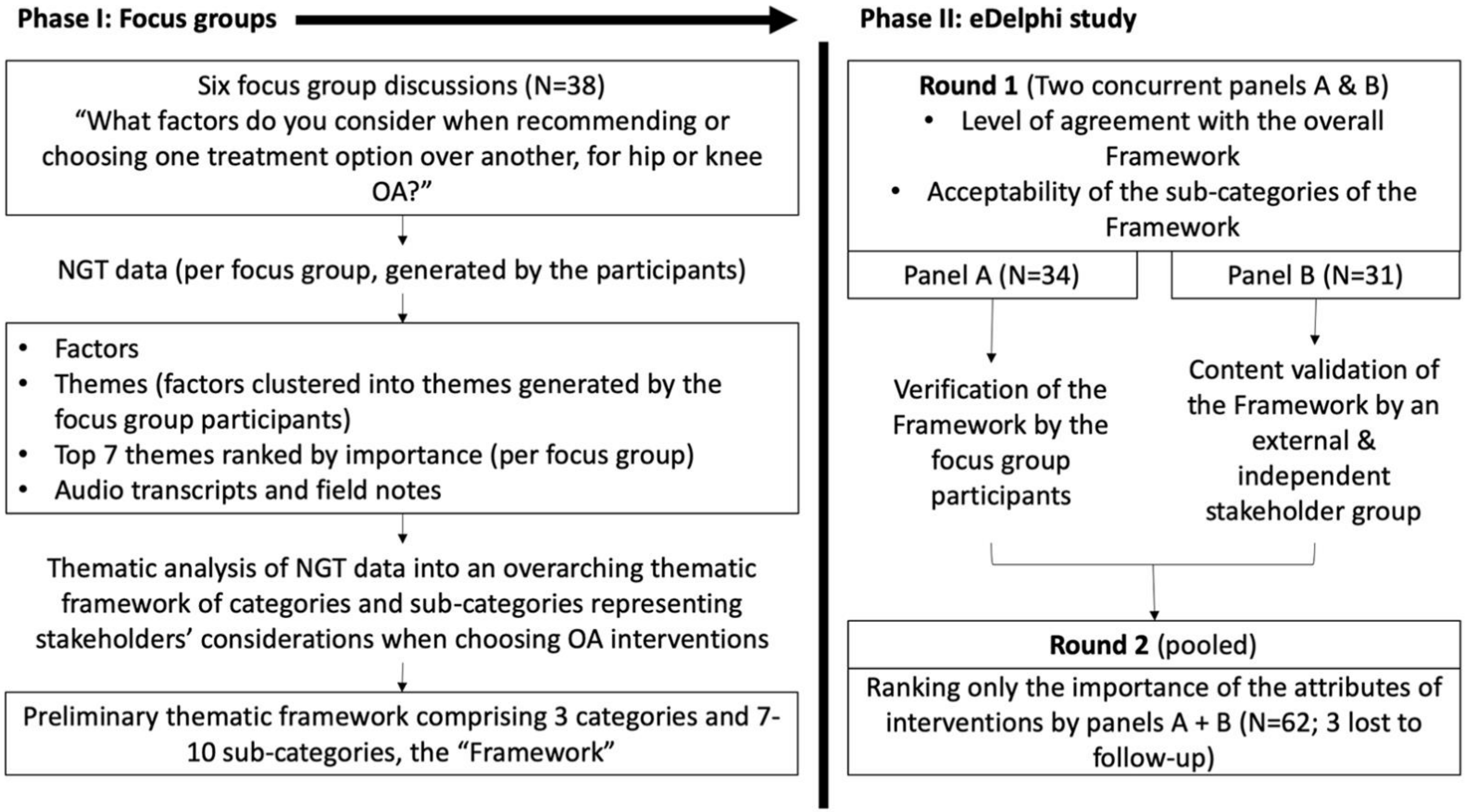


Figure 13. Summary of the study design

4.4.2 Sampling And Recruitment

4.4.2.1Focus groups

We sampled 5 stakeholder groups: health care consumers, providers, policy-makers, Māori health advocates and OA experts. Convenience and snowball sampling generated lists of stakeholders, who were screened against inclusion criteria adapted from a previous qualitative study of cross-sectoral experts in musculoskeletal MoC[355] (Table 8). Purposive sampling of eligible stakeholders was used to maximise the diversity of participant perspectives, to reduce the risk of limited perspectives being represented within stakeholder groups. The Māori health advocates group were included as Māori experience overall disproportionately worse health outcomes relative to the non-Māori population in Aotearoa/NZ.[150] For each of the above stakeholder groups, focus groups with 3-11 participants were formed.[356]

Stakeholders from professional, government and non-government organisations and health care provider agencies were invited to participate by telephone and email. Consumers were targeted using local newspaper advertisements and recruitment flyers at local community centres. 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/). OA experts were initially identified using an online database (http://expertscape.com/) and reviewed by two authors (JC and JHA) with regards to their eligibility. Because of the small population of OA experts in NZ, Australian OA experts were also included on the basis that beliefs about the management of OA should not differ between the two countries given similarities of the health systems in these neighbouring countries and consistency of international clinical practice guidelines.[14, 202]

Table 8. Inclusion criteria for the focus group discussions (phase 1)

Stakeholders	Health policy/strategy, advocate or health service/programme delivery in NZ	Osteoarthritis expert in clinical or health services research in NZ or Australia	Clinical practitioner	Consumer [d]	Māori health advocate
Inclusion criteria	Experience in health service or programme delivery, coordination, management or funding related to musculoskeletal health care and/or chronic disease (that implicitly includes OA) for at least 1 year at a senior level in NZ [c] Experience at a senior [a] level in musculoskeletal and/or chronic disease (that implicitly includes OA) health policy or advocacy for at least 1 year in NZ At least one participant reflecting each of the sub-categories below	Awareness of guidelines for OA At least 5 years experience in undertaking clinical and/or health services research in arthritis or health care-related implementation science at a senior level [b] with evidence of peer-reviewed publication(s) in the area At least one participant reflecting each of the sub-categories below	Experience in delivery of clinical care for people with arthritic conditions at a senior practitioner level [b] for at least 5 years in NZ Awareness of guidelines for OA At least one participant reflecting each of the sub-categories below	Meets the NICE clinical criteria for diagnosis of OA (i.e.: 45 years or over; has activity-related joint pain, and either; has no morning joint related stiffness or morning stiffness lasting less than 30 minutes) At least one participant reflecting each of the sub-categories below	At least one participant reflecting each of the sub-categories below
Sub-categories	Community services health policy or strategy decision-maker Public health services health policy or strategy decision-maker (District	Model of care or clinical practice guideline development expert Clinical epidemiologist in OA	General Practice Physiotherapy Rheumatology Orthopaedic surgery Community pharmacy	At least 40% male At least 40% female National-level advocacy for arthritis health care	At least one male At least one female Māori service provider

	Health Boards and Primary Health Organisations) Assessment, treatment and rehabilitation services health policy or strategy decision-maker National-level advocacy for arthritis health care Health service management or coordination at secondary hospital setting Health service management or coordination at primary care setting	Academia and education in OA

---

[a] At least senior officer or manager level of employment; [b] Fellowship level for medical practitioners (e.g. FRNZCGP); senior clinician level for other disciplines; associate professor level for researchers; [c] At least at the manager or head of department level; [d] Self-reported diagnosis of OA or determined by the National Institute for Health and Clinical Excellence guideline recommendation for diagnosis of OA.[56]

4.4.2.2 eDelphi survey

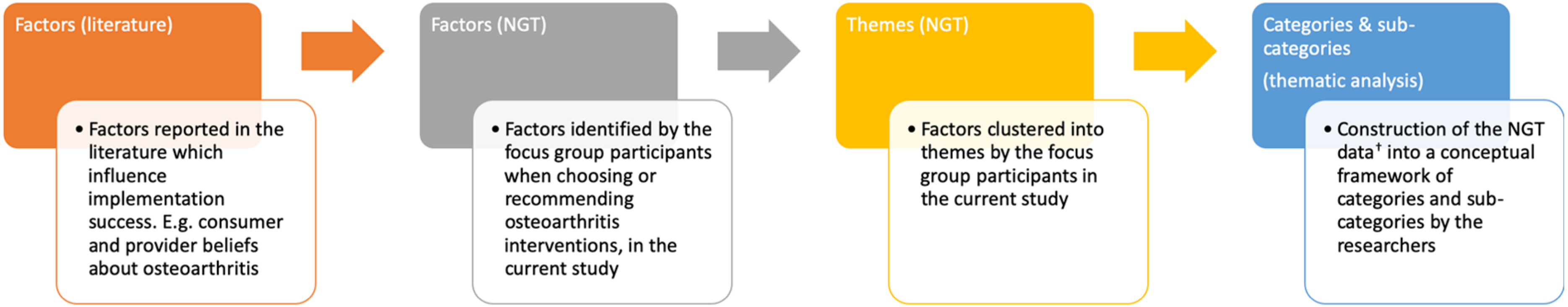
The dual-panel eDelphi survey was conducted using Qualtrics (Provo, USA). Panel A included phase 1 participants. Panel B, convened to concurrently assess the content validity of the Framework, comprised participants satisfying the same inclusion criteria but who were unable to take part in the focus groups.

4.5 PROTOCOLS & DATA ANALYSIS

4.5.1 Focus group protocol (phase 1)

In phase 1, focus groups were facilitated using the mixed-method NGT. The NGT structures group discussion in response to a question that can then be prioritised through group consensus in real-time. Focus groups generated data comprising: factors (representing the factors affecting stakeholders' choice of OA interventions), themes (factors grouped into themes, by participants), top 7 themes ranked by importance (by the participants), audio transcripts and field notes (hereinafter "NGT data"; see Figure 13). From these data, thematic analysis resulted in a framework (hereinafter "the Framework") of categories and sub-categories (term descriptors are seen in Figure 14).

Face-to-face and virtual (digital) focus groups were conducted for each stakeholder group separately using the NGT.[325] Each focus group was audio-recorded and transcribed verbatim for analysis by JC. The question posed to participants to discuss was deliberately broad to capture all possible characteristics of interventions: "What factors do you consider when recommending or choosing one treatment option over another, for hip or knee OA?" Consumers were posed the question in the context of their own experience: "What factors do you use to choose one treatment option over another for your OA?"



Factors (literature)

- Factors reported in the literature which influence implementation success. E.g. consumer and provider beliefs about osteoarthritis

Factors (NGT)

- Factors identified by the focus group participants when choosing or recommending osteoarthritis interventions, in the current study

Themes (NGT)

- Factors clustered into themes by the focus group participants in the current study

Categories & sub-categories

(thematic analysis)

- Construction of the NGT data† into a conceptual framework of categories and sub-categories by the researchers

NGT= Nominal Group Technique
†factors, themes, themes ranked by importance, audio transcripts and field notes

Figure 14. Description of terms used in this paper for factors, themes, categories and sub-categories

The first author (JC) was trained in the NGT by a researcher (CC), experienced in the method; CC facilitated the first focus group, assisted by JC, who facilitated all subsequent sessions. Two hours were allocated for each group, structured using five standard NGT steps:[325] (1) introduction and explanation (~5 min), (2) individual and silent generation of factors (15 min), (3) sharing factors identified in step 2, individually with no group discussion (20 min), (4) group discussion and participant-led clustering of factors into themes (40 min), and (5) voting and ranking of themes (20 min). These 5 steps are detailed in SUPPLEMENT 3.

In anticipation of the focus groups generating potentially many themes, participants were asked to identify and rank their top 7 themes – on the basis that this number of characteristics is commonly used in health valuation studies employing multi-criteria decision analysis (MCDA).[268] In the case of OA experts, participating in a virtual environment using a modified NGT, steps 1 and 2 were performed online using Qualtrics. Responses were collated and clustered offline by JC. For step 4, the preliminary clustered ideas were then presented to the OA experts, using Zoom videoconferencing (San Jose, USA) and RealtimeBoard (www.realtimeboard.com).

4.5.2 Qualitative data analysis

For the thematic analyses we used the abductive approach[300] which builds on the general inductive approach described by Thomas.[329] From this perspective we accepted *a priori* that the research team's primary frame of reference (evidence-based practice and evidence-based policy-making)[172, 357] would influence the thematic analyses.

First, the NGT data generated from the focus groups were organised using the software package Atlas.ti (Ver.7, Berlin, 2015).

Second, JC performed a thematic analysis of the data following the steps described by Thomas (2006),[329] and guided by the advice of McMillan and colleagues (2014) to aggregate diverse and multiple-group NGT data.[326] The iterative process of open-coding raw data began with reading all results and transcripts from the consumer and health care provider focus groups, verbatim. Preliminary codes emerging from the text resulted in the creation of categories and sub-categories (hereinafter, “themes” and “factors” refer to patterns identified in the raw data; “categories” and “sub-categories” refer to the analysts’ constructions of the data). The remaining focus group transcripts were coded using this abductively-derived coding framework, until a broad thematic framework of participants’ considerations emerged. Next, the analysis focussed on identifying specific NGT data that were related to the overarching aim of the study. These data were open coded by JC until a primary thematic framework of stakeholders’ considerations emerged.

Thirdly, a thematic framework representing participants’ choice-making with respect to OA interventions was developed. Using the same analytical approach undertaken by JC, two researchers experienced in qualitative analysis (AMB, JHA) independently coded and analysed the transcribed NGT results and illustrative quotes to develop two independent frameworks. JC then mapped these two frameworks onto the primary framework to identify areas of concordance and discordance. In two meetings (arbitrated by PH) these 3 analysts reviewed the three frameworks and the results of the mapping exercise to agree on a consensus framework (hereinafter referred to as the Framework) consisting of categories and sub-categories.

Consistent with the a priori research question, i.e. to focus on the characteristics of interventions, the sub-categories describing the characteristics of interventions were further refined by consensus (JC, JHA, AMB) into discrete sub-categories that were generalisable, complete and non-redundant – desirable characteristics of criteria in MCDA.[290, 358]

4.5.3 eDelphi protocol (phase 2)

Phase 2 was a two-round, dual-panel online Delphi (eDelphi) survey to verify the Framework and to rank the sub-categories describing characteristics of interventions (Figure 13). The protocol was based on a previously published process.[352]

A two-round eDelphi was planned a priori, regardless of the overall level of agreement with the Framework after round 1. Two panels were planned for verification the Framework and validation of the results. Participants anonymously completed the survey and were eligible for a NZ$100 gift voucher upon completion of both rounds. Email reminders were sent after each round to increase the response rate.

The aims of round 1 were to: i) verify that no characteristics of interventions important to stakeholders were missing from the analysis (Panel A) and; ii) establish the content validity of the Framework (Panel B). Both panels provided ratings of agreement for the overall Framework (assessed using a 5-point Likert scale: strongly agree, agree, neutral, disagree, strongly disagree) and acceptability of the derived sub-categories (assessed using a 3-point Likert scale: completely agree, partially agree, disagree). Free text fields allowed participants to comment on any component of the Framework, and participants were prompted to explain their answer on any disagreed (disagree/strongly disagree) component. Participants were also asked to add any essential characteristics of interventions missing from the Framework. Demographic data were captured in round 1. Round 2 focussed on ranking the importance of the characteristics of interventions. Panels A and B were pooled for this quantitative analysis.

4.5.4 eDelphi survey data analysis

An a priori consensus for round 1 of the eDelphi was defined as ≥80% complete or partial agreement for the sub-categories (3-point Likert scale: completely agree, partially agree, disagree), and ≥80% agree or strongly agree for the overall framework (5-point Likert scale: strongly agree, agree, neither agree or disagree, disagree, strongly disagree). The qualitative aspect of round 1 comprised content-analysis of participants' free-text comments to identify new categories or sub-categories emerging from either panel. We performed further analysis of free-text comments across all the sub-categories rated as 'disagree' from >20% of respondents (per group) to clarify the descriptions of the characteristics of interventions carried into round 2.

The aim of round 2 was to rank the characteristics of interventions by importance. Consistent with the *a priori* research question, from this point on the eDelphi survey was limited to the characteristics of interventions; other categories and sub-categories of the Framework were not evaluated further. We calculated the average rank of the characteristics of interventions, from most- to least-important, within stakeholder groups and for the overall sample. Level of agreement was assessed using Kendall's W within and across the groups, where 0.1, 0.3, 0.5, 0.7 are very weak, weak, moderate and strong levels of agreement respectively.[342] To ensure the final set of characteristics was representative of all groups (which differed in size), the top two ranked characteristics for each group were included in the final aggregated list of characteristics, regardless of their overall ranking, providing that adequate within-group agreement was demonstrated (Kendall's W >0.1, p<0.05). Data were analysed using IBM SPSS Statistics (ver.24, Armonk, NY).

4.6 RESULTS

4.6.1 Focus group discussions

Six focus group meetings were held (N=38, 61% female; n=3-11 participants per group, median 6 [IQR 3.3]), each lasting approximately 90 minutes (range 60-120 min). Roles, experience and number participating in each stakeholder group are summarised in Table 9. Steps 2 and 3 of the

NGT produced 364 factors. After the removal of duplicates, 258 unique factors emerged across focus group discussions (mean 43 per group; median 39; range 29-62). Step 4 produced a median of 9 (range 5-15) participant-clustered themes per group (total 56 overlapping themes). Step 5 produced themes ranked by importance which broadly overlapped across groups concerning consumer-related factors (e.g. clinical status and consumer preferences), themes related to efficacy and safety of intervention and health system factors (e.g. intervention accessibility and culturally appropriate practices). The OA expert group, conducted by the modified NGT that did not include step 5, produced themes echoed in the other groups.

The results of the NGT exercise (the NGT data) are reported in Supplement 3, Table S1. Initial independent thematic analysis by the researchers (JC, AMB, JHA) produced three frameworks, each with 3-10 categories, comprising 5-17 sub-categories. Two rounds of discussion leading to consensus resulted in a thematic framework of three over-arching categories: 1) characteristics of consumers (10 sub-categories), 2) characteristics of interventions (10 sub-categories), and 3) characteristics of health systems (7 sub-categories) (Figure 15). A summary description of the Framework is presented in Table 10 (sub-category descriptions and supporting illustrative quotes are in Supplement 3, Table S2 & Table S3).

Table 9. Demographic characteristics of the focus group (phase 1) and eDelphi survey (phase 2) participants at rounds 1 and 2

	Focus groups (Phase 1) (N=38)			eDelphi Round 1 (Phase 2)						eDelphi Round 2		
				Panel A[a]			Panel B (N=31)			Pooled[b]		
Stakeholder group[c]	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)
Consumers	**11 (29), 69 ± 10 (50-88)**	**5 (13)**	**16 ± 16 (1-50)**	**9 (26), 69 ± 10 (50-88)**	**5 (15)**	**14 ± 13 (1-35)**	**7 (23), 67 ± 9 (50-75)**	**5 (16)**	**10 ± 8 (1-25)**	**15 (24), 68 ± 10 (50-88)**	**9 (15)**	**13 ± 11 (1-35)**
Māori health advocates[d]	**5 (13), 37 ± 12 (25-57)**	**4 (11)**	**5 ± 4.4 (2-10)**	**5 (15), 37 ± 12 (25-57)**	**4 (12)**	**5 ± 4 (2-10)**	**-**	**-**	**-**	**5 (8), 37 ± 12 (25-57)**	**4 (6)**	**5 ± 4 (2-10)**
Health care providers[e]	**10 (26), 49 ± 13 (33-67)**	**5 (13)**	**22 ± 9 (11-30)**	**9 (26), 49 ± 13 (33-67)**	**4 (12)**	**21 ± 9 (11-30)**	**14 (45), 49 ± 10 (30-66)**	**5 (16)**	**20 ± 10 (8-37)**	**22 (36), 48 ± 11 (30-67)**	**9 (15)**	**20 ± 9 (8-35)**
General practitioner	4 (40), 55 ± 11 (41-64)	2 (20)	25 ± 7 (15-30)	3 (33), 56 ± 13 (41-64)	1 (11)	26 ± 8 (15-30)	4 (29), 47 ± 9 (39-59)	3 (21)	18 ± 8 (12-30)	8 (36), 53 ± 12 (39-67)	4 (18)	22 ± 9 (12-30)
Physiotherapist	5 (50), 46 ± 14 (33-67)	2 (20)	21 ± 11 (11-30)	5 (56), 46 ± 14 (33-67)	2 (22)	21 ± 11 (11-30)	4 (29), 44 ± 8 (30-56)	1 (7)	12 ± 4 (8-15)	8 (36), 40 ± 8 (30-55)	3 (14)	15 ± 7 (8-30)
Orthopaedic surgeon	-	-	-	-	-	-	3 (21), 56 ± 10 (46-66)	0 (0)	24 ± 12 (15-37)	2 (9), 51 ± 6 (46-55)	0 (0)	18 ± 4 (15-20)
Pharmacist	1(10), 39	1 (10)	15	1 (11), 39	1 (11)	15	-	-	-	1 (5), 39	1 (5)	15

Rheumatologist	-	-	-	-	-	-	3 (21), 57 ± 4 (52-59)	1 (7)	30 ± 5 (25-35)	3 (14), 57 ± 4 (52-59)	1 (5)	30 ± 5 (25-35)
Health policy-makers[f]	7 (18), 53 ± 10 (34-62)	5 (13)	6 ± 7 (1-20)	6 (18), 52 ± 11 (34-62)	5 (15)	8 ± 7 (1-20)	4 (13), 53 ± 7 (44-59)	4 (13)	14 ± 11 (6-30)[g]	9 (15), 53 ± 9 (34-62)	9 (15)	10 ± 10 (1-30)
Heath policy or programme developers	1(14), 62	1(14)	3 ± 0 (0)	-	-	-	-	-	-	-	-	-
Health system funders	2 (29), 60 ± 1 (59-60)	1 (14)	20 ± 0 (0)	-	-	-	-	-	-	-	-	-
Health service delivery, coordinators or managers	4 (57), 48 ± 11 (34-59)	3 (57)	3 ± 3 (1-7)	-	-	-	-	-	-	-	-	-
Osteoarthritis experts	**5 (13), 50 ± 7 (41-60)**[h]	**4 (11)**	**12 ± 5 (8-15)**	**5 (13), 50 ± 7 (41-60)** [h]	**4 (12)**	**12 ± 3 (8-15)**	**6 (19), 52 ± 8 (40-60)**[h]	**3 (10)**	**16 ± 10 (5-34)**	**11 (17), 51 ± 7 (40-60)**	**7 (11)**	**14 ± 8 (5-34)**

[a] response rate of N=34 (89%); [b] response rate of N=62 (95%); [c] defined by primary role; [d] two stakeholders were interviewed face-to-face; [e] two focus group meetings were conducted for the health care provider group; [f] sub-group level data was not collected for eDelphi survey; [g] p= 0.036 (Mann-Whitney U); [h] n=4 Australians

Table 10. Summary descriptions of the framework of factors influencing stakeholders choice of OA intervention derived from the focus group discussions

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
Characteristics of the intervention Stakeholders gathered information about interventions to assess for suitability against a patient’s circumstances or feasibility within a health system. The characteristics or attributes of interventions, applicable to any number of interventions, were used to compare and contrast between interventions. Key characteristics of interventions included: evidence for effectiveness and safety, consideration of the benefit and trade-off, cost-effectiveness (to the consumer), immediacy of treatment and access to treatment. Information about the intervention options were also gathered from health professionals, peer/family experience and advertising sources (e.g. newspaper and online advertising) in the decision to select a treatment option for consumers.	• Access to treatment for the consumer. The suitability of access considered in terms of an intervention’s distance to nearest provider and wait time. • Active versus passive intervention options. The extent to which an intervention allows a patient to self-manage their condition over the long term, versus passive care. • Cost of the intervention. The financial costs of intervention relevant to the use or provision of health care and society. • Duration of the intervention effect. The ability to delay or avoid more invasive interventions later through improvements in quality of life (e.g. physical function). • Evidence about the intervention appropriateness. Providing the right treatment or services for OA at the right time (e.g. surgical intervention for early- or advanced-stage OA). • Evidence about the intervention effect. Considered on the basis of evidence outside a clinical trial, including short-term and long-term evidence of effectiveness and change in health system and societal costs.	• “Availability; for example, I explored physiotherapy and exercise options but I live an hour or so out of town…to access services whether it be the public system, whether it be close to; in my case travel’s a factor.” Consumer #6 • Capacity and confidence and motivation might be different; I might put that as a different ‘cause somebody might have the capacity and confidence, but they’re not motivated to change … if you look at the goal-setting care plans, the way they’re done, those sorts of things [factors related to self-management] would definitely be taken into consideration.” Osteoarthritis expert #1 • “A whole lot around cost; what the particular cost is. The difference to the lifetime cost system and also the patient’s lifetime costs. The value for money. (Health Policy #5) • “I made a note about this just while conversation was going on earlier about surgery and last resorts because from my point of view I would much prefer surgery to a drug medication that goes on and on and on.” Consumer #7

Key Category and summative description	**Sub-categories**	**Illustrative quotes supporting each sub-category**
	• Evidence about the intervention harms. The risk of short- and long-term side effects of intervention. • Immediacy of the intervention effect. The time delay between starting and experiencing the benefits of intervention. • Quality of the evidence (+ *views and opinions*). The extent to which one could be confident that the effects of the treatment or service described were real. *Views and opinions* concern the value which different people assign to different sources of evidence. For example, peer or family advice and/or personal accounts were viewed as more trustworthy for some participants than health professionals' or advertisement (e.g. newspaper, Facebook) advice, claims or recommendations about the benefits or harms of intervention.	• "I guess in a sense we might say that all the evidence or whatever we might come up with on paper says that this is the right treatment, but actually for certain groups of people in society it might be the wrong treatment because it might be unfeasible because they're rural or can't afford to travel. So therefore it's the wrong treatment for that person or even that group at that time." Health Policy Maker #2 • "Well, efficacy, simply does the drug work; or does the treatment work? And with more complex interventions, say we've got multidisciplinary clinics, or whatever, then you'd need effectiveness which is showing that it works within a context." Osteoarthritis expert #2 • "It doesn't matter whether it's a drug that's actually prescribed by the doctor or this Arthrem or any of those other things; you are putting something into your body and things like that. Is it going to be worth it? Is it not? You're weighing this up all the time." Consumer #4 • "The problem is though you see on the advertisements and things like that and these people stand up and say, 'This is the best thing since sliced bread,' and you think 'Right, okay, I'm going to give this a go.' How much are they being paid to say that and in actual fact have they had that working for them; because you've got

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
		no way of checking. All it is, is you think “Well if it's worked for that person do I spend the money and give that a go,” but you have no knowledge of it." Consumer #9
Characteristics of the consumer When choosing or recommending an intervention for OA, stakeholders considered it important to match an intervention to the biopsychosocial profile of the patient. Key factors included an assessment of a patient’s severity of OA, clinical comorbidity profile, preferences (inclusive of attitudes, beliefs, ability to self-manage, motivation and treatment goals) and access to interventions. Social considerations included support to/from the family and/or community responsibilities.	• Clinical status of individual’s osteoarthritis. The recommendation or selection of intervention based on objective tests, patient reported symptoms and age. • Presence of other conditions. For example, diabetes, heart disease and mental health, and the medical care being received for these conditions which may influence intervention choice. • Patient attitudes and beliefs. Intervention options need to reflect a consumer’s (and their family’s) beliefs about OA, as well as beliefs or expectations about the intervention options. • Capacity to self-manage. Consumer’s health literacy, attitudes and beliefs about OA (e.g. perception of pain), and capacity to self-manage and navigate the health system. • Treatment goals. Treatment options need to relate to the immediate and longer-term functional goals of the consumer now and into the future. • Desperation (e.g. from pain). Once consumers felt that they had exhausted the therapeutic	• "The trouble is there’s the objective parts of what you find out about the patient and then there’s the subjective part that they contribute which they talk about. So, to me those are the two separations and I would see them as being quite important to me. So, your x-rays, your scans, your blood tests and all that, I see them as being objective versus what the patient says and how incapacitated they say they are, all that sort of stuff which a subjective thing is.” Health care provider #1 • "…a couple of things that always worry me are whether they’ve got a chronic pain syndrome and whether they’ve got a chemical addiction already when you’re considering what you’re gonna do. Those two things influence me quite a bit with what I’m going to suggest that they might contemplate doing." Health care provider #1 • "Desire. Like after all is said and done you can explain the benefits, present evidence based practice, and if they don’t want to take it they

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
	options immediately available to them to manage the symptoms of OA they sought to improve their quality of life by trying other interventions or medicines (e.g. herbal remedies) with poor evidence supporting their safety, benefit or interaction with other interventions. • Access to treatment. The consumer’s geographic location and the number of treatment options available (to choose). • Affordability for the consumer. The immediate out of pocket expenses to the consumer. • Social support factors. Consideration of consumers’ place in the community and wider community and support requirements if a disability is present. • Feasibility to the consumer. Whether or not a treatment option is acceptable and feasible to the consumer to adopt in the short and long term, based on the biopsychosocial circumstances of the consumer, their goals and expectations of treatment, and access to health services.	just don’t want to take it. They’ve got to want to; they’ve got to want the treatment. And someone might just want to die and not take the treatment; like I’m over this, I’m sick of being old. I have patients say that: ‘Don’t get old dear, don’t get old’…or you might be like the 20th health professional to contact them about this and they just stop…They’re just over it." Māori advocate #1 • “I think health literacy, education and level of patient knowledge all relate to what the patient needs to bring to the treatment or treatment decision.” Health policy #2 • "I put how I’m feeling; it's time of the day sometimes. Like how I’m feeling; if I’m in a good space or something like that and I’m going somewhere and it's painful or whatever I might decide to not do something. I might decide not to get sort of very heavy about it. Or otherwise I may be having a bad day and I weigh up the pros and cons that way. It’s how I’m feeling on the day." Consumer #3 • “Being open for information too rather than just closing your mind off. I think you reach the stage where you will listen to anybody in the hope that something may work.” Consumer #7 • “Yeah so it's not so much about equity; it's about would somebody have to travel a big distance

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
		and that would be difficult for them. It's where the costs fall for the treatment option and which of them are most accessible for that person." Health policy #2 • "The other thing too perhaps and I don't know quite how you put it in, but I'm still part time working and so my choices up to a point is how it can still allow me to work…Because if something's not going to allow me to continue what I'm doing then I wouldn't be interested in it?" Consumer #10 • "I would say family support to and of as well. Like for some Māori it depends on their beliefs and their upbringing but some people would choose a spokesperson; and so everything is dealt with that spokesperson that the patient has nominated as the spokesperson. So you wouldn't necessarily – and this is really weird as a health professional that you talk to someone other than the patient, 'cause it's really the patient that you're dealing with and their experience. But to some Māori they do choose a spokesperson and so everything gets [9.10]. So let's say it's an older person; they might choose their youngest son, or they might choose the youngest daughter and that's the person that you deal with, with all cares concerning or all treatments concerning the

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
		patient. So you actually don't really deal with patient; which is different." Māori advocate #1
Characteristics of the health system Participants considered whether there was the need for the society to invest in health care for OA and, whether the interventions considered were feasible to adopt or implement from the perspective of the health system and the consumer. Factors considered by stakeholders included the burden of disease profile for OA, socio-political interest, health system capacity, access and equity, culturally appropriate services and, the cost-effectiveness of intervention to society.	• Political landscape – is the disease a compelling problem? Before recommending interventions for OA, policy-makers considered if there was community and political interest in addressing the burden of OA. • Costs and benefits to society. Before recommending treatments for OA, policy-makers considered if there was community and political interest in addressing the burden of OA. • National health system capacity. Policy-makers also considered the capacity of a health system to support OA health care including policy direction, funding availability, evidence in support of funding or implementing a new service or intervention and workforce capacity and capability • Local health system capacity. Health policy-makers considered the feasibility for a local health system to provide care services for OA in terms of the costs, communication capacity among care providers, the current availability of services (public and private), access to services and scope for integration with other existing services	• "We've talked about the burden of disease but actually there's a far more less palatable political reality, which is about is there a burning platform? Is there some advocacy for this? Is this both accepted and publicly accepted as being needed? And it might be quite small. So I wonder if burden of disease is only one part of is there a burning platform or public support for this…Sometimes the burning platform is because of the cost to the system. Sometimes it's because of the consumer voice, from some effective advocates, and that's all captured in the how is this framed in the public's mind or little P politically. Because there's a political element to it which might just be local community; I don't mean national politics." Health Policy #2 • "A whole lot around cost; what the particular cost is. The difference to the lifetime cost system and also the patient's lifetime costs. The value for money. Health Policy #5 • "…I think the question is what resources are available? What funding is available? Because you're likely to cut your service, cut your coat according to your cloth. You're likely to come up

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
	• Access and equity. Consideration of the burden imposed on the consumer to reach an intervention or service for OA was reasonable (e.g. geographical remoteness, fit with lifestyle) and, whether or not it was affordable to those who need it (fairness/equity of access) • Culturally appropriate services. Whether or not a health care service for OA was culturally appropriate for diverse communities. • Characteristics of the health service provider. The beliefs, experiences and preferences of the health service provider.	with a service that is feasible within the funding available and other resources available." Health Policy #5 • "...it's actually I guess in a way saying more about the role of the District Health Board. [A DHB] plan for commissioned services in its own district, right? Once it might be a national policy or capability that we're going to provide these services in the country and then the DHBs end up providing those services in their own district according to what's available in their district to do that." Health Policy #2 • "I wonder if accessibility at a system level becomes kind of fairness or equity; that at the same time as you want to know can the system deliver this, can you set up treatment options that will carry through, there's something about access and about fairness of access that you might want to pick up as well. (health policy #1)…Yeah…I just was thinking it's more than access; it's about fairness of access which I call equity, but I mean we might frame it differently." Health policy #5 • "…educating the family is important also because that patient potentially is living with them, and so they will be offering the cares in some way, shape or form; so having the family, the patient and the whole team, the MDT, all on

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
		the same page. Like Māori don’t tend to – this could be generalising but in my experience Māori don’t put their loved ones into a rest home; so they’re at home living with the younger daughter or the niece or somebody. So having those people involved in all aspects is really helpful ‘cause that will help the patient overall.“ Māori Advocate #2

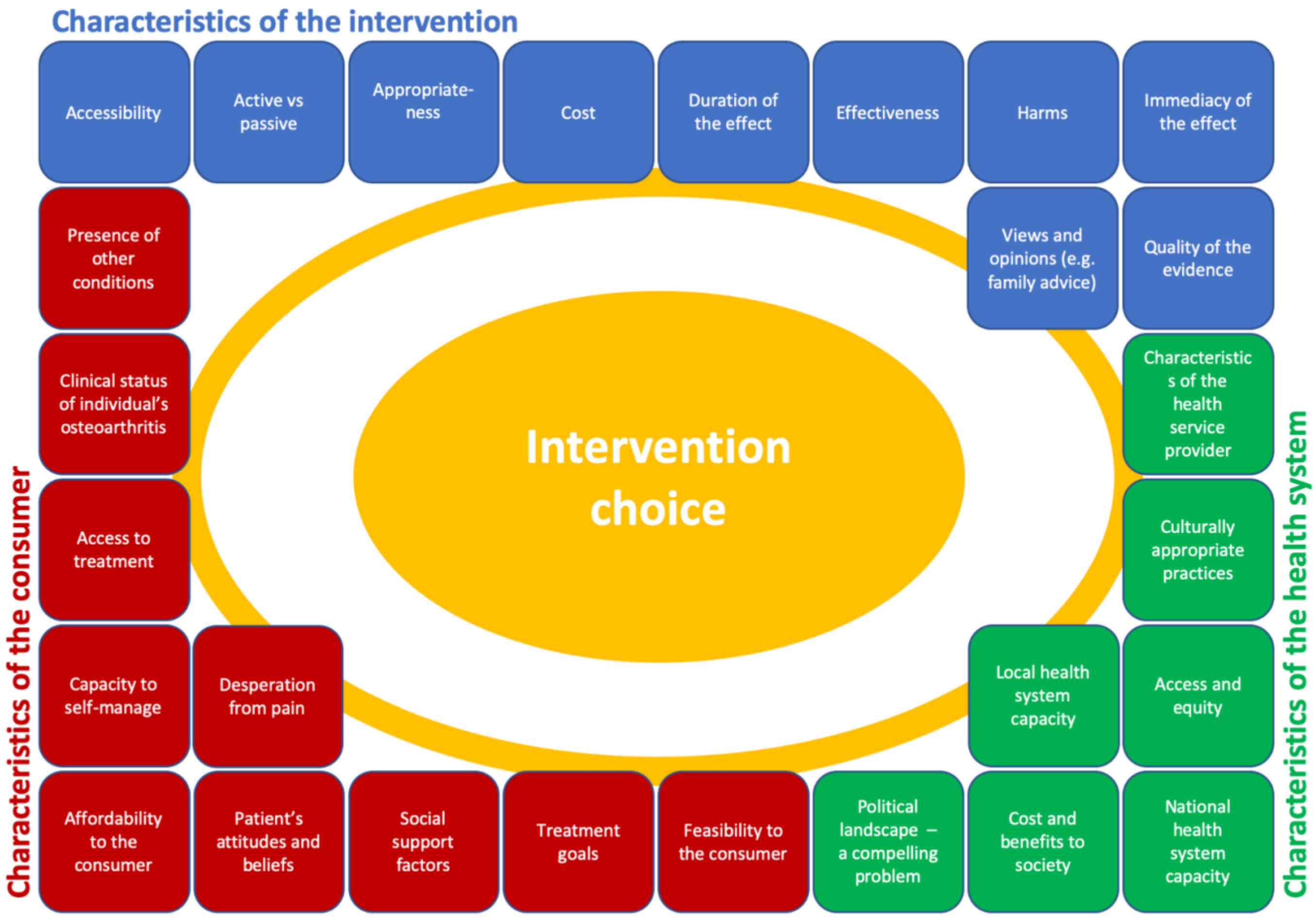


Figure 15. Thematic representation of the three over-arching categories and sub-categories, derived from the 56 themes generated by the focus discussion, that stakeholders consider in their choice of hip or knee osteoarthritis intervention

4.6.2 eDelphi survey

Delphi round 1 included 65 participants (4 phase 1 participants lost to follow-up in Panel A, N=34; Panel B, N=31). In round 2, 3 participants were lost to follow-up due to personal circumstances (95% retention rate). The panels' demographic characteristics are summarised in Table 9.

4.6.2.1eDelphi round 1 outcomes

The overall acceptability of all the categories of the Framework was confirmed by panels A (91% strongly agree or agree) and B (94% strongly agree or agree) and by the stakeholder groups (Supplement 4, Table S4 – Table S5). Analysis of pooled panel and stakeholder-group agreement scores for each sub-category revealed consensus across all sub-categories, except for the "desperation" sub-category (77% partial or complete agreement), confirming the content of 26 of the 27 sub-categories (Supplement 4, Table S6). Analysis of agreement within stakeholder groups identified lack of agreement for the sub-categories (Supplement 4, Table S7): [n (%) stakeholder group] 5 (45%) OA experts for "desperation", 3 (30%) policy-makers for "views relating to the characteristics of the intervention" and 5 (22%) providers for "local health system capacity". Analysis of stakeholder's free-text comments to the overall framework did not reveal any new categories or sub-categories. The three sub-categories with >20% disagreement are discussed here. Content analysis of free text comments from 4 of the 5 OA experts (1 did not leave a comment) who disagreed with the "desperation" sub-category posited that consumers may be desperate to try new treatments but did not consider it good practice; i.e. they agreed with the meaning of the sub-category per se. Likewise, the comments from three providers who disagreed with the "local health system capacity" sub-category (2 did not leave comments) concerned local health system barriers, rather than disagreement with the sub-category. One policy-maker commented about their disagreement for the "views relating to the characteristics of the intervention" sub-category, however it was unrelated.

In summary, the eDelphi panellists' agreement with the overall Framework and the sub-categories (except the "desperation" sub-category) indicate strong alignment between the stakeholder-generated synthesis and the qualitative analysis, with validation provided by an independent stakeholder panel. Within the "characteristics of interventions" category, consensus was not reached by the stakeholder groups about the "views relating to the characteristics of the interventions" sub-category. Two participants' comments guided the analysis team to conclude that these views related to the evidence concerning interventions – from a range of sources – therefore these two sub-categories could be merged within the "quality of the evidence" sub-category for round 2. Thus, apart from this one change, no other modifications were made to the Framework, confirming its completeness and content validity, and leaving nine discrete characteristics of interventions (sub-categories) to be ranked by participants.

4.6.2.2eDelphi round 2 outcomes

There was significant but weak agreement in the rankings across the groups (N=62; between-group W=0.333; p≤0.00). A weak-to-moderate level of agreement in rankings was observed within stakeholder groups except for the Māori health advocacy group (N=5), which did not demonstrate adequate within-group agreement (Kendall's W=0.089, 'very weak'; p=0.893). The top two ranked characteristics of interventions across the resulting stakeholder groups were: evidence about the intervention effectiveness, evidence about the intervention appropriateness, quality of the evidence, and immediacy of treatment effect. The Māori advocacy group was the only group to rank immediacy of treatment effect among their top two characteristics; however we were unable to achieve significant agreement within this group. Our failure to observe reliable estimates in this stakeholder group was likely due to the low sample size, despite our extensive efforts, so data from the Māori health advocate group were pooled with those of the healthcare consumer group. The aggregate ranking of the characteristics of interventions, across all groups, was (in decreasing order of importance): effectiveness, appropriateness, quality of the evidence, accessibility, harm, cost, duration, passivity and immediacy of effect (Table 11).

Table 11. Characteristics of interventions after eDelphi round 2, in decreasing order of importance (top to bottom), for each stakeholder group. Data presented as the mean characteristic and rank by stakeholder group.

Stakeholder group (W; p-value)					
Consumers	**Māori health advocates**	**Health care providers**	**Health policy-makers**	**Osteoarthritis experts**	**Pooled**
Effect (2.53)	Immediacy (3.80)	Effect (2.24)	Quality (2.33)	Effect (1.83)	Effect (2.40)
Appropriate (3.87)	Effect (4.00)	Appropriate (2.33)	Effect (2.44)	Quality (3.08)	Appropriate (3.18)
Access (4.27)	Appropriate (4.20)	Quality (4.05)	Appropriate (3.22)	Appropriate (3.33)	Quality (4.00)
Duration (4.87)	Access (5.20)	Access (4.14)	Cost (5.44)	Harms (4.08)	Access (4.71)
Quality (5.13)	AvP (5.20)	Harms (5.19)	Access (5.78)	Access (5.25)	Harms (5.34)
Immediacy (5.73)	Cost (5.20)	Cost (5.71)	Harms (5.89)	Cost (5.58)	Cost (5.65)
Cost (5.87)	Duration (5.60)	Duration (6.29)	Duration (6.00)	Duration (6.33)	Duration (5.85)
Harms (5.93)	Quality (5.60)	AvP (6.95)	AvP (6.67)	AvP (7.17)	AvP (6.77)
AvP (6.80)	Harms (6.20)	Immediacy (8.10)	Immediacy (7.22)	Immediacy (8.33)	Immediacy (7.10)
(0.222; 0.001)	(0.089; 0.893)	(0.533; <0.00)	(0.452; <0.00)	(0.588; <0.00)	(0.333; <0.00)

The mean rank for the characteristic is in parentheses (lower numbers represent greater importance).
W= Kendall's W; AvP= Active vs Passive interventions; Effect= Evidence about the intervention effectiveness; Appropriate = Evidence about the intervention appropriateness; Access = Access to the intervention by the consumer; Quality = Quality of the evidence; Duration = Duration of the intervention effect; Immediacy = Immediacy of the intervention effect; Cost = Cost of the intervention; Harms = Evidence about the intervention harms.

4.7 DISCUSSION

To the authors' knowledge, this is the first study to investigate, using a cross-sectoral and multi-disciplinary sample, what stakeholders consider when choosing or recommending interventions for hip or knee OA. A Framework of three over-arching categories of factors considered by stakeholders were identified: characteristics of interventions, characteristics of consumers, and characteristics of health systems. We identified, and ranked by importance to stakeholders, nine characteristics of interventions.

The categories and sub-categories of the Framework developed here overlap with other conceptual frameworks,[27, 170] underlining the Framework's construct validity. A review[27] of the factors influencing implementation success in primary care identified themes, many of which meshed with our Framework, including: "external context", "organisation", "professional" and "characteristics of interventions". Constructs of the Consolidated Framework for Implementation Research[170] also overlapped with our Framework: "intervention characteristics", "outer setting", "inner setting" and "characteristics of individuals". Further, in a study of arthritis MoCs[232] the following concepts were also considered important: "patient self-management", "provider skills and expertise", "health care delivery", "health system", "community and public awareness" and "primary prevention". While supporting existing (generic) frameworks,[27, 170, 232] the present Framework provides important context-relevant information about factors relevant to the New Zealand health system for selecting and delivering OA care, which have not been established previously. Establishing context-relevant factors are important in informing implementation initiatives.[352]

The nine characteristics of interventions of our Framework span clinical (appropriateness, effectiveness, harms and the quality of the evidence) and health system-related characteristics (accessibility and cost of the intervention) that align with the principles of contemporary MoCs for OA; i.e. "the right care, delivered at the right time, by the right team, in the right place and with the right resources".[352, 355] Alignment of these characteristics with stakeholder preferences is important to optimise the quality of OA care delivered within a health system. The description of the characteristics of interventions as discrete sub-categories that were generalisable, complete and non-redundant makes them particularly useful for informing MCDA methods, which go beyond rank-ordering (such as in this study) to more sophisticated approaches that explicitly evaluate the characteristics quantitatively,[266] which may help decision-makers identify which interventions offer the greatest value to the health system and society or, in the case of consumers, the individual.

This research highlights a discordance between what influences stakeholders' choice of intervention relative to the evidence typically provided to guide decision-making (e.g. CPGs). Of the nine characteristics of interventions, 'effectiveness' and 'quality of the evidence' are routinely evaluated in the traditional evidence-based approach, and 'appropriateness' and 'harms' are often considered. The identification of these characteristics was, therefore, unsurprising. However, five other characteristics also arose, including: accessibility, cost and passivity of the intervention, immediacy and duration of the intervention effect. These additional characteristics are highly relevant to health consumers and providers, but typically lie outside the evidence-based paradigm and in healthcare provider and policy-makers' considerations in practice.[353] The GRADE approach to clinical guideline formalisation does encourage guideline committees to also consider these contextual factors when making recommendations, however their consideration remains subjective.[359] These results may therefore carry implications for guideline developers including greater transparency in decision-making. Further, a principle aim of CPGs is to provide recommendations,[14, 202] which are an indicator of 'appropriateness' of the intervention for a given application. However these recommendations are generally formed on the basis of the characteristics of ‘effectiveness’, ‘quality of the evidence’, and ‘harms’, meaning they are non-independent of the characteristic ‘appropriateness’.

Matching stakeholders’ preferences for characteristics of interventions with the interventions provided and delivered in a health system may be fundamental to maximising uptake, in a context where resources are limited but many choices exist. In this study for example, immediacy of the intervention effect was considered more important by the consumer and Māori health advocacy groups, whereas the other (delivery sector) groups ranked it consistently last. These findings reinforce the importance of considering context-specific preferences in efforts to translate evidence into practice and policy[27, 171] that may be particularly important in the current policy landscape with calls for a national OA Model of Care and programmatic funding for OA care from the Ministry of Health.[8, 157]

The Framework’s categories reflect the themes generated by the focus groups. Many of the themes overlapped between groups and were not mutually exclusive, highlighting the complex nature of intervention choice. The Māori and health care provider groups, in particular, generated themes specifically about culturally appropriate practices and health services which incorporated “whānau” (family) and community needs (e.g. education about OA) that were not specifically raised by the policy-maker group. Briggs and colleagues[355] note that inclusion of family support factors into MoCs are poorly integrated, perhaps due to inadequate consumer participation in co-design. Due to low sample size, our results with respect to the Māori health advocate group’s preferences for characteristics of interventions should be interpreted with caution, however these preliminary results suggest Māori have culturally moderated healthcare considerations which must be considered by healthcare providers in NZ to achieve successful implementation and reduce health disparities, and that providers recognise its importance. The NZ Ministry of Health’s Māori Health Strategy[360] exemplifies active Māori participation in planning, development and delivery of appropriate health services for Māori people, reinforcing the benefits of inclusive policy-making.

Our study is not without limitations. Recruitment relied on convenience and snowball methods, which are prone to selection bias, but we filtered participants through purposive sampling which, while intended to enhance diversity, may have inadvertently biased our results towards the views and opinions held by the authors. Moreover, the sample size for each stakeholder group was modest, which may have constrained the breadth of considerations reported in this study when choosing or recommending OA interventions, and the ranked importance of the characteristics of interventions. The necessity to use a modified NGT process for the OA Expert group, conducted remotely due to their geographic dispersion, limited comparison of the NGT results with the other groups. Analysis of NGT data from multiple groups that vary in sample size and produce diverse themes is complex,[326] however we included several verification and validation processes in our design to ensure validity, robustness, and generalisability of these results for the NZ population.

4.8 CONCLUSION

Intervention choice is complex and influenced by the characteristics of interventions, consumers and the health system. This finding has implications for developers of multi-level care strategies, such as MoCs. Currently NZ has no such strategy for managing OA, making the reporting of these context specific characteristics timely.[8] Future research could incorporate the nine discrete characteristics of interventions identified in the current study into MCDA, to support complex decision-making, such as those faced by guideline developers.

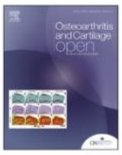
SUPPLEMENT 1

Osteoarthritis and Cartilage Open 2 (2020) 100062

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131



Choosing interventions for hip or knee osteoarthritis - What matters to stakeholders? A mixed-methods study

Jason Chua[a], Andrew M. Briggs[b], Paul Hansen[c], Cathy Chapple[d], J. Haxby Abbott[a,*]

[a] *Centre for Musculoskeletal Outcomes Research, University of Otago, New Zealand*
[b] *School of Physiotherapy and Exercise Science, Curtin University, Australia*
[c] *Department of Economics, University of Otago, New Zealand*
[d] *School of Physiotherapy, University of Otago, New Zealand*

ARTICLE INFO

Keywords:
Choice behavior
Focus groups
Qualitative research
Mixed-methods research
Musculoskeletal diseases
Osteoarthritis

SUMMARY

Objective: To identify what factors drive choices among interventions for hip or knee osteoarthritis (OA), and to rank the characteristics of interventions.

Methods: In this phased, mixed-methods design involving cross-sectoral and multi-disciplinary stakeholders – healthcare consumers, providers, policy-makers, Māori health advocates and OA experts – we used the Nominal Group Technique in focus groups to generate data. We conducted thematic analysis of the focus group data to inform a framework of categories and sub-categories describing factors and characteristics influencing the choice of OA interventions. We then used a dual-panel, two-round e– Delphi survey to verify the framework and rank the characteristics of interventions.

Results: From six focus groups ($n = 38$ participants), 364 factors were identified and clustered into 56 themes (mean 9 themes per focus group; range 5–15). Thematic analysis revealed a framework of 3 core categories: characteristics of interventions (10 sub-categories), characteristics of consumers (10 sub-categories) and characteristics of the health system (7 sub-categories). In Delphi round 1, the framework was verified by each of two panels ($n = 65$, $\geq 80\%$ acceptability). In round 2, two characteristics of interventions were combined, resulting in 9 characteristics (in decreasing order of importance): effectiveness, appropriateness, quality of the evidence, accessibility, harm, cost, duration, passivity, and immediacy of intervention effect.

Conclusion: Stakeholders make choices among interventions for hip or knee OA within a framework of characteristics of interventions, of consumers, and of the health system. We identified and ranked 9 key characteristics of interventions that stakeholders consider when choosing or recommending interventions for hip or knee OA.

1. Introduction

Clinical practice guidelines (CPGs) for the management of hip and knee osteoarthritis (OA) consistently recommend core management strategies of exercise, weight loss, education and self-management [1,2]. Yet the care received by people with OA is often at odds with these core strategies [3,4]. Successful implementation of interventions and health system strategies for OA are influenced by numerous factors, often involving multiple actors [5]. For example external context (e.g. health policy and infrastructure); organizational culture (e.g. culture and leadership); professional attributes (e.g. beliefs and attitudes to change) [5]; and consumers' and providers' attitudes about OA and OA interventions [6,7].

Fundamental to an intervention's successful implementation is its 'fit' within a local health system (context), which depends on the compatibility of the characteristics of interventions with stakeholders' preferences for those characteristics [5]. The preferences and expectations [8] of health care consumers and providers about characteristics such as treatment effectiveness, cost and accessibility may impact the intervention's overall suitability within a particular context. For these reasons, characteristics of interventions must be considered when developing or implementing local health system strategies, such as Models of Care (MoC) [9]. Preferences for these characteristics are often studied at one level of the health system and are less focussed. For example 'healthcare decision-making' among clinical decision-makers and policy-makers [10], or separately among patients [11]. An important knowledge gap

---

[*] Corresponding author. Centre for Musculoskeletal Outcomes Research, University of Otago, Dunedin, 9054, New Zealand.
E-mail address: haxby.abbott@otago.ac.nz (J.H. Abbott).

https://doi.org/10.1016/j.ocarto.2020.100062
Received 6 January 2020; Accepted 29 March 2020

2665-9131/© 2020 Osteoarthritis Research Society International (OARSI). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

remains concerning the characteristics multi-level stakeholders consider, and their preferences for these characteristics.

The aims of this study were to: 1) identify the factors considered by relevant stakeholders, across a range of health care settings, when choosing or recommending OA interventions and express these in a conceptual framework; and 2) define and rank the main characteristics of interventions according to the relative importance placed on them by stakeholders.

2. METHODS

2.1. Design

In this two-phase, mixed-methods study (Fig. 1), phase 1 consisted of focus group discussions conducted using the Nominal Group Technique (NGT) [12]. The NGT data were pooled and thematically analysed to identify priority themes [12,13]. Phase 2 was a dual-panel, two-round, eDelphi study. Round 1 aimed to verify the framework and the focus of round 2 was to rank the importance of the characteristics of interventions. Our design included these validation processes: (i) independent parallel coding of NGT data with reduction into a composite framework to minimise researcher bias (ii) a dual-panel eDelphi which included an external and independent panel to verify the content validity of the Framework.

The study was conducted in New Zealand (NZ) between November 2016 and August 2017. The methods and results for phase 1 are presented according to COREQ-32 [14] criteria for reporting qualitative studies (Supplement 1). Study phase 2 follows guidance on reporting and conducting Delphi studies (CREDES) [15] and the recommendations of Diamond and colleagues (2014) [16]. Ethical approval was obtained from the University of Otago Human Ethics Committee (D16-329).

2.2. Sampling and recruitment

2.2.1. Focus groups

We sampled 5 stakeholder groups: health care consumers, providers, policy-makers, Māori health advocates and OA experts. Convenience and snowball sampling generated lists of stakeholders, who were screened against inclusion criteria adapted from a previous qualitative study of cross-sectoral experts in musculoskeletal MoC [17] (Table 1). Purposive sampling of eligible stakeholders was used to maximise the diversity of participant perspectives, to reduce the risk of limited perspectives being represented within stakeholder groups. The Māori health advocates group were included as Māori experience overall disproportionately worse health outcomes relative to the non-Māori population in Aotearoa/NZ [18]. For each of the above stakeholder groups, focus groups with 4–8 participants were formed [19].

Stakeholders from professional, government and non-government organisations and health care provider agencies were invited to participate by telephone and email. Consumers were targeted using local newspaper advertisements and recruitment flyers at local community centres. 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/). OA experts were initially identified using an online database (http://expertscape.com/) and reviewed by two authors (JC and JHA) with regards to their eligibility. Because of the small population of OA experts in NZ, Australian OA experts were also included on the basis that beliefs about the management of OA should not differ between the two countries given similarities of the health systems in these neighbouring countries and consistency of international clinical practice guidelines [1,2].

2.2.2. eDelphi survey

The dual-panel eDelphi survey was conducted using Qualtrics (Provo, USA). Panel A included phase 1 participants. Panel B, convened to concurrently assess the content validity of the Framework, comprised participants satisfying the same inclusion criteria but who were unable to take part in the focus groups.

3. Protocols & data analysis

3.1. Focus group protocol (phase 1)

In phase 1, focus groups were facilitated using the mixed-method NGT. The NGT structures group discussion in response to a question that can then be prioritised through group consensus in real-time. Focus groups generated data comprising: factors (representing the factors affecting stakeholders' choice of OA interventions), themes (factors grouped into themes, by participants), top 7 themes ranked by importance (by the participants), audio transcripts and field notes (hereinafter

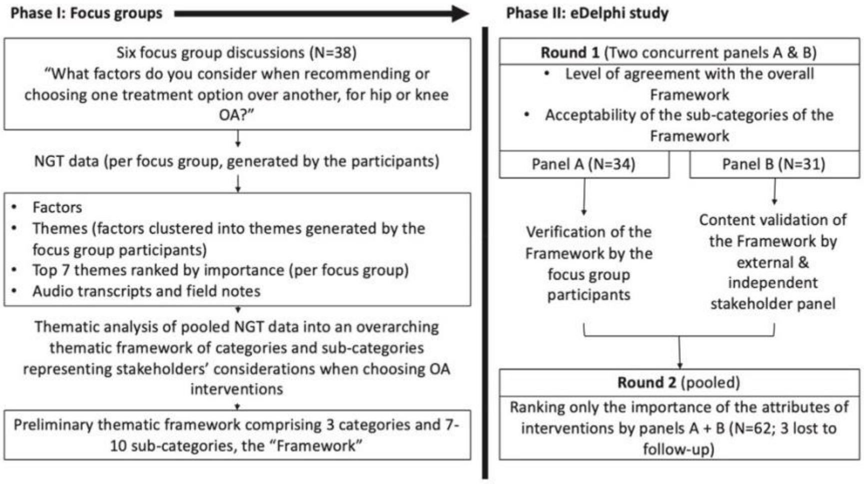


Fig. 1. Summary of the study design.

Table 1
Inclusion criteria for the focus group discussions (phase 1).

Stakeholders	Health policy/strategy, advocate or health service/programme delivery in NZ	Osteoarthritis expert in clinical or health services research in NZ or Australia	Clinical practitioner	Consumer[d]	Maori health advocate
Inclusion criteria	• Experience in health service or programme delivery, coordination, management or funding related to musculoskeletal health care and/or chronic disease (that implicitly includes OA) for at least 1 year at a senior level in NZ[c] • Experience at a senior[a] level in musculoskeletal and/or chronic disease (that implicitly includes OA) health policy or advocacy for at least 1 year in NZ • At least one participant reflecting each of the sub-categories below	• Awareness of guidelines for OA • At least 5 years experience in undertaking clinical and/or health services research in arthritis or health care-related implementation science at a senior level[b] with evidence of peer-reviewed publication(s) in the area • At least one participant reflecting each of the sub-categories below	• Experience in delivery of clinical care for people with arthritic conditions at a senior practitioner level[b] for at least 5 years in NZ • Awareness of guidelines for OA • At least one participant reflecting each of the sub-categories below	• Meets the NICE clinical criteria for diagnosis of OA (i.e.: 45 years or over; has activity-related joint pain, and either; has no morning joint related stiffness or morning stiffness lasting less than 30 min) • At least one participant reflecting each of the sub-categories below	• At least one participant reflecting each of the sub-categories below
Sub-categories	• Community services health policy or strategy decision-maker • Public health services health policy or strategy decision-maker (District Health Boards and Primary Health Organisations) • Assessment, treatment and rehabilitation services health policy or strategy decision-maker • National-level advocacy for arthritis health care • Health service management or coordination at secondary hospital setting • Health service management or coordination at primary care setting	• Model of care or clinical practice guideline development expert • Clinical epidemiologist in OA • Academia and education in OA	• General Practice • Physiotherapy • Rheumatology • Orthopaedic surgery • Community pharmacy	• At least 40% male • At least 40% female • National-level advocacy for arthritis health care	• At least one male • At least one female • Maori service provider

[a] At least senior officer or manager level of employment.

[b] Fellowship level for medical practitioners (e.g. FRNZCGP); senior clinician level for other disciplines; associate professor level for researchers.

[c] At least at the manager or head of department level.

[d] Self-reported diagnosis of OA or determined by the National Institute for Health and Clinical Excellence guideline recommendation for diagnosis of OA [36].

"NGT data"; see Fig. 1). From these data, thematic analysis resulted in a framework (hereinafter "the Framework") of categories and sub-categories (term descriptors are seen in Fig. 2).

Face-to-face and virtual (digital) focus groups were conducted for each stakeholder group separately using the NGT [12]. Each focus group was audio-recorded and transcribed verbatim for analysis. The question posed to participants to discuss was deliberately broad to capture all possible characteristics of interventions: "What factors do you consider when recommending or choosing one treatment option over another, for hip or knee OA?" Consumers were posed the question in the context of their own experience: "What factors do you use to choose one treatment option over another for your OA?"

The first author (JC) was trained in the NGT by a researcher (CC), experienced in the method; CC facilitated the first focus group, assisted by JC, who facilitated all subsequent sessions. Two hours were allocated for each group, structured using five standard NGT steps [12]: (1) introduction and explanation (~5 min), (2) individual and silent generation of factors (15 min), (3) sharing factors identified in step 2, individually with no group discussion (20 min), (4) group discussion and participant-led clustering of factors into themes (40 min), and (5) voting and ranking of themes (20 min). These 5 steps are detailed in Supplement 2.

In anticipation of the focus groups generating potentially many themes, participants were asked to identify and rank their top 7 themes – on the basis that this number of characteristics is commonly used in health valuation studies employing multi-criteria decision analysis (MCDA) [20]. In the case of OA experts, participating in a virtual environment using a modified NGT, steps 1 and 2 were performed online using Qualtrics. Responses were collated and clustered offline by JC. For step 4, the preliminary clustered ideas were then presented to the OA experts, using Zoom videoconferencing (San Jose, USA) and Realtime-Board (www.realtimeboard.com).

3.2. Qualitative data analysis

For the thematic analyses we used the abductive approach [21] which builds on the general inductive approach described by Thomas [22]. From this perspective we accepted a priori that the research team's primary frame of reference (evidence-based practice and evidence-based policy-making) [23,24] would influence the thematic analyses.

First, the NGT data generated from the focus groups were organised using the software package Atlas.ti (Ver.7, Berlin, 2015).

Second, JC performed a thematic analysis of the data following the steps described by Thomas (2006) [22], and guided by the advice of McMillan and colleagues (2014) to aggregate diverse and multiple-group NGT data [13]. The iterative process of open-coding raw data began with reading all results and transcripts from the consumer and health care provider focus groups, verbatim. Preliminary codes emerging from the text resulted in the creation of categories and sub-categories (hereinafter, "themes" and "factors" refer to patterns identified in the raw data; "categories" and "sub-categories" refer to the analysts' constructions of the data). The remaining focus group transcripts were coded using this abductively-derived coding framework, until a broad thematic framework of participants' considerations emerged. Next, the analysis focussed on identifying specific NGT data that were related to the overarching aim of the study. These data were open coded by JC until a primary thematic framework of stakeholders' considerations emerged.

Thirdly, a thematic framework representing participants' choice-making with respect to OA interventions was developed. Using the same analytical approach undertaken by JC, two researchers experienced in qualitative analysis (AMB, JHA) independently coded and analysed the transcribed NGT results and illustrative quotes to develop two independent frameworks. JC then mapped these two frameworks onto the primary framework to identify areas of concordance and discordance. In two meetings (arbitrated by PH) these 3 analysts reviewed the three frameworks and the results of the mapping exercise to agree on a consensus framework (hereinafter referred to as the Framework) consisting of categories and sub-categories.

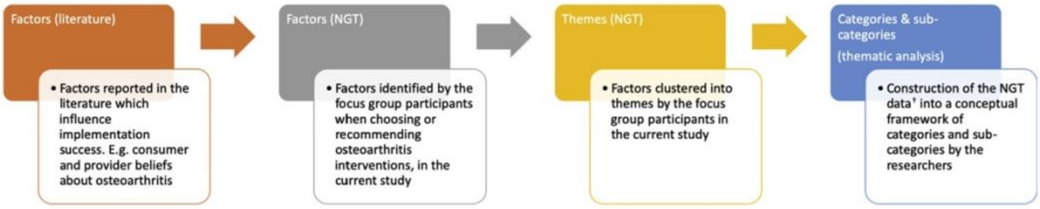
Consistent with the a priori research question, i.e. to focus on the characteristics of interventions, the sub-categories describing the characteristics of interventions were further refined by consensus (JC, JHA, AMB) into discrete sub-categories that were generalisable, complete and non-redundant – desirable characteristics of criteria in MCDA [25,26].

3.3. eDelphi protocol (phase 2)

Phase 2 was a two-round, dual-panel online Delphi (eDelphi) survey to verify the Framework and to rank the sub-categories describing characteristics of interventions (Fig. 1). The protocol was based on a previously published process [9].

A two-round eDelphi was planned a priori, regardless of the overall level of agreement with the Framework after round 1. Two panels were planned for verification the Framework and validation of the results. Participants anonymously completed the survey and were eligible for a NZ$100 gift voucher upon completion of both rounds. Email reminders were sent after each round to increase the response rate.

The aims of round 1 were to: i) verify that no characteristics of interventions important to stakeholders were missing from the analysis (Panel A) and; ii) establish the content validity of the Framework (Panel B). Both panels provided ratings of agreement for the overall Framework (assessed using a 5-point Likert scale: strongly agree, agree, neutral, disagree, strongly disagree) and acceptability of the derived sub-categories (assessed using a 3-point Likert scale: completely agree, partially agree, disagree). Free text fields allowed participants to comment on any component of the Framework, and participants were prompted to explain their answer on any disagreed (disagree/strongly disagree) component. Participants were also asked to add any essential characteristics of interventions missing from the Framework. Demographic data were captured in round 1. Round 2 focussed on ranking the importance of the characteristics of interventions. Panels A and B were pooled for this quantitative analysis.



NGT= Nominal Group Technique
†factors, themes, themes ranked by importance, audio transcripts and field notes

Fig. 2. Description of terms used in this paper for factors, themes, categories and sub-categories.

3.4. *eDelphi survey data analysis*

An a priori consensus for round 1 of the eDelphi was defined as ≥80% complete or partial agreement for the sub-categories (3-point Likert scale: completely agree, partially agree, disagree), and ≥80% agree or strongly agree for the overall framework (5-point Likert scale: strongly agree, agree, neither agree or disagree, disagree, strongly disagree). The qualitative aspect of round 1 comprised content-analysis of participants' free-text comments to identify new categories or sub-categories emerging from either panel. We performed further analysis of free-text comments across all the sub-categories rated as 'disagree' from >20% of respondents (per group) to clarify the descriptions of the characteristics of interventions carried into round 2.

The aim of round 2 was to rank the characteristics of interventions by importance. Consistent with the a priori research question, from this point on the eDelphi survey was limited to the characteristics of interventions; other categories and sub-categories of the Framework were not evaluated further. We calculated the average rank of the characteristics of interventions, from most-to least-important, within stakeholder groups and for the overall sample. Level of agreement was assessed using Kendall's W within and across the groups, where 0.1, 0.3, 0.5, 0.7 are very weak, weak, moderate and strong levels of agreement respectively [27]. To ensure the final set of characteristics was representative of all groups (which differed in size), the top two ranked characteristics for each group were included in the final aggregated list of characteristics, regardless of their overall ranking, providing that adequate within-group agreement was demonstrated (Kendall's W > 0.1, $p < 0.05$). Data were analysed using IBM SPSS Statistics (ver.24, Armonk, NY).

4. RESULTS

4.1. *Focus group discussions*

Six focus group meetings were held (N = 38, 61% female; n = 3–11 participants per group, median 6 [IQR 3.3]), each lasting approximately 90 min (range 60–120 min). Roles, experience and number participating in each stakeholder group are summarised in Table 2. Steps 2 and 3 of the NGT produced 364 factors. After the removal of duplicates, 258 unique factors emerged across focus group discussions (mean 43 per group; median 39; range 29–62). Step 4 produced a median of 9 (range 5–15) participant-clustered themes per group (total 56 overlapping themes). Step 5 produced themes ranked by importance which broadly overlapped across groups concerning consumer-related factors (e.g. clinical status and consumer preferences), themes related to efficacy and safety of intervention and health system factors (e.g. intervention accessibility and culturally appropriate practices). The OA expert group, conducted by the modified NGT that did not include step 5, produced themes echoed in the other groups. The results of the NGT exercise (the NGT data) are reported in Supplement 2, Table S1.

Initial independent thematic analysis by the researchers (JC, AMB, JHA) produced three frameworks, each with 3–10 categories, comprising 5–17 sub-categories. Two rounds of discussion leading to consensus resulted in a thematic framework of three over-arching categories: 1) characteristics of consumers (10 sub-categories), 2) characteristics of interventions (10 sub-categories), and 3) characteristics of health systems (7 sub-categories) (Fig. 3). A summary description of the Framework is presented in Table 3(sub-category descriptions and supporting illustrative quotes are in Supplement 2, Tables S2–S3).

4.2. *eDelphi survey*

Delphi round 1 included 65 participants (4 phase 1 participants lost to follow-up in Panel A, N = 34; Panel B, N = 31). In round 2, 3 participants were lost to follow-up due to personal circumstances (95% retention rate). The panels' demographic characteristics are summarised in Table 2.

4.3. *eDelphi round 1 outcomes*

The overall acceptability of all the categories of the Framework was confirmed by panels A (91% strongly agree or agree) and B (94% strongly agree or agree) and by the stakeholder groups (Supplement 3,

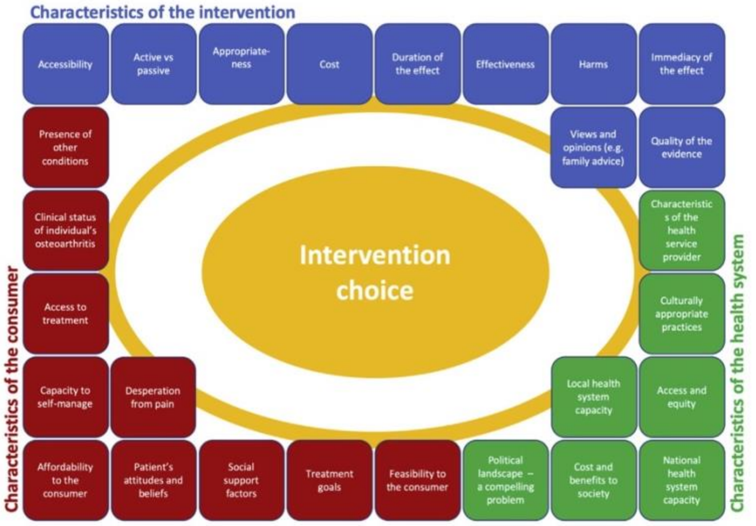


Fig. 3. Thematic representation of the three over-arching categories and sub-categories, derived from the 56 themes generated by the focus discussion, that stakeholders consider in their choice of hip or knee osteoarthritis intervention.

Table 2
Demographic characteristics of the focus group (phase 1) and eDelphi survey (Phase 2) participants at rounds 1 and 2.

	Focus groups (Phase 1)			eDelphi Round 1 (Phase 2)						eDelphi Round 2		
	(N = 38)			Panel A[a]			Panel B (N = 31)			Pooled[b]		
Stakeholder group[c]	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)	n (%), age ± SD (range) in years	Female n (%)	Years in current role ± SD (range)
Consumers	**11 (29), 69 ± 10 (50–88)**	**5 (13)**	**16 ± 16 (1–50)**	**9 (26), 69 ± 10 (50–88)**	**5 (15)**	**14 ± 13 (1–35)**	**7 (23), 67 ± 9 (50–75)**	**5 (16)**	**10 ± 8 (1–25)**	**15 (24), 68 ± 10 (50–88)**	**9 (15)**	**13 ± 11 (1–35)**
Māori health advocates[d]	**5 (13), 37 ± 12 (25–57)**	**4 (11)**	**5 ± 4.4 (2–10)**	**5 (15), 37 ± 12 (25–57)**	**4 (12)**	**5 ± 4 (2–10)**	-	-	-	**5 (8), 37 ± 12 (25–57)**	**4 (6)**	**5 ± 4 (2–10)**
Health care providers[e]	**10 (26), 49 ± 13 (33–67)**	**5 (13)**	**22 ± 9 (11–30)**	**9 (26), 49 ± 13 (33–67)**	**4 (12)**	**21 ± 9 (11–30)**	**14 (45), 49 ± 10 (30–66)**	**5 (16)**	**20 ± 10 (8–37)**	**22 (36), 48 ± 11 (30–67)**	**9 (15)**	**20 ± 9 (8–35)**
General practitioner	4 (40), 55 ± 11 (41–64)	2 (20)	25 ± 7 (15–30)	3 (33), 56 ± 13 (41–64)	1 (11)	26 ± 8 (15–30)	4 (29), 47 ± 9 (39–59)	3 (21)	18 ± 8 (12–30)	8 (36), 53 ± 12 (39–67)	4 (18)	22 ± 9 (12–30)
Physiotherapist	5 (50), 46 ± 14 (33–67)	2 (20)	21 ± 11 (11–30)	5 (56), 46 ± 14 (33–67)	2 (22)	21 ± 11 (11–30)	4 (29), 44 ± 8 (30–56)	1 (7)	12 ± 4 (8–15)	8 (36), 40 ± 8 (30–55)	3 (14)	15 ± 7 (8–30)
Orthopaedic surgeon	–	–	–	–	–	–	3 (21), 56 ± 10 (46–66)	0 (0)	24 ± 12 (15–37)	2 (9), 51 ± 6 (46–55)	0 (0)	18 ± 4 (15–20)
Pharmacist	1 (10), 39	1 (10)	15	1 (11), 39	1 (11)	15	–	–	–	1 (5), 39	1 (5)	15
Rheumatologist	–	–	–	–	–	–	3 (21), 57 ± 4 (52–59)	1 (7)	30 ± 5 (25–35)	3 (14), 57 ± 4 (52–59)	1 (5)	30 ± 5 (25–35)
Health policy-makers[f]	**7 (18), 53 ± 10 (34–62)**	**5 (13)**	**6 ± 7 (1–20)**	**6 (18), 52 ± 11 (34–62)**	**5 (15)**	**8 ± 7 (1–20)**	**4 (13), 53 ± 7 (44–59)**	**4 (13)**	**14 ± 11 (6–30)**[g]	**9 (15), 53 ± 9 (34–62)**	**9 (15)**	**10 ± 10 (1–30)**
Heath policy or programme developers	1 (14), 62	1 (14)	3 ± 0 (0)	–	–	–	–	–	–	–	–	–
Health system funders	2 (29), 60 ± 1 (59–60)	1 (14)	20 ± 0 (0)	–	–	–	–	–	–	–	–	–
Health service delivery, coordinators or managers	4 (57), 48 ± 11 (34–59)	3 (57)	3 ± 3 (1–7)	–	–	–	–	–	–	–	–	–
Osteoarthritis experts	**5 (13), 50 ± 7 (41–60)**[h]	**4 (11)**	**12 ± 5 (8–15)**	**5 (13), 50 ± 7 (41–60)**[h]	**4 (12)**	**12 ± 3 (8–15)**	**6 (19), 52 ± 8 (40–60)**[h]	**3 (10)**	**16 ± 10 (5–34)**	**11 (17), 51 ± 7 (40–60)**	**7 (11)**	**14 ± 8 (5–34)**

[a] Response rate of N 34 (89%).
[b] Response rate of N 62 (95%).
[c] Defined by primary role.
[d] Two stakeholders were interviewed face-to-face.
[e] Two focus group meetings were conducted for the health care provider group.
[f] Sub-group level data was not collected for eDelphi survey.
[g] p 0.036 (Mann-Whitney U).
[h] n 4 Australians.

Table 3

Summary descriptions of the framework of factors influencing stakeholders choice of OA intervention derived from the focus group discussions.

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
Characteristics of the intervention Stakeholders gathered information about interventions to assess for suitability against a patient's circumstances or feasibility within a health system. The characteristics or attributes of interventions, applicable to any number of interventions, were used to compare and contrast between interventions. Key characteristics of interventions included: evidence for effectiveness and safety, consideration of the benefit and trade-off, cost-effectiveness (to the consumer), immediacy of treatment and access to treatment. Information about the intervention options were also gathered from health professionals, peer/family experience and advertising sources (e.g. newspaper and online advertising) in the decision to select a treatment option for consumers.	• Access to treatment for the consumer. The suitability of access considered in terms of an intervention's distance to nearest provider and wait time. • Active versus passive intervention options. The extent to which an intervention allows a patient to self-manage their condition over the long term, versus passive care. • Cost of the intervention. The financial costs of intervention relevant to the use or provision of health care and society. • Duration of the intervention effect. The ability to delay or avoid more invasive interventions later through improvements in quality of life (e.g. physical function). • Evidence about the intervention appropriateness. Providing the right treatment or services for OA at the right time (e.g. surgical intervention for early- or advanced-stage OA). • Evidence about the intervention effect. Considered on the basis of evidence outside a clinical trial, including short-term and long-term evidence of effectiveness and change in health system and societal costs. • Evidence about the intervention harms. The risk of short- and long-term side effects of intervention. • Immediacy of the intervention effect. The time delay between starting and experiencing the benefits of intervention. • Quality of the evidence (+*views and opinions*). The extent to which one could be confident that the effects of the treatment or service described were real. *Views and opinions* concern the value which different people assign to different sources of evidence. For example, peer or family advice and/or personal accounts were viewed as more trustworthy for some participants than health professionals' or advertisement (e.g. newspaper, Facebook) advice, claims or recommendations about the benefits or harms of intervention.	• *"Availability; for example, I explored physiotherapy and exercise options but I live an hour or so out of town ... to access services whether it be the public system, whether it be close to; in my case travel's a factor."* Consumer #6 • *Capacity and confidence and motivation might be different; I might put that as a different 'cause somebody might have the capacity and confidence, but they're not motivated to change ... if you look at the goal-setting care plans, the way they're done, those sorts of things [factors related to self-management] would definitely be taken into consideration."* Osteoarthritis expert #1 • *"A whole lot around cost; what the particular cost is. The difference to the lifetime cost system and also the patient's lifetime costs. The value for money. (Health Policy #5)* • *"I made a note about this just while conversation was going on earlier about surgery and last resorts because from my point of view I would much prefer surgery to a drug medication that goes on and on and on."* Consumer #7 • *"I guess in a sense we might say that all the evidence or whatever we might come up with on paper says that this is the right treatment, but actually for certain groups of people in society it might be the wrong treatment because it might be unfeasible because they're rural or can't afford to travel. So therefore it's the wrong treatment for that person or even that group at that time."* Health Policy Maker #2 • *"Well, efficacy, simply does the drug work; or does the treatment work? And with more complex interventions, say we've got multidisciplinary clinics, or whatever, then you'd need effectiveness which is showing that it works within a context."* Osteoarthritis expert #2 • *"It doesn't matter whether it's a drug that's actually prescribed by the doctor or this Arthrem or any of those other things; you are putting something into your body and things like that. Is it going to be worth it? Is it not? You're weighing this up all the time."* Consumer #4 • *"The problem is though you see on the advertisements and things like that and these people stand up and say, 'This is the best thing since sliced bread,' and you think 'Right, okay, I'm going to give this a go.' How much are they being paid to say that and in actual fact have they had that working for them; because you've got no way of checking. All it is, is you think "Well if it's worked for that person do I spend the money and give that a go," but you have no knowledge of it."* Consumer #9
Characteristics of the consumer When choosing or recommending an intervention for OA, stakeholders considered it important to match an intervention to the biopsychosocial profile of the patient. Key factors included an assessment of a patient's severity of OA, clinical comorbidity profile, preferences (inclusive of attitudes, beliefs, ability to self-manage, motivation and treatment goals) and access to interventions. Social considerations included support to/from the family and/or community responsibilities.	• Clinical status of individual's osteoarthritis. The recommendation or selection of intervention based on objective tests, patient reported symptoms and age. • Presence of other conditions. For example, diabetes, heart disease and mental health, and the medical care being received for these conditions which may influence intervention choice. • Patient attitudes and beliefs. Intervention options need to reflect a consumer's (and their family's) beliefs about OA, as well as beliefs or expectations about the intervention options. • Capacity to self-manage. Consumer's health literacy, attitudes and beliefs about OA (e.g. perception of pain), and capacity to self-manage and navigate the health system. • Treatment goals. Treatment options need to relate to the immediate and longer-term functional goals of the consumer now and into the future. • Desperation (e.g. from pain). Once consumers felt that they had exhausted the therapeutic options immediately available to them to manage the symptoms of OA they sought to improve their quality of life by trying other interventions or medicines (e.g. herbal remedies) with poor evidence supporting their safety, benefit or interaction with other interventions.	• *"The trouble is there's the objective parts of what you find out about the patient and then there's the subjective part that they contribute which they talk about. So, to me those are the two separations and I would see them as being quite important to me. So, your x-rays, your scans, your blood tests and all that, I see them as being objective versus what the patient says and how incapacitated they say they are, all that sort of stuff which a subjective thing is."* Health care provider #1 • *"... a couple of things that always worry me are whether they've got a chronic pain syndrome and whether they've got a chemical addiction already when you're considering what you're gonna do. Those two things influence me quite a bit with what I'm going to suggest that they might contemplate doing."* Health care provider #1 • *"Desire. Like after all is said and done you can explain the benefits, present evidence based practice, and if they don't want to take it they just don't want to take it. They've got to want to; they've got to want the treatment. And someone might just want to die and not take the treatment; like I'm over this, I'm sick of being old. I have patients say that: 'Don't get old dear, don't get old'* ... or you might be like the 20th health professional to contact them about this and they just stop ... They're just over it." Maori advocate #1

(continued on next page)

Table 3 (*continued*)

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
	• Access to treatment. The consumer's geographic location and the number of treatment options available (to choose). • Affordability for the consumer. The immediate out of pocket expenses to the consumer. • Social support factors. Consideration of consumers' place in the community and wider community and support requirements if a disability is present. • Feasibility to the consumer. Whether or not a treatment option is acceptable and feasible to the consumer to adopt in the short and long term, based on the biopsychosocial circumstances of the consumer, their goals and expectations of treatment, and access to health services.	• *“I think health literacy, education and level of patient knowledge all relate to what the patient needs to bring to the treatment or treatment decision.”* Health policy #2 • *“I put how I'm feeling; it's time of the day sometimes. Like how I'm feeling; if I'm in a good space or something like that and I'm going somewhere and it's painful or whatever I might decide to not do something. I might decide not to get sort of very heavy about it. Or otherwise I may be having a bad day and I weigh up the pros and cons that way. It's how I'm feeling on the day.”* Consumer #3 • *“Being open for information too rather than just closing your mind off. I think you reach the stage where you will listen to anybody in the hope that something may work.”* Consumer #7 • *“Yeah so it's not so much about equity; it's about would somebody have to travel a big distance and that would be difficult for them. It's where the costs fall for the treatment option and which of them are most accessible for that person.”* Health policy #2 • *“The other thing too perhaps and I don't know quite how you put it in, but I'm still part time working and so my choices up to a point is how it can still allow me to work … Because if something's not going to allow me to continue what I'm doing then I wouldn't be interested in it?”* Consumer #10 • *“I would say family support to and of as well. Like for some Maori it depends on their beliefs and their upbringing but some people would choose a spokesperson; and so everything is dealt with that spokesperson that the patient has nominated as the spokesperson. So you wouldn't necessarily and this is really weird as a health professional that you talk to someone other than the patient, ‘cause it's really the patient that you're dealing with and their experience. But to some Maori they do choose a spokesperson and so everything gets [9.10]. So let's say it's an older person; they might choose their youngest son, or they might choose the youngest daughter and that's the person that you deal with, with all cares concerning or all treatments concerning the patient. So you actually don't really deal with patient; which is different.”* Maori advocate #1
Characteristics of the health system Participants considered whether there was the need for the society to invest in health care for OA and, whether the interventions considered were feasible to adopt or implement from the perspective of the health system and the consumer. Factors considered by stakeholders included the burden of disease profile for OA, socio-political interest, health system capacity, access and equity, culturally appropriate services and, the cost-effectiveness of intervention to society.	• Political landscape is the disease a compelling problem? Before recommending interventions for OA, policy-makers considered if there was community and political interest in addressing the burden of OA. • Costs and benefits to society. Before recommending treatments for OA, policy-makers considered if there was community and political interest in addressing the burden of OA. • National health system capacity. Policy-makers also considered the capacity of a health system to support OA health care including policy direction, funding availability, evidence in support of funding or implementing a new service or intervention and workforce capacity and capability • Local health system capacity. Health policy makers considered the feasibility for a local health system to provide care services for OA in terms of the costs, communication capacity among care providers, the current availability of services (public and private), access to services and scope for integration with other existing services • Access and equity. Consideration of the burden imposed on the consumer to reach an intervention or service for OA was reasonable (e.g. geographical remoteness, fit with lifestyle) and, whether or not it was affordable to those who need it (fairness/equity of access) • Culturally appropriate services. Whether or not a health care service for OA was culturally appropriate for diverse communities. • Characteristics of the health service provider. The beliefs, experiences and preferences of the health service provider.	• *“We've talked about the burden of disease but actually there's a far more less palatable political reality, which is about is there a burning platform? Is there some advocacy for this? Is this both accepted and publicly accepted as being needed? And it might be quite small. So I wonder if burden of disease is only one part of is there a burning platform or public support for this … Sometimes the burning platform is because of the cost to the system. Sometimes it's because of the consumer voice, from some effective advocates, and that's all captured in the how is this framed in the public's mind or little P politically. Because there's a political element to it which might just be local community; I don't mean national politics.”* Health Policy #2 • *“A whole lot around cost; what the particular cost is. The difference to the lifetime cost system and also the patient's lifetime costs. The value for money.* Health Policy #5 • *“… I think the question is what resources are available? What funding is available? Because you're likely to cut your service, cut your coat according to your cloth. You're likely to come up with a service that is feasible within the funding available and other resources available.”* Health Policy #5 • *"… it's actually I guess in a way saying more about the role of the District Health Board. [A DHB] plan for commissioned services in its own district, right? Once it might be a national policy or capability that we're going to provide these services in the country and then the DHBs end up providing those services in their own district according to what's available in their district to do that.”* Health Policy #2

(*continued on next page*)

Table 3 (*continued*)

Key Category and summative description	Sub-categories	Illustrative quotes supporting each sub-category
		• *"I wonder if accessibility at a system level becomes kind of fairness or equity; that at the same time as you want to know can the system deliver this, can you set up treatment options that will carry through, there's something about access and about fairness of access that you might want to pick up as well. (health policy #1) … Yeah … I just was thinking it's more than access; it's about fairness of access which I call equity, but I mean we might frame it differently."* Health policy #5 • *"… educating the family is important also because that patient potentially is living with them, and so they will be offering the cares in some way, shape or form; so having the family, the patient and the whole team, the MDT, all on the same page. Like Māori don't tend to – this could be generalising but in my experience Māori don't put their loved ones into a rest home; so they're at home living with the younger daughter or the niece or somebody. So having those people involved in all aspects is really helpful 'cause that will help the patient overall."* Māori Advocate #2

Tables S1–S2). Analysis of pooled panel and stakeholder-group agreement scores for each sub-category revealed consensus across all sub-categories, except for the "desperation" sub-category (77% partial or complete agreement), confirming the content of 26 of the 27 sub-categories (Supplement 3, Table S3). Analysis of agreement within stakeholder groups identified lack of agreement for the sub-categories (Supplement 3, Table S4): [n (%) stakeholder group] 5 (45%) OA experts for "desperation", 3 (30%) policy-makers for "views relating to the characteristics of the intervention" and 5 (22%) providers for "local health system capacity".

Analysis of stakeholder's free-text comments to the overall framework did not reveal any new categories or sub-categories. The three sub-categories with >20% disagreement are discussed here. Content analysis of free text comments from 4 of the 5 OA experts (1 did not leave a comment) who disagreed with the "desperation" sub-category posited that consumers may be desperate to try new treatments but did not consider it good practice; i.e. they agreed with the meaning of the sub-category per se. Likewise, the comments from three provider's who disagreed with the "local health system capacity" sub-category (2 did not leave comments) concerned local health system barriers, rather than disagreement with the sub-category. One policy-maker commented about their disagreement for the "views relating to the characteristics of the intervention" sub-category, however it was unrelated.

In summary, the eDelphi panellists agreement with the overall Framework and it's sub-categories (excepting the "desperation" sub-category) indicate strong alignment between the stakeholder-generated synthesis and the qualitative analysis, with validation provided by an independent stakeholder panel. Within the "characteristics of interventions" category, consensus was not reached by the stakeholder groups about the "views relating to the characteristics of the interventions" sub-category. Two participants' comments guided the analysis team to conclude that these views related to the evidence concerning interventions – from a range of sources – therefore these two sub-categories could be merged within the "quality of the evidence" sub-category for round 2. Thus, apart from this one change, no other modifications were made to the Framework, confirming its completeness and content validity, and leaving nine discrete characteristics of interventions (sub-categories) to be ranked by participants.

4.4. *eDelphi round 2 outcomes*

There was significant but weak agreement in the rankings across the groups (N = 62; between-group W = 0.333; p=<0.00). A weak-to-moderate level of agreement in rankings was observed within stakeholder groups except for the Māori health advocacy group (N = 5), which did not demonstrate adequate within-group agreement (Kendall's W = 0.089, 'very weak'; p = 0.893). The top two ranked characteristics of interventions across the resulting stakeholder groups were: evidence about the intervention effectiveness, evidence about the intervention appropriateness, quality of the evidence, and immediacy of treatment effect. The Māori advocacy group was the only group to rank immediacy of treatment effect among their top two characteristics; however we were unable to achieve significant agreement within this group. Our failure to observe reliable estimates in this stakeholder group was likely due to the low sample size, despite our extensive efforts, so data from the Māori health advocate group were pooled with those of the healthcare consumer group. The aggregate ranking of the characteristics of interventions, across all groups, was (in decreasing order of importance): effectiveness, appropriateness, quality of the evidence, accessibility, harm, cost, duration, passivity and immediacy of effect (Table 4).

5. Discussion

To the authors' knowledge, this is the first study to investigate, using a cross-sectoral and multi-disciplinary sample, what stakeholders consider when choosing or recommending interventions for hip or knee

Table 4
Characteristics of interventions after eDelphi round 2, in decreasing order of importance (top to bottom), for each stakeholder group. Data presented as the mean characteristic and rank by stakeholder group.

Stakeholder group (W; p-value)					
Consumers	Maori health advocates	Health care providers	Health policy-makers	Osteoarthritis experts	Pooled
Effect (2.53)	Immediacy (3.80)	Effect (2.24)	Quality (2.33)	Effect (1.83)	Effect (2.40)
Appropriate (3.87)	Effect (4.00)	Appropriate (2.33)	Effect (2.44)	Quality (3.08)	Appropriate (3.18)
Access (4.27)	Appropriate (4.20)	Quality (4.05)	Appropriate (3.22)	Appropriate (3.33)	Quality (4.00)
Duration (4.87)	Access (5.20)	Access (4.14)	Cost (5.44)	Harms (4.08)	Access (4.71)
Quality (5.13)	AvP (5.20)	Harms (5.19)	Access (5.78)	Access (5.25)	Harms (5.34)
Immediacy (5.73)	Cost (5.20)	Cost (5.71)	Harms (5.89)	Cost (5.58)	Cost (5.65)
Cost (5.87)	Duration (5.60)	Duration (6.29)	Duration (6.00)	Duration (6.33)	Duration (5.85)
Harms (5.93)	Quality (5.60)	AvP (6.95)	AvP (6.67)	AvP (7.17)	AvP (6.77)
AvP (6.80)	Harms (6.20)	Immediacy (8.10)	Immediacy (7.22)	Immediacy (8.33)	Immediacy (7.10)
(0.222; 0.001)	**(0.089; 0.893)**	**(0.533; < 0.00)**	**(0.452; < 0.00)**	**(0.588; < 0.00)**	**(0.333; < 0.00)**

The mean rank for the characteristic is in parentheses (lower numbers represent greater importance). W Kendall's W; AvP Active vs Passive interventions; Effect Evidence about the intervention effectiveness; Appropriate Evidence about the intervention appropriateness; Access Access to the intervention to the consumer; Quality Quality of the evidence; Duration Duration of the intervention effect; Immediacy Immediacy of the intervention effect; Cost Cost of the intervention; Harms Evidence about the intervention harms.

OA. A Framework of three over-arching categories of factors considered by stakeholders were identified: characteristics of interventions, characteristics of consumers, and characteristics of health systems. We identified, and ranked by importance to stakeholders, nine characteristics of interventions.

The categories and sub-categories of the Framework developed here overlap with other conceptual frameworks [5,28], underlining the Framework's construct validity. A review [5] of the factors influencing implementation success in primary care identified themes, many of which meshed with our Framework, including: "external context", "organisation", "professional" and "characteristics of interventions". Constructs of the Consolidated Framework for Implementation Research [28] also overlapped with our Framework: "intervention characteristics", "outer setting", "inner setting" and "characteristics of individuals". Further, in a study of arthritis MoCs [29] the following concepts were also considered important: "patient self-management", "provider skills and expertise", "health care delivery", "health system", "community and public awareness" and "primary prevention". While supporting existing (generic) frameworks [5,28,29], the present Framework provides important context-relevant information about factors relevant to the New Zealand health system for selecting and delivering OA care, which have not been established previously. Establishing context-relevant factors are important in informing implementation initiatives [9].

The nine characteristics of interventions of our Framework span clinical (appropriateness, effectiveness, harms and the quality of the evidence) and health system-related characteristics (accessibility and cost of the intervention) that align with the principles of contemporary MoCs for OA; i.e. "the right care, delivered at the right time, by the right team, in the right place and with the right resources" [9,17]. Alignment of these characteristics with stakeholder preferences is important to optimise the quality of OA care delivered within a health system. The description of the characteristics of interventions as discrete sub-categories that were generalisable, complete and non-redundant makes them particularly useful for informing MCDA methods, which go beyond rank-ordering (such as in this study) to more sophisticated approaches that explicitly evaluate the characteristics quantitatively [30], which may help decision-makers identify which interventions offer the greatest value to the health system and society or, in the case of consumers, the individual.

This research highlights a discordance between what influences stakeholders' choice of intervention relative to the evidence typically provided to guide decision-making (e.g. CPGs). Of the nine characteristics of interventions, 'effectiveness' and 'quality of the evidence' are routinely evaluated in the traditional evidence-based approach, and 'appropriateness' and 'harms' are often considered. The identification of these characteristics was, therefore, unsurprising. However, five other characteristics also arose, including: accessibility, cost and passivity of the intervention, immediacy and duration of the intervention effect. These additional characteristics are highly relevant to health consumers and providers, but typically lie outside the evidence-based paradigm and in healthcare provider and policy-makers' considerations in practice [10]. The GRADE approach to clinical guideline formalisation does encourage guideline committees to also consider these contextual factors when making recommendations, however their consideration remains subjective [31]. These results may therefore carry implications for guideline developers including greater transparency in decision-making. Further, a principle aim of CPGs is to provide recommendations [1,2], which are an indicator of 'appropriateness' of the intervention for a given application. However these recommendations are generally formed on the basis of the characteristics of 'effectiveness', 'quality of the evidence', and 'harms', meaning they are non-independent of the characteristic 'appropriateness'.

Matching stakeholders' preferences for characteristics of interventions with the interventions provided and delivered in a health system may be fundamental to maximising uptake, in a context where resources are limited but many choices exist. In this study for example, immediacy of the intervention effect was considered more important by the consumer and Māori health advocacy groups, whereas the other (delivery sector) groups ranked it consistently last. These findings reinforce the importance of considering context-specific preferences in efforts to translate evidence into practice and policy [5,32] that may be particularly important in the current policy landscape with calls for a national OA Model of Care and programmatic funding for OA care from the Ministry of Health [33,34].

The Framework's categories reflect the themes generated by the focus groups. Many of the themes overlapped between groups and were not mutually exclusive, highlighting the complex nature of intervention choice. The Māori and health care provider groups, in particular, generated themes specifically about culturally appropriate practices and health services which incorporated "whānau" (family) and community needs (e.g. education about OA) that were not specifically raised by the policy-maker group. Briggs and colleagues [17] note that inclusion of family support factors into MoCs are poorly integrated, perhaps due to inadequate consumer participation in co-design. Due to low sample size, our results with respect to the Māori health advocate group's preferences for characteristics of interventions should be interpreted with caution, however these preliminary results suggest Māori have culturally moderated healthcare considerations which must be considered by healthcare providers in NZ to achieve successful implementation and reduce health disparities, and that providers recognise its importance. The NZ Ministry of Health's Māori Health Strategy [35] exemplifies active Māori participation in planning, development and delivery of appropriate health services for Māori people, reinforcing the benefits of inclusive policy-making.

Our study is not without limitations. Recruitment relied on convenience and snowball methods, which are prone to selection bias, but we filtered participants through purposive sampling which, while intended to enhance diversity, may have inadvertently biased our results towards the views and opinions held by the authors. Moreover, the sample size for each stakeholder group was modest, which may have constrained the breadth of considerations reported in this study when choosing or recommending OA interventions, and the ranked importance of the characteristics of interventions. The necessity to use a modified NGT process for the OA Expert group, conducted remotely due to their geographic dispersion, limited comparison of the NGT results with the other groups. Analysis of NGT data from multiple groups that vary in sample size and produce diverse themes is complex [13], however we included several verification and validation processes in our design to ensure validity, robustness, and generalisability of these results for the NZ population.

Intervention choice is complex and influenced by the characteristics of interventions, consumers and the health system. This finding has implications for developers of multi-level care strategies, such as MoCs. Currently NZ has no such strategy for managing OA, making the reporting of these context specific characteristics timely [33]. Future research could incorporate the nine discrete characteristics of interventions identified in the current study into MCDA, to support complex decision-making, such as those faced by guideline developers.

Author contributions

JC, AMB, PH and JHA contributed to the conception, design and primary analysis of data in the study. JC and CC contributed to preparation and data collection. JC had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All the authors contributed to revisions of the manuscript, the interpretation of the data and critically reviewed the manuscript for publication. JHA conceived of the programme of enquiry, and is the corresponding author for this manuscript. All authors read and approved the final manuscript.

Role of the funding source

The funding source had no role in the design or conduct of the study; collection, analysis, or interpretation of the data; or writing of the report.

Declaration of Competing Interest

AMB declares travel support funding from the University of Otago <$10,000. AMB is supported by a fellowship awarded by the Australian National Health and Medical Research Council (113548).

Acknowledgements

This study is supported by a grant from the Health Research Council of New Zealand (15/263).

Thank you to the participants in this study, including: Associate Professor Ilana Ackerman, Mrs Win Anderson, Mr Stuart Barson, Professor Kim Bennell, Professor Rhiannon Braund, Professor Rachelle Buchbinder, Mr Michael Chin, Mr Don Chuah, Mr Nick Conn, Ms Zoe Cotter, Dr Ben Darlow, Ms Lisa Denny, Ms Janice Donaldson, Professor Anthony Dowell, Associate Professor Michelle Dowsey, Mrs Virginia Duncum, Associate Professor Rebecca Grainger, Mrs Ruth Groffman, Associate Professor David Gwynne-Jones, Dr Katherine Harding, Mrs Pip Herd, Mrs Lynne Hill, Professor Rana Hinman, Dr Ben Hudson, Mr Liam Hutton, Dr Grant Johnston, Mr Arnold Jones, Dr Joanne Kemp, Associate Professor Peter Larmer, Mr John Matheson, Dr Fran McGrath, Dr Jill McIlraith, Mr Graeme Moginie, Ms Cathie Morton, Mr Logan Murray, Mr Leighton Murray, Mr John Neilson, Ms Stacey Porter, Dr Peter Radue, Mr Dallas Roderique, Ms Mary Scally, Mrs Carolyn Sims, Professor Helen Slater, Associate Professor Simon Stebbings, Professor Nicholas Taylor, Mrs Shan White, Ms Uru-Manuka Williams and Ms Becy Wilson.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2020.100062.

References

[1] P.J. Larmer, N.D. Reay, E.R. Aubert, P. Kersten, Systematic review of guidelines for the physical management of osteoarthritis, Arch. Phys. Med. Rehabil. 95 (2014) 375 389.

[2] A.E. Nelson, K.D. Allen, Y.M. Golightly, A.P. Goode, J.M. Jordan, A systematic review of recommendations and guidelines for the management of osteoarthritis: the Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative, Semin. Arthritis Rheum. 43 (2014) 701 712.

[3] M. Basedow, A. Esterman, Assessing appropriateness of osteoarthritis care using quality indicators: a systematic review, J. Eval. Clin. Pract. 21 (2015) 782 789.

[4] K.B. Hagen, G. Smedslund, N. Østerås, G. Jamtvedt, Quality of community-based osteoarthritis care: a systematic review and meta-analysis, Arthritis Care Res. 68 (2016) 1443 1452.

[5] R. Lau, F. Stevenson, B.N. Ong, K. Dziedzic, S. Treweek, S. Eldridge, et al., Achieving change in primary care-causes of the evidence to practice gap: systematic reviews of reviews, Implement. Sci. 11 (2016) 40.

[6] T. Egerton, L.E. Diamond, R. Buchbinder, K.L. Bennell, S.C. Slade, A systematic review and evidence synthesis of qualitative studies to identify primary care clinicians' barriers and enablers to the management of osteoarthritis, Osteoarthritis Cartilage 25 (2017) 625 638.

[7] T. Smith, R. Purdy, S. Lister, C. Salter, R. Fleetcroft, P.G. Conaghan, Attitudes of people with osteoarthritis towards their conservative management: a systematic review and meta-ethnography, Rheumatol. Int. 34 (2014) 299 313.

[8] A.M. Briggs, M.J. Cross, D.G. Hoy, L. Sànchez-Riera, F.M. Blyth, A.D. Woolf, et al., Musculoskeletal health conditions represent a global threat to healthy aging: a report for the 2015 world health organization world report on ageing and health, Gerontol. 56 (2016) S243 S255.

[9] A.M. Briggs, J.E. Jordan, M. Jennings, R. Speerin, P. Bragge, J. Chua, et al., Supporting the evaluation and implementation of musculoskeletal models of care: a globally informed framework for judging readiness and success, Arthritis Care Res. 69 (2017) 567 577 (Hoboken).

[10] N. Tanios, M. Wagner, M. Tony, R. Baltussen, J. van Til, D. Rindress, et al., Which criteria are considered in healthcare decisions? Insights from an international survey of policy and clinical decision makers, Int. J. Technol. Assess. Health Care 29 (2013) 456 465.

[11] E.M. Selten, J.E. Vriezekolk, R. Geenen, W.H. van der Laan, R.G. van der Meulen-Dilling, M.W. Nijhof, et al., Reasons for treatment choices in knee and hip osteoarthritis: a qualitative study, Arthritis Care Res. 68 (9) (2016) 1260 1267, https://doi.org/10.1002/acr.22841 (Hoboken).

[12] M. Potter, S. Gordon, P. Hamer, The nominal group technique: a useful consensus methodology in physiotherapy research, N. Z. J. Physiother. 32 (2004) 126.

[13] S.S. McMillan, F. Kelly, A. Sav, E. Kendall, M.A. King, J.A. Whitty, et al., Using the Nominal Group Technique: how to analyse across multiple groups, Health Serv. Outcome Res. Methodol. 14 (2014) 92 108.

[14] A. Tong, P. Sainsbury, J. Craig, Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups, Int. J. Qual. Health Care 19 (2007) 349 357.

[15] S. Jünger, S.A. Payne, J. Brine, L. Radbruch, S.G. Brearley, Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review, Palliat. Med. 31 (2017) 684 706.

[16] I.R. Diamond, R.C. Grant, B.M. Feldman, P.B. Pencharz, S.C. Ling, A.M. Moore, et al., Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies, J. Clin. Epidemiol. 67 (2014) 401 409.

[17] A.M. Briggs, J.E. Jordan, R. Speerin, M. Jennings, P. Bragge, J. Chua, et al., Models of care for musculoskeletal health: a cross-sectional qualitative study of Australian stakeholders' perspectives on relevance and standardised evaluation, BMC Health Serv. Res. 15 (2015) 1 13.

[18] A. Loyola-Sanchez, K. Hurd, C. Barnabe, Healthcare utilization for arthritis by indigenous populations of Australia, Canada, New Zealand, and the United States: a systematic review, Semin. Arthritis Rheum. 46 (2017) 665 674.

[19] J. Kitzinger, Introducing focus groups, BMJ Br. Med. J. (Clin. Res. Ed.) 311 (1995) 299 302.

[20] K. Marsh, T. Lanitis, D. Neasham, P. Orfanos, J. Caro, Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature, Pharmacoeconomics 32 (2014) 345 365.

[21] D.L. Morgan, Paradigms lost and pragmatism regained: methodological implications of combining qualitative and quantitative methods, J. Mix. Methods Res. 1 (2007) 48 76.

[22] D.R. Thomas, A general inductive approach for analyzing qualitative evaluation data, Am. J. Eval. 27 (2006) 237 246.

[23] D.L. Sackett, W.M. Rosenberg, J.M. Gray, R.B. Haynes, W.S. Richardson, Evidence based medicine: what it is and what it isn't, BMJ 312 (1996) 71 72.

[24] J.A.M. Gray, Evidence based policy making, BMJ 329 (2004) 988.

[25] J.F.P. Bridges, A.B. Hauber, D. Marshall, A. Lloyd, L.A. Prosser, D.A. Regier, et al., Conjoint analysis applications in health a checklist: a report of the ISPOR good research practices for conjoint analysis task force, Value Health 14 (2011) 403 413.

[26] K. Marsh, M. IJzerman, P. Thokala, R. Baltussen, M. Boysen, Z. Kalo, et al., Multiple criteria decision analysis for health care decision making-emerging good practices: report 2 of the ISPOR MCDA emerging good practices task force, Value Health 19 (2016) 125 137.

[27] R.C. Schmidt, Managing Delphi surveys using nonparametric statistical techniques*, Decis. Sci. J. 28 (1997) 763 774.

[28] L.J. Damschroder, D.C. Aron, R.E. Keith, S.R. Kirsh, J.A. Alexander, J.C. Lowery, Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science, Implement. Sci. 4 (2009).

[29] C. MacKay, P. Veinot, E.M. Badley, Characteristics of evolving models of care for arthritis: a key informant study, BMC Health Serv. Res. 8 (2008) 147.

[30] P. Thokala, N. Devlin, K. Marsh, R. Baltussen, M. Boysen, Z. Kalo, et al., Multiple criteria decision analysis for health care decision making-an introduction: report 1 of the ISPOR MCDA emerging good practices task force, Value Health 19 (2016) 1 13.

[31] GRADE handbook. In: Schünemann H, Brożek J, Guyatt G, Oxman A Eds. Handbook for Grading the Quality and the Strength of Recommendations Using the GRADE Approach, retrieved 26 July, 2018, from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

[32] A.M. Briggs, M. Chan, H. Slater, Models of Care for musculoskeletal health: moving towards meaningful implementation and evaluation across conditions and care settings, Best Pract. Res. Clin. Rheumatol. 30 (2016) 359 374.

[33] J. Baldwin, A.M. Briggs, W. Bagg, P. Larmer, An osteoarthritis model of care should be a national priority for New Zealand, N. Z. Med. J. 30 (2017) 78 86.

[34] Ministry of Health. The Mobility Action Programme. Retrieved 26 November 2019, from https://www.health.govt.nz/our-work/preventative-health-wellness/mobility-action-programme.

[35] Ministry of Health, The Guide to He Korowai Oranga: Maori Health Strategy 2014, Ministry of Health, Wellington, 2014.

[36] National Clinical Guideline Centre, National Institute for health and clinical excellence: guidance, in: Osteoarthritis: Care and Management in Adults London, National Institute for Health and Care Excellence, UK, 2014.

SUPPLEMENT 2

COREQ-32 Checklist

Topic	Item No.	Guide Questions/Description	Reported on Page No.
Domain 1: Research team and reflexivity			
Personal characteristics			
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	133
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	NA
Occupation	3	What was their occupation at the time of the study?	NA
Gender	4	Was the researcher male or female?	NA
Experience and training	5	What experience or training did the researcher have?	133
Relationship with participants			
Relationship established	6	Was a relationship established prior to study commencement?	128
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	128
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	133
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	133
Participant selection			
Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	128
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	128
Sample size	12	How many participants were in the study?	136-137
Non-participation	13	How many people refused to participate or dropped out? Reasons?	138-39
Setting			
Setting of data collection	14	Where was the data collected? e.g., home, clinic, workplace	N/A
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	N/A

Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	Table 9
Data collection			
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	131
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	N/A
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	131
Field notes	20	Were field notes made during and/or after the interview or focus group?	131
Duration	21	What was the duration of the inter views or focus group?	136
Data saturation	22	Was data saturation discussed?	N/A
Transcripts returned	23	Were transcripts returned to participants for comment and/or correction?	N/A
Domain 3: analysis and findings			
Data analysis			
Number of data coders	24	How many data coders coded the data?	134
Description of the coding tree	25	Did authors provide a description of the coding tree?	N/A
Derivation of themes	26	Were themes identified in advance or derived from the data?	133-34
Software	27	What software, if applicable, was used to manage the data?	133
Participant checking	28	Did participants provide feedback on the findings?	135
Reporting			
Quotations presented	29	Were participant quotations presented to illustrate the themes /findings? Was each quotation identified? e.g. participant number	S3 & S4
Data and findings consistent	30	Was there consistency between the data presented and the findings?	Figure 13, Table 10
Clarity of major themes	31	Were major themes clearly presented in the findings?	Figure 13, Table 10
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	S1, Table S1-S2

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

SUPPLEMENT 3

Steps of the Nominal Group Technique

Introduction and explanation (step 1, 5 minutes)

The purpose and process of the meeting is explained to the participants by the facilitator. This step minimises ambiguity or misinterpretation of the study question and/or process.

Silent generation of ideas (step 2, 15 minutes)

Silent and independent generation of ideas. Participants list as many factors they can think of during this step in the context of choosing or recommending interventions for hip or knee osteoarthritis (OA). Discussion is discouraged during this step. When all participants finish generating ideas, this signals the end of step 2.

Sharing ideas (step 3, 20 minutes)

Participants share their ideas one at a time in a round-robin fashion, until no further factors are generated by the group. Group discussion is limited to allow the full presentation of individual participants' factor(s). Factors were transcribed to a board, where similar ideas were grouped by the facilitator, in preparation for Step 4.

Group discussion (step 4, 40 minutes)

Participants are encouraged by the facilitator to review, clarify, consider, add or remove factors shared during step 3. During this process, participants are encouraged to add new factors not considered in Step 1 and cluster similar factors together, according to the consensus of the group. The output of this step results in similar factors grouped together and given an overarching name (theme). All viewpoints are considered and dominant personalities minimised by the facilitator.

Voting and ranking (step 5, 20 minutes).

Participants rank order their top 7 themes from most important to least important. The individual results are collated and then reverse scored, where 7 points are allocated to the 1st ranked item, through 1 point to the 7th ranked item. An excel spreadsheet was used to determine the overall score of each item after it was ranked. This information was shared with the group, signifying the end of the discussion. However, if there was unclear consensus after the first round of voting and ranking, as signalled by two or more top 7 themes with equal rankings, then a final session of discussion followed by voting and ranking was conducted to reach consensus.

The following text provided context to the question for those who completed the modified NGT:

Choosing between different types of treatment options for osteoarthritis involves making trade-offs between factors which are important to you. For example, if you were choosing between cars, the factors you might consider are: colour, size, fuel consumption, engine size, comfort, brand desirability, status and so forth.

What are the factors you use to choose one treatment option over another, for osteoarthritis? This includes any factors you consider at any point in the course of the disease, i.e. the early/mild, mid/moderate and late/severe stages of the condition; when considering any type of treatment option.

Factors can be described as treatment attributes, considerations or differences between any number of treatments options (in the car example above, considerations or differences include colour, size, fuel consumption, cost, engine size, comfort, brand desirability and status).

Example data collection sheet

QUESTION (consumers): What factors do you use to choose one treatment option over another, for your osteoarthritis?

PART 1 INSTRUCTIONS

1. Silently white down as many factors as you can think of without speaking to anybody.
2. Use only one post-it note per factor.
3. Place all the post-it notes on your table in a grid, so you can refer to them later.
4. Include a description of the idea you have come up with, if you want to.
5. Once you have finished, please place your pen on the table.

PART 2 USE ONLY - INSTRUCTIONS

Rank the most important factors to you, from most important (1) to least important (7):

1. (Most important)
2.
3.
4.
5.
6.
7. (Least important)

PART 3 USE ONLY - INSTRUCTIONS

Rank the most important factors to you, from most important (1) to least important (7):

1. (Most important)
2.
3.
4.
5.
6.
7. (Least important)

Table S1. Summary of outputs from the Nominal Group Technique by group

Group	**Theme** [factors]	Importance ranking (lower is better)
Consumers	Wanting to keep mobile. [no factors]	1
Consumers	**What your health professional tells you.** **-** Recommendation from PT, DR, Surgeons, feedback professionals, arthritis NZ info, pharmacist recommendations, Surgeon recommendations, drs advice	2
Consumers	**What are my current symptoms?** - pain/discomfort, ease of pain, pain threshold, - physical difficulty, - disturbed sleep, - ease of movement, - comfort	3
Consumers	**Medication issues.** - overall health considerations - wariness of medication dependence - non-pharmaceutical options preferred/ avoid medication dependency - ease of taking medication e.g. Not big tablets, drugless, pharmacy medication - personal satisfaction at making progress with treatment. Doesn't seem to work/try something else	4
Consumers	**Doing own research/ being informed.** **-** knowledge of potential outcomes - own research of treatment options, access research, reading information, google - whether supported by recognised research, articles online	5
Consumers	Having accurate diagnosis.	6
Consumers	Weighing up options.	7
Consumers	What do I have to do or want to do in my day.	
Consumers	**Consider or look at anything.** **- g**ive anything a try, willingness to try anything that may work - regular medication	
Consumers	**What other people tell you.** **-** friends/family names of drugs they use for pain etc. relating their own relevant experiences - recommendations from people with similar conditions, comments from others, other peoples comments on your walking	
Consumers	How I'm feeling.	
Consumers	**When to opt for surgery.** - desire to be restored to prior physical fitness and capability - resort to surgery or operation as last resort - immediacy, want it fixed	

Group	**Theme** [factors]	Importance ranking (lower is better)
Consumers	**Advertising.** - confusion over advertising - Facebook currently recruiting for a new study with treatment for OA - Facebook home and anecdotal remedies - information from adverts on tv, paper, supermarket, other advertisements	
Consumers	**Cost considerations.** - availability - cost involved - cost of public or private and if surgery is required	
Consumers	**Response to treatment.** - can pain be reduced to manageable level by drugs or physio? - results - exercise (advised by GP) sensible eg. Gardening/walking, - physiotherapist recommended exercise, weight loss reducing load on joints - keeping moving exercise most important - exercise to reduce difficulties - diet to help your condition e.g. celery seeds/fish/caua	
Health Policy-makers	**Evidence.** - evidence based - evidence of effectiveness - patient safety - measure of life adjusted years - benefits vs risks - cost of recovery time to patient - avoiding surgery - early intervention - least intrusive/invasive - previous treatment tried - risks of intervention outcomes to patients	1
Health Policy-makers	**Costs and benefits.** - cost - difference lifetime cost to system & patients lifetime costs. E.g. value for money - societal cost - cost-effectiveness of treatment options - interaction between patient, family and services - time to recovery or response to treatment	2
Health Policy-makers	**The burden of disease.** - Is there a burning platform or public support for OA? - knowledge of chronic long term nature of disease from patient - disease severity - impact on function of the disease - patient co-morbidities	3

Group	**Theme** [factors]	Importance ranking (lower is better)
Health Policy-makers	**System feasibility/ capacity/ resources.** **-** health literacy capacity within the health system to support the patient - what treatment options are fully/privately/publically funded - equity of access to services or outcomes - what funding is available?	4
Health Policy-makers	**Patient factors.** **-** patient preference, patient expectations - social factors - ability to carry work - ability to carry out daily activities - how much is weight/pain/stiffness in influencing function - investment/expectations by family and the patient (i.e. co-pay) - health literacy and education on treatment, level of patient knowledge - patient employment factors - time to treatment (from the patient perspective) - ability to receive and conform to treatment - compliance with treatment - patient satisfaction	5
Health Policy-makers	**Provider quality, capacity and capability.** **-** Ability to communicate with health professionals and people - treatment/patient plan - individualised - waiting for particular treatment - accessibility and availability - looking at the suite of options available - e.g. MDT approach	6
Health care providers (1/2)	**clinical severity and diagnosis.** **-** disease progression and whether it's early or impairment late stage - pain -night pain/24hr history of pain -area of impairment - e.g. location & tissue involvement? - treatment efficacy - functional disability - diagnostic difficulties - either patient's or provider's using x-ray as support which, may or may not be helpful - patient's level of perceived severity - patient's movement patterns and limitations	1
Health care providers (1/2)	**Subjective measures and evidence available.** - Quality of life questionnaire/ impact on ADLs - QoL depression scales (e.g. DASS), - psychological component of a general clinical assessment - clinical diagnosis - radiological evidence - oxford questionnaire	2

Group	**Theme** [factors]	Importance ranking (lower is better)
Health care providers (1/2)	**Patient demographics.** **-** comorbidities - weight - age - patient and health provider perception of importance of treating OA against other comorbidities/conditions - occupational type - client education of OA in the past - patient's age and stage - patient preferences - patients goals and motivation - lifestyle applicability - how likely to get traction	3
Health care providers (1/2)	**Patient preferences**. - Patient factors, - exercise preference, - convenience of treatment to patient	4
Health care providers (1/2)	**Efficacy and safety of treatment.** **-** Treatment safety, - trial pharmaceutical interventions, - the toxic nature of existing pharmacology - limited efficacy	5
Health care providers (1/2)	**Objective measures.** **-** Biomechanics, restricted ROM, weight bearing tolerance/gait, objective assessment - ROM and quads, functional tests	6
Health care providers (1/2)	**access to services.** **-** options of treatment available	7
Health care providers (1/2)	**Prior management.** - Success of otherwise of any previous treatment	
Health care providers (1/2)	**Patient pathways or clinical pathways.** - Availability of patient pathways - e.g. variability within DHBs, pathway for presentation.	
Health care providers (1/2)	**Provider preferences.** - Own previous experience with treatment (therapist factors)	
Health care providers (2/2)	**Patient characteristics of the arthritis**. - Attitude to pain and what is means, e.g. pain = bad - Compliance issues. E.g. patient's ability to take medication as prescribed; willingness to engage in non-pharmaceutical plans e.g. exercise prescriptions, weight loss; motivation of the patient and ongoing support they have - Patient concerns - Patient preferences. E.g. acceptance of medication, religion (i.e. Jehovah Witnesses with their attitudes to blood products); patient experience, expertise or interest in other treatments (Physio, podiatry, reiki etc.)	1

Group	**Theme** [factors]	Importance ranking (lower is better)
	- The outcome the patient has set for themselves. E.g. activity level expectation - Severity of pain; e.g. night time vs day, mild that can be managed with distraction; pain with activity, at rest or at night; activity restriction levels - impact on daily activities - Impact of non-work quality of life activities - Impact/effect on ability to work - Level of function. E.g. degree/amount of functional limitation on walking distance, stairs , getting up and down from chair - Functional demands - Current quality of life - Effect on other joints; generalised vs local - Walking aids - Analgesia requirements	
Health care providers (2/2)	**Patient characteristics of the co-morbidities**. - Drug/other allergies - Patient age and life expectancy - Home risk factors for falls - Patient co-morbidities that may limit therapeutic options or exacerbates the problem. E.g. stomach ulcers, obesity, social isolation - Severity of the diseases - Yellow flags that may hinder rehabilitation and treatment - General medical condition of patient - Medication used for other conditions - Other disabilities that may compound the overall disability. I.e. mental status, deafness, poor vision	2
Health care providers (2/2)	**Place in their family and wider community- support characteristics.** - Ethnicity as it effects outcome of interventions - Disability especially mobility within the home and to other areas of town - Need to drive - Family concerns and dynamics - Financial and work status - Patient responsibilities i.e. for the care of other people	3
Health care providers (2/2)	**Treatment characteristics.** **-** How effective treatment is - How tolerated treatment is - How quality of life would be altered (for better if operation goes well, or for worse if doesn't go well) when thinking of referral for surgery - Risks of intervention. E.g. potential for addiction if opioid-like meds use e.g. codeine - Treatment research - Cost: E.g. whether free (DHB funded) physio is readily available, access to gym facilities at cost person can manage or free (e.g. PHO provided) - Cost of treatment. E.g. celecoxib is unfunded; public versus private treatment - Patient access to private healthcare: either self-funded or has insurance	4

Group	**Theme** [factors]	Importance ranking (lower is better)
Health care providers (2/2)	**Potential flags for poor outcomes.** **-** Previous treatment or surgery - What treatments the patient has tried in the past - Past history – e.g. chronic pain, addiction – intervention history	5
Health care providers (2/2)	**Objective measure of the disease severity.** **-** X-ray, CT & MRI findings - Inflammatory vs degenerative arthritis - Stage/diagnosis of OA i.e. early/mild or mod etc. - Clinical examination. E.g. amount of deformity, limb alignment, ligament status, range of motion, instability	6
Health care providers (2/2)	**Access.** - Access. E.g. access to gym facilities - Ability to get to appointments - Transport - Special skills of treatment providers. I.e. they may have a better success rate than their peers. - District Health Board entry criteria for possible treatments - patient support systems - availability of non-surgical treatment options - geographical location and intervention rates of joint replacements - availability of private/public surgery	7
Content area experts	**Intervention/treatment factors** (safety): - side effects, treatment harms, after effects.	Did not rank themes
Content area experts – subthemes in coloured text	**Intervention/treatment factors** (treatment benefits - short and long term): - efficacy: 'how well it works in a scientific environment' - Evidence of effectiveness: how well it works in the clinical setting - risk-benefit ratio, evidence about the treatment benefits and harms including evidence about whether benefits persist over the longer term - evidence for the appropriate treatment of the condition. E.g. cardiovascular exercise, joint specific strengthening and neuromuscular exercise, education & advice (including weight loss, if appropriate).	
Content area experts	**Intervention/treatment factors** (comorbidities and other disease related): - how comorbidities influence patient and system prioritisation - presence of any comorbidities that may mean some treatments are not appropriate for an individual (e.g. consideration of analgesia and NSAIDs and surgery) - multiple medical factors influencing treatments offered, e.g. BMI, activity level, patient age, comorbidities, balance, presence of low mood or depression, overweight, obesity, mental wellbeing.	
Content area experts	**Intervention/treatment factors** (likelihood of concordance to treatment): - practicability, e.g. ease of use and convenience - capacity to commit to time required for treatment (e.g. exercise) - patient lifestyle	

Group	**Theme** [factors]	Importance ranking (lower is better)
	- sustainability of treatment to patient - type of exercise (e.g. home based vs supervised; land based vs water based) - consumer health priorities, goals and concerns, e.g. other chronic condition self-management or carer role - treatment intensity/acceptability (the patients goals and main problems as articulated by them, ability to continue working) - quality of life in terms of goals for work/recreation	
Content area experts	**Self management** (consumer health literacy): - level of education - English as a second language - culturally and linguistically diverse communities	
Content area experts	**Self management** (Patient concordance factors): - a person's preferences and goals (and the alignment of goals with treatment) - patient adherence to treatment - patient capacity, confidence and priorisation of goals in relation to other conditions to adhere to treatment recommendations /participate in treatment - being a carer - ability to take treatment	
Content area experts	**Self management** (factors that promote long-term self-management): - active versus passive treatment - education and education about the disease and available treatment - treatments that allow a patient to self-manage their OA over the long term rather than be a passive recipient of therapies 'applied to them'. - intervention factors: people's attitudes	
Content area experts	**Patient clinical factors** (disease factors): - location of OA - inflammation - radiological features - radiological severity - symptom severity - rate of progression - pattern of symptom development	
Content area experts	**Patient clinical factors** (Function and level of disability status): - fatigue - functional limitations and limitations in movement - muscle weakness - relationship between function, personal activities of daily living, occupational capacity, sporting capacity - sleep - sexual relations	
Content area experts	**Patient clinical factors** (Pain): - level of pain - pain severity/irritability to guide the type of treatment program and dosage prescribed - chronic pain vs acute flare	

Group	**Theme** [factors]	Importance ranking (lower is better)
	- location or disruption of pain - what's an acceptable level of pain - presence of neuropathic pain - other pain e.g. low back pain	
Content area experts	**Patient clinical factors** (Psychological factors): - expectations - anxiety and depression - patient expectations and experience	
Content area experts	**Patient clinical factors** (Patient's treatment history): - patient's previous experience with different treatments and which have been more/less effective - including an assessment of those which have been tried but perhaps not appropriately and thus warrant another go - prior use of known effective therapies - contributing factors to the patient's main problem and which are amenable to treatment and will have the most impact on their personal problems/goals from effective treatment	
Content area experts	**System factors** (access and equity): - feasibility of patient access to treatment - availability of treatment (e.g. bike, treadmill, weights) in patient's own environment and the clinical environment - need for travel - cost of treatment - cost/capacity to pay for treatment or service to patient (e.g. insurance status) - carer support - requirement for referral - e.g. is a GP referral required - depending on health system - geography: metropolitan vs rural, regional and remote - waiting time	
Māori advocates	**Patient factors** - treatment compliance - mental health, cognition, memory, whanau support, age (old vs young), life expectancy - Patient fatigue with health advice influences treatment compliance - routine of the patient and appropriateness of treatment (in relation to their daily activities) - patient ability to cope with treatment side effects (long vs. short term) vs benefits of treatment, mediated by symptoms of the disease - what's worked in the past and what hasn't - patient preferences for treatment. e.g. may not support pharmacological intervention, patient desire to undertake treatment - patient has to want the treatment - physical ability of patient to apply treatment. e.g. finger dexterity to open treatment packaging - existing responsibilities and/or lifestyle requirements - e.g. to the Marae	1

Group	**Theme** [factors]	Importance ranking (lower is better)
Māori advocates	**Culturally appropriate practices** - whanau support (both immediate family and the community) - family support to and of whanau. E.g. a patient's spokesperson (target kids/grandkids to support patient) - whanau is involved throughout the whole process of care. I.e. family should be involved along the way - Ensure there's family education along the way - so they can assist/help the patient, for example if whanau provide carer or family support - providing education to whanau in the home/community - cultural beliefs, i.e. does the patient believe that the treatment will work? - is the treatment/practice culturally appropriate for Māori, e.g. tapu and noa (putting a heat pack on your feet and then using it on your head). - important to share health information with whanau - what to do (e.g. medication & timing) and why (e.g. poor treatment outcomes or risk of harmful effects)	2
Māori advocates	**Treatment factors** - treatment costs, e.g. frequency of treatment costs. Is the treatment subsidised by the government? Will the treatment impact pension? - treatment packaging, e.g. blister packs - potential side-effects (long term vs short term side effects) - delivery method, e.g. oral, intravenous or topical, needling	3
Māori advocates	**Environmental factors** - what is current best practice (health providers' perspective)? (i.e. evidence-based practice) - weather: more aggressive treatment choices are mediated by season (e.g. winter vs summer) - location: where does the patient live? Can they receive the service? Can their progress with treatment be monitored? are they geographically isolated on an island or farm? - does the patient rely on someone coming to them? e.g. allied health services - individualised/personalised care plan influenced by environment	4
Māori advocates	**External recommendations** - Peer influence - health professional reputation, family expectations (patient does not want to be a burden); whanau have a big impact on patient concordance/adoption of treatment requirements - Peer's experience of the treatment choice - the information they provide (e.g. family and friends) - health professional's personal views/experience with the disease	5

Table S2. Subcategory descriptions of the Framework

Category 1 - Characteristics of the interventions: the evidence (and views relating to the characteristics) that inform the decision to choose, adopt or implement an intervention option.
For the second category, the selection of intervention options for OA was based upon the evaluation of various characteristics, or attributes describing the interventions themselves. For consumers, conflicting sources of information about the attributes of the interventions was sometimes confusing, or unhelpful.
Access to treatment for the consumer: The access to intervention was a key consideration of stakeholders, concerning both the ability to access interventions, as well as the suitability of intervention options to the consumer (e.g. distance to nearest provider and wait time), which overlapped with the feasibility to the health system in terms of equity of access (see category 3: *Characteristics of the health system*).
Active versus passive intervention options: Stakeholders considered the extent to which an intervention allow a patient to self-manage their condition over the long term (i.e. active care strategies), versus passive care. The selection was influenced by the patient's capacity to engage in effective self-management, i.e. intrinsic patient factors (health literacy, locus of control, and patient concordant factors).
Cost of the intervention: Stakeholders considered the financial costs relevant to the use or provision of health care, and the societal cost for osteoarthritis. For example, out of pocket expenses and, the investment, was an important for consumers. Cost considerations also included the cost of intervention (e.g. private vs. public health care), ongoing intervention cost (e.g. ongoing pharmaceutical cost), cost of travel, cost of recovery time and the investment cost acceptable to the consumer and/or family. Societal costs concerned the cost effectiveness of a whole service, and the benefits for societal return (e.g. increased workforce capacity).
Duration of the intervention effect: Considered improvements in physical function and quality of life and the ability to avoid more invasive interventions later. Surgery, for example, was considered to provide an immediate therapeutic effect, albeit with a higher risk profile, while exercise and weight loss took more time to deliver an effect.
Immediacy of the intervention effect: The length of time a intervention option had to be sustained, or the time delay, between starting and experiencing the benefits of intervention was also influenced intervention decisions. Surgery, for example, was considered rapid by some consumers, while drug therapy required ongoing maintenance.
Evidence about the intervention appropriateness: Providing the right interventions or services for OA at the right time was considered. For example, stakeholders considered the appropriateness of providing a interventions or services at the early or advanced stage of OA, such as in the case of providing joint replacement surgery.
Evidence about the intervention effect: Evidence for effectiveness of intervention was also considered on the basis of evidence outside a clinical trial (i.e. real-world impact versus clinical trial), including short-term and long-term evidence of effectiveness and change in health system and societal costs.

Evidence about the intervention harms: The risk of short- and long-term side effects of intervention was considered alongside the benefits of a intervention (e.g. risk of addiction versus pain reduction) and considered to be mediated by a patient's preference for the type and level of risk acceptable to them. This was particularly relevant for pharmacologic (e.g. safety, toxicity, benefit, dependency) and surgical interventions, and less so for 'conservative' care options.
Quality of the evidence: Concerned the extent to which one could be confident that the effects of the intervention or service described were real. Consumers, for example, weighed up the trustworthiness of different sources of information ranging from magazine advertisements to more scientific or peer-reviewed sources of information.
Views relating to the characteristics of interventions: The characteristics of the intervention options could mean different things to different people. The opinions, from different sources, about the features of intervention options that influence intervention selection. For example, peer or family advice and/or personal accounts were viewed as more trustworthy than health professionals' or advertisement (e.g. newspaper, Facebook) advice, claims or recommendations about the benefits or harms of the intervention. Consumers noted they experienced many sources of information that often overstated their effectiveness (e.g. return to previous quality of life), with insufficient information to support their claimed effectiveness.
Category 2 - Characteristics of the consumer: matching the circumstances of consumers (including disease status and comorbidities) to the characteristics of intervention options.
For the first key category, stakeholders considered interventions that would address factors across the biopsychosocial domains important. In this context, the selection of the intervention needed to be matched with the biopsychosocial profile relevant to the consumer, inclusive of their comorbidities.
Clinical status of individual's osteoarthritis. The recommendation or selection of intervention options was based on objective tests (e.g. x-ray or other diagnostics), what the consumer says (subjective reporting relating to symptoms and disability), and age. Consumers also considered a diagnosis of OA an important factor informing their intervention selection.
Presence of other conditions. Intervention choice was influenced by other conditions or diseases (e.g. diabetes, heart disease, mental health), the medical care being received for these conditions and, previous intervention history. Interventions for OA that targeted more than one disease or condition were also favoured by osteoarthritis experts.
Consumer attitudes and beliefs. Intervention options need to reflect a consumer's (and their family's) beliefs about OA, as well as beliefs or expectations about the intervention options. Attitudes incorporate beliefs and knowledge about OA, willingness to pay for intervention, willingness to comply with intervention and, religious and cultural beliefs.
Capacity to self-manage: Consumer's health literacy, attitudes and beliefs about OA (e.g. perception of pain), and capacity to self-manage and navigate the health system influenced the selection of intervention options. These factors were relevant for the selection of intervention options promoting the long-term self-management of OA, including the ability for consumers to concord to recommendations for intervention, or requirements of the intervention(s).

Intervention goals: Intervention options need to relate to the immediate and longer-term functional goals of the consumer now and into the future. Longer-term goals related to lifestyle choices (e.g. sport, ability to care for grandchildren, family/partner expectations, sleep) and work capacity. How OA was prioritised with other conditions and disability needs (e.g. mobility within the home and to other areas) also influenced the selection of interventions or recommendations.

Desperation (e.g. from pain). Once consumers felt that they had exhausted the therapeutic options immediately available to them to manage the symptoms of OA, they sought to improve their quality of life by trying other interventions or medicines (e.g. herbal remedies) with poor evidence supporting their safety, benefit or interaction with other interventions.

Access to intervention: The consumer's geographic location (i.e. closeness to available intervention providers) and the number of intervention options available (to choose) influenced the choice of intervention. Access to care also considered the consumer's ability to meet the cost and/or time requirements of the intervention (equity of access), including the burden imposed by the intervention selection on the carer or other family members, and the role of the consumer with OA in the broader family.

Affordability for the consumer: The immediate out of pocket expenses to the consumer was an important consideration in the selection or recommendation of an intervention. Cost considerations included the cost of intervention (e.g. private vs. public health care), ongoing intervention costs (e.g. ongoing drug cost), the cost of travel, the cost of recovery time and, the investment cost acceptable to the consumer and/or family.

Social support factors: Consideration of consumers' place in the community and wider community and support requirements if a disability is present. For Māori, recognising the role of a spokesperson (who acts as a conduit between the patient and health care provider); involving and understanding Whānau (family) expectations along the health care journey; and understanding patient responsibilities to their community (e.g. carer role or marae responsibilities) strongly influenced the feasibility of an intervention.

Feasibility to the consumer: Whether or not a intervention option is acceptable and feasible to the consumer to adopt in the short and long term, based on the biopsychosocial circumstances of the consumer, their goals and expectations of intervention, and access to health services.

Category 3: Characteristics of the health system: factors underlying the ability of the New Zealand health system to provide access to interventions, including the priority and feasibility of investing in, adopting and implementing interventions for OA.

The first thematic category considered the factors necessary for health care related to OA (e.g. an intervention, or whole service for OA), to be feasibly adopted and/or implemented by a national or local health system. Key subcategories are described, below.

Political landscape - a compelling problem. Before recommending interventions for OA, policy-makers considered if there was community and political interest in addressing the burden of OA. With this health care resources for OA could be prioritised more than for other diseases or conditions with a lower community/political profile. Key ingredients for change referred to a well-articulated disease profile (disease epidemiology/burden of disease) and, political 'burning platform,' such as local and national policy, advocacy groups and the public health cost of OA.

Costs and benefits to society: Whether the change to the services offered (overall, not just one specific intervention option) would deliver significant value for money, compared with the status quo, was a consideration of health policy-makers. That is, would changes result in returns in non-health areas, such as increased work productivity and decreased sickness benefit costs that might offset the increased cost of providing services.
National health system capacity. Health policy-makers also considered the capacity of a health system to support OA health care including policy direction, funding availability, evidence in support of funding or implementing a new service or intervention (e.g. disability adjusted life years, cost-effectiveness) and workforce capacity and capability (e.g. health care provider literacy about OA).
Local health system capacity. Health policy-makers considered the feasibility for a local health system to provide care services for OA in terms of the costs, communication capacity among care providers, the current availability of services (public and private), access to services and scope for integration with other existing services (e.g. availability of existing patient pathways).
Access and Equity. Consideration of the burden imposed on the consumer to reach an intervention or service for OA was reasonable (e.g. geographical remoteness, fit with lifestyle) and, whether or not it was affordable to those who need it (fairness/equity of access). Services or interventions which received public subsidies affected equity of access (e.g. public vs. private health care). Accessibility also encompassed local health system access criteria (e.g. District Health Board referral requirements for patient pathways to health care), the interventions available for consideration, and the wait time involved in accessing intervention (e.g. surgical wait list).
Culturally appropriate services. Whether or not a health care service for OA was culturally appropriate for diverse communities was considered by stakeholders when choosing or recommending interventions for OA. For Māori, health services which incorporated whanau (family) and community needs (e.g. health education about OA) and culturally sensitive practices, such as respecting Tapu and Noa along the health care journey, were considered important health service characteristics.
Characteristics of the health service provider. The beliefs, experiences and preferences of the health service provider also influenced intervention option selection. In particular, the provider's experience with intervention options and their understanding of evidence-based care for OA.

Table S3. Supporting quotations for the subcategory descriptions of the Framework

Category 1: Characteristics of the consumer “…patient factors. It's a basis, it's your basis to start any treatment. It's your platform. I don’t know what to call it, but it's the very, very beginning of everything. Where are we starting from? How much do we know?” (Health policy #1)
Clinical status of individual’s osteoarthritis "The trouble is there’s the objective parts of what you find out about the patient and then there’s the subjective part that they contribute which they talk about. So, to me those are the two separations and I would see them as being quite important to me. So, your x-rays, your scans, your blood tests and all that, I see them as being objective versus what the patient says and how incapacitated they say they are, all that sort of stuff which a subjective thing is.” (Health care provider #1)
Presence of other conditions "…a couple of things that always worry me are whether they’ve got a chronic pain syndrome and whether they’ve got a chemical addiction already when you’re considering what you’re gonna do. Those two things influence me quite a bit with what I’m going to suggest that they might contemplate doing." (Health care provider #1) “The other thing I thought is, that there is an issue also that may not come up in what you've got; and that is in terms of how comorbidities may influence more than just side-effects or treatment, because they may influence patient factors around prioritisation and complexity of medical management. So, there's a sort of system; there's a patient and system impact. (Osteoarthritis expert #1) “…you want to actually focus on, potentially on treatments that are going to target multiple comorbid conditions, and it might be weight loss, or it might be exercise.” (Osteoarthritis expert #1) “What we’ve tended to find is that you might have someone with OA but they’ve also got heart failure, or chronic airways disease, or diabetes; and they don’t prioritise their OA above their other conditions. They’ll take their cardiac medications, and have their cardiac investigations; but they won’t necessarily go to exercise therapy, or see a nutritionist.” (Osteoarthritis expert #2)
Patient attitudes and beliefs "Desire. Like after all is said and done you can explain the benefits, present evidence based practice, and if they don’t want to take it they just don’t want to take it. They’ve got to want to; they’ve got to want the treatment. And someone might just want to die and not take the treatment; like I’m over this, I’m sick of being old. I have patients say that: “Don’t get old dear, don’t get old…or you might be like the 20th health professional to contact them about this and they just stop…They’re just over it." (Māori advocate & health care provider #1)
Capacity to self-manage “I think health literacy, education and level of patient knowledge all relate to what the patient needs to bring to the treatment or treatment decision.” (Health policy #2)
Treatment goals "I put how I’m feeling; it's time of the day sometimes. Like how I’m feeling; if I’m in a good space or something like that and I’m going somewhere and it's painful or whatever I might decide to not do something. I might decide not to get sort of very heavy about it. Or otherwise I may be having a bad day and I weigh up the pros and cons that way. It’s how I’m feeling on the day." (Consumer #3)
Desperation (e.g. from pain) “Being open for information too rather than just closing your mind off. I think you reach the stage where you will listen to anybody in the hope that something may work.” (Consumer #7)

“The problem is though you see on the advertisements and things like that and these people stand up and say, “This is the best thing since sliced bread,” and you think “Right, okay, I’m going to give this a go.” How much are they being paid to say that and in actual fact have they had that working for them; because you’ve got no way of checking. All it is, is you think “Well if it's worked for that person do I spend the money and give that a go,” but you have no knowledge of it.” (Consumer #9)
Access to treatment “I’m just thinking about [for people who] live rurally all these things are harder. There may not be a Weight Watchers where you live, there won’t be Jenny Craig, you may or may not have a gym you can go to, there may or may not be a physio that the DHB funds so you can get free rehab…it all is affected by what’s available.” (Health care provider #2) “Yeah so it's not so much about equity; it's about would somebody have to travel a big distance and that would be difficult for them. It's where the costs fall for the treatment option and which of them are most accessible for that person.” (Health policy #2) “…And the other thing is, being a Carer. So, if you’ve got someone who’s actually a carer for someone else, that often influences their ability to take up [the intervention].” (Osteoarthritis expert #2)
Affordability for the consumer “The other thing too perhaps and I don’t know quite how you put it in, but I’m still part time working and so my choices up to a point is how it can still allow me to work…Because if something’s not going to allow me to continue what I’m doing then I wouldn’t be interested in it?” (Consumer #10) “I should think recovery actually comes into that, because some of these treatments for example if you have a wedge, or whatever, say of your knee, the recovery is about 12 weeks; and if you’ve got someone in employment and they don’t have insurance it's not an option for them. So…for what we’re talking about here; that recovery time actually has a bit impact on what treatment you would offer.” (Health policy #3) …”and I suppose it links in with employment…and ADLs at home, your ability…” (Health Policy #4) “It's also the cost of recovery time; so it's recovery time and what the costs are of that, which might be the ability to take up your special role or it might be a financial cost.“ (Health policy #2)
Social support factors “I would say family support to and of as well. Like for some Māori it depends on their beliefs and their upbringing but some people would choose a spokesperson; and so everything is dealt with that spokesperson that the patient has nominated as the spokesperson. So you wouldn’t necessarily – and this is really weird as a health professional that you talk to someone other than the patient, ‘cause it's really the patient that you’re dealing with and their experience. But to some Māori they do choose a spokesperson and so everything gets [9.10]. So let's say it's an older person; they might choose their youngest son, or they might choose the youngest daughter and that’s the person that you deal with, with all cares concerning or all treatments concerning the patient. So you actually don’t really deal with patient; which is different." (Māori advocate #1)
Feasibility to the consumer “For example if you’ve got someone living in a rural area and you’re saying to them you’re going to need week to week physio for 10 or 12 weeks and they live in a rural area that might not be an option for them. It might not be feasible…I think that’s the difference of the feasibility of the overall suite of interventions for the system to live as opposed to what is feasible for that person.“ (Health policy #5)
Category 2: Characteristics of the interventions "I think a big way to cover it is to be well informed. You’ve got this condition, it's been around for centuries no doubt. People must have done studies on it. Perhaps the Arthritis Foundation are a good

place to go and say, “Okay I’ve got this complaint, what have people tried and what’s worked. What should I do?” and if you don’t like taking pills or prescriptions acupuncture, physio and try and go down that track and find out what you can do. So it's all about being informed.” (Consumer #4)
Access to treatment for the consumer “Availability; for example, I explored physiotherapy and exercise options but I live an hour or so out of town and I would have possibly taken on further if it had been easier and I’d got feedback and things on it. There were other factors too but availability and of course the same as the true with the operations, is the big thing, isn’t it; is the availability… to access services whether it be the public system, whether it be close to; in my case travel’s a factor. Things like physiotherapy or people go to the physio pool here; those sorts of things.” (Consumer #6) “I do actually think this equity thing actually is rearing its head more and more [1.10.02]. It does come from that commitment from the patient as well, because any of these treatments involve a commitment from a patient but there is an issue that some will be able to commit more than others. You could offer a 65 year old or a 62 year old a hip replacement but they simply might not be able to accept it, even though they need it, because of the recovery and the impact for them. But if you’ve got someone who has much better access to holiday pay and sick pay. (Health Policy #1) “on the other hand you might also have somebody who’s reasonably well off who gets offered a hip replacement and they have the ability to take it but actually they have a much better investment in making what they got… you know make their hips last a lot longer; so they have more of an ability to get stuck into the exercise or lose their weight or whatever. I think that investment thing is a very big area. We know people like that; because they had health insurance the first thing they were offered was surgery where really that’s probably what they need in ten years’ time and not now”. (Health policy #1)
Active versus passive intervention options Capacity and confidence and motivation might be different; I might put that as a different ‘cause somebody might have the capacity and confidence, but they’re not motivated to change … if you look at the goal-setting care plans, the way they’re done, those sorts of things [factors related to self-management] would definitely be taken into consideration. I think it’s more and more the way in which health professionals are approaching the way they develop their care plans with patients.” (Osteoarthritis expert #1)
Cost of the intervention “One of the things which I do think is an influence is the cost involved. Okay, while you’re prepared to spend there is a limit to which you can do it and you think well if this is going to cost you x number of dollars is it going to work?” (Consumer #4) “Cost also as I’ve noted comes to whether it's private or it's you paying yourself or whether you can do it publicly. It's just cost me a bundle of money to do it. I wouldn’t have even got into the public system.”(Consumer #5)
Duration of the intervention effect “I made a note about this just while conversation was going on earlier about surgery and last resorts because from my point of view I would much prefer surgery to a drug medication that goes on and on and on. I would really like the thing fixed and done with and then not continuing with medication; that would be my preference. And so, I’d say medication, like going on pills, would be… the last resort for me rather than surgery.” (Consumer #7)
Immediacy of the intervention effect “People were talking about an operation for example as a last resort; whereas for me I want to get something done before all the other bits fall apart as they are starting to and the deterioration accumulates.

So, I want to step in now and that's why I've chosen to go private because public is a way out of my reach and it's that immediacy which has made me choose that option." (Consumer #7)
Evidence about the intervention appropriateness "What is best practice? … You know, what's the least intrusive and invasive treatment is actually what is the best practice in this situation?" Health Policy Line 126 "I guess in a sense we might say that all the evidence or whatever we might come up with on paper says that this is the right treatment, but actually for certain groups of people in society it might be the wrong treatment because it might be unfeasible because they're rural or can't afford to travel. So therefore it's the wrong treatment for that person or even that group at that time." (Health Policy Maker #2)
Evidence about the intervention effect "Well, efficacy, simply does the drug work; or does the treatment work? And with more complex interventions, say we've got multidisciplinary clinics, or whatever, then you'd need effectiveness which is showing that it works within a context." (Osteoarthritis expert #2)
Evidence about the intervention harms "It doesn't matter whether it's a drug that's actually prescribed by the doctor or this Arthrem or any of those other things; you are putting something into your body and things like that. Is it going to be worth it? Is it not? You're weighing this up all the time." (Consumer #4) "You've got to be very careful with surgery when you're considering a surgical operation because you could finish up in a worse state." (Consumer #8) "…some risks are more or less acceptable to some patients; so the risks of interventions is different from the potential benefits; it's actually the nature of the risks maybe more or less acceptable to certain people; so a risk of catastrophic outcome as opposed to no improvement would be more or less acceptable to different folks." (Health policy #2)
Quality of the evidence and views relating to the characteristics of interventions "…I've got supermarket medication. I don't know whether it works but it's all sort of natural stuff and if it doesn't work well it doesn't do you any harm; that's the way I look at it. But I would like more scientific background to that type of stuff because they're making a considerable amount of money out of it." (Consumer #1) "The problem is though you see on the advertisements and things like that and these people stand up and say, "This is the best thing since sliced bread," and you think "Right, okay, I'm going to give this a go." How much are they being paid to say that and in actual fact have they had that working for them; because you've got no way of checking. All it is, is you think "Well if it's worked for that person do I spend the money and give that a go," but you have no knowledge of it." (Consumer #9) "If it works you're likely to give it a go from someone else's experience or you're not. If I had a patient that was chartered Tramadol and I was about to give it but the daughter was there, she goes, "No dad don't take that, that doesn't work it made me violently ill." And so I have to then step in and say, "Yes it made you violently ill but you are not your father, you don't know how this treatment is going to go for him if he's never had it." So had to give more information to the dad. Actually he ended up taking the Tramadol and it actually worked for him. So had he listened to his daughter who's had ill experiences with it, there's a treatment that he would have just negated, hit it on the head, purely because of somebody else's opinion or experience." (Māori advocate #2)

"I get confused over the amount of advertising on products and I wonder whether they really work.“ (Consumer #1)
Category 3: Characteristics of the health system “There’s a very basic policy parameter that’s often called feasibility which is the extent to which the system can do something. So if you’re thinking about screening programmes, introducing a new screening programme or changing it, and think about the debates for example about bowel screening, one of the really basic policy criteria is system feasibility.” (Health policy #1)
Political landscape - a compelling problem “We’ve talked about the burden of disease but actually there’s a far more less palatable political reality, which is about is there a burning platform? Is there some advocacy for this? Is this both accepted and publicly accepted as being needed? And it might be quite small. So I wonder if burden of disease is only one part of is there a burning platform or public support for this…Sometimes the burning platform is because of the cost to the system. Sometimes it's because of the consumer voice, from some effective advocates, and that’s all captured in the how is this framed in the public’s mind or little P politically. Because there’s a political element to it which might just be local community; I don’t mean national politics." (Health Policy #2)
Costs and benefits to society “A whole lot around cost; what the particular cost is. The difference to the lifetime cost system and also the patient’s lifetime costs. The value for money. (Health Policy #5) “…under the evidence of effectiveness, you may have it, but it's something about cost effectiveness. So somebody was talking about the cost difference to the person, to the patient, to the health system and to society as a whole; and if those components are taken into account for each of the treatment options then that covers what I mean by cost effectiveness.” (Health Policy #2)
National health system capacity “…I think the question is what resources are available? What funding is available? Because you’re likely to cut your service, cut your coat according to your cloth. You’re likely to come up with a service that is feasible within the funding available and other resources available.” (Health Policy #5) ...What are the components that need to be in place to make this work? So making sure that whatever’s being funded or promoted is actually doable, feasible and their funding for it is incredibly important.” (Health Policy #6)
Local health system capacity "...it's actually I guess in a way saying more about the role of the District Health Board. [A DHB] plan for commissioned services in its own district, right? Once it might be a national policy or capability that we’re going to provide these services in the country and then the DHBs end up providing those services in their own district according to what’s available in their district to do that." (Health Policy #2)
Access and Equity “I guess in a sense we might say that all the evidence or whatever we might come up with on paper says that this is the right treatment, but actually for certain groups of people in society it might be the wrong treatment because it might be unfeasible because they’re rural or can’t afford to travel. So therefore it's the wrong treatment for that person or even that group at that time…” (Health Policy #5) “Availability; for example, I explored physiotherapy and exercise options but I live an hour or so out of town and I would have possibly taken on further if it had been easier and I’d got feedback and things on it. There were other factors too but availability and of course the same as the true with the operations, is the big thing, isn’t it; is the availability” (Consumer #6)

"I think one of the questions under accessibility and availability is what treatment options are fully funded, partly funded or privately funded. So it's kind of where the costs fall and that can have a big impact on what options you would choose…This is a more immediate question that says of the options available are they all equally financially available to this person, or what would be?" (Health Policy #2) "I wonder if accessibility at a system level becomes kind of fairness or equity; that at the same time as you want to know can the system deliver this, can you set up treatment options that will carry through, there's something about access and about fairness of access that you might want to pick up as well. (health policy #1)…Yeah…I just was thinking it's more than access; it's about fairness of access which I call equity, but I mean we might frame it differently." (Health policy #5)
Culturally appropriate services "…educating the family is important also because that patient potentially is living with them, and so they will be offering the cares in some way, shape or form; so having the family, the patient and the whole team, the MDT, all on the same page. Like Māori don't tend to – this could be generalising but in my experience Māori don't put their loved ones into a rest home; so they're at home living with the younger daughter or the niece or somebody. So having those people involved in all aspects is really helpful 'cause that will help the patient overall. " (Māori Advocate #2)
Characteristics of the health service provider "..I don't see the responsibility for health literacy being all on the patient or their family. So one of the aspects that I think is important is how well the local system, so in terms of options for treatment, how well the local system will facilitate self-management in the person. So it's a health literacy capacity within the system that's available and there will be choices in terms of different types of intervention which would facilitate a sense of control and self-management in the person and some of those will work better than others. So it's the enhancing health literacy or enhancing self-management as a choice between systems." (Health policy #2)

SUPPLEMENT 4

Table S4. Level of agreement with the overall Framework by panel

Panel	Level of agreement with the framework (%)					Median*
	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree	
Panel A (n=34)	0	2.9	5.9	50	41.2	4
Panel B (n = 31)	0	0	6.5	45.2	48.4	4

*5= strongly agree, 4= agree, 3= neutral, 2= disagree and 1= strongly disagree

Table S5. Level of agreement with the overall Framework by panel and stakeholder group

	Level of agreement, n (%)					
Panel A (n=34)	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Median*
Consumer	(0)	(0)	1 (11)	3 (33)	5 (56)	5
Health care provider	(0)	(0)	(0)	6 (67)	3 (33)	4
Health policy maker	(0)	(0)	(0)	3 (50)	3 (50)	5
Māori Health advocate	(0)	(0)	1 (20)	2 (40)	2 (40)	4
Osteoarthritis expert	(0)	1 (20)	(0)	3 (60)	1 (20)	4
Panel B (n= 31)						
Consumer	(0)	(0)	2 (29)	3 (43)	2 (29)	4
Health care provider	(0)	(0)	(0)	5 (36)	9 (64)	5
Health policy maker	(0)	(0)	(0)	4 (100)	(0)	4
Māori Health advocate	(0)	(0)	(0)	(0)	(0)	
Osteoarthritis expert	(0)	(0)	(0)	2 (33)	4 (67)	5
Pooled panels A + B (n=65)						
Consumer	(0)	(0)	3 (19)	6 (38)	7 (44)	4
Health care provider	(0)	(0)	(0)	11 (48)	12 (52)	5
Health policy maker	(0)	(0)	(0)	7 (70)	3 (30)	4
Māori advocacy	(0)	(0)	1 (20)	2 (40)	2 (40)	4
Osteoarthritis expert	(0)	1 (9)	(0)	5 (45)	5 (45)	4

*range 1-5, where 5 = strongly agree, 4= agree, 3 = neutral, 2 = disagree and 1 = strongly disagree

Table S6. Level of agreement with the sub-categories of the Framework by individual panels A and B

	Round 1 level of agreement Panel A % (n= 34)				**Round 1 level of agreement Panel B % (n= 31)**			
Category	**Median***	**Completely agree**	**Partially agree**	**Disagree**	**Median***	**Completely agree**	**Partially agree**	**Disagree**
CATEGORY 1. Characteristics of the consumer								
Clinical status of individual's OA	3	73.5	26.5		2	48.4	51.6	
Presence of other conditions	3	67.6	26.5	5.9	3	80.6	16.1	3.2
Patient's attitudes and beliefs	3	88.2	11.8		3	61.3	38.7	
Capacity to self-manage	3	85.3	11.8	2.9	3	83.9	16.1	
Treatment goals	3	91.2	8.8		3	83.9	16.1	
Desperation	2	44.1	44.1	11.8	2	29.0	48.4	22.6
Access to treatment	3	73.5	23.5	2.9	3	71.0	19.4	9.7
Affordability for the consumer	3	82.4	17.6		3	71.0	22.6	6.5
Social support factors	3	55.9	41.2	2.9	3	67.7	22.6	9.7
CATEGORY 2. Characteristics of the interventions								
Access to the treatment	3	73.5	26.5		3	71	25.8	
Active vs passive treatments	3	61.8	35.3	2.9	3	61.3	35.5	3.2
Cost of the treatment	3	70.6	29.4		3	71.0	25.8	3.2
Duration of the treatment effect	3	76.5	20.6	2.9	3	74.2	22.6	3.2
Treatment appropriateness	3	82.4	17.6		3	87.1	12.9	
Treatment effectiveness	3	88.2	11.8		3	90.3	9.7	
Treatment harms	3	91.2	8.8		3	80.6	19.4	
Immediacy of the treatment effect	3	61.8	35.3	2.9	3	64.5	29	6.5
Quality of the evidence	3	73.5	23.5	2.9	3	87.1	12.9	
Views relating to the characteristics of the treatment	3	73.5	14.7	11.8	2	41.9	41.9	16.1

CATEGORY 3. Characteristics of the health system								
Political landscape	3	58.8	32.4	8.8	3	67.7	19.4	12.9
Costs and benefits	3	73.5	26.5		3	83.9	9.7	6.5
National health system capacity	3	73.5	23.5	2.9	3	71.0	16.1	12.9
Local health system capacity	3	61.8	32.4	5.9	3	74.2	12.9	12.9
Access and equity	3	79.4	17.6	2.9	3	77.4	6.5	16.1
Culturally appropriate services	3	73.5	20.6	5.9	3	74.2	16.1	9.7
Characteristics of the health service provider	3	82.4	17.6		3	71.0	16.1	12.9

*Where 3= completely agree, 2= partly agree and 1= disagree

Table S7. Level of agreement with the sub-categories of the Framework by pooled and individual panels, A and B

	Pooled (panels A + B)			Panel A			Panel B		
	Frequency n (%)								
Stakeholder group	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**
Clinical status of individual's OA									
Consumer		2 (13)	14 (88)		1 (11)	8 (89)		1 (14)	6 (86)
Health care provider		11 (48)	12 (52)		3 (33)	6 (67)		8 (57)	6 (43)
Health policy maker		1 (10)	9 (90)			6 (100)		1 (25)	3 (75)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert		9 (82)	2 (18)		3 (60)	2 (40)		6 (100)	
Presence of other conditions									
Consumer	3 (19)	3 (19)	10 (63)	2 (22)	2 (22)	5 (56)	1 (14)	1 (14)	5 (71)
Health care provider		3 (13)	20 (87)		1 (11)	8 (89)		2 (14)	12 (86)
Health policy maker		3 (30)	7 (70)		2 (33)	4 (67)		1 (25)	3 (75)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert		3 (27)	8 (73)		2 (40)	3 (60)		1 (17)	5 (83)
Patient's attitudes and beliefs									
Consumer		2 (13)	14 (88)		1 (11)	8 (89)		1 (14)	6 (86)
Health care provider		8 (35)	15 (65)		1 (11)	8 (89)		7 (50)	7 (50)
Health policy maker		5 (50)	5 (50)		1 (17)	5 (83)		4 (100)	
Māori advocacy			5 (100)			5 (100)			
Osteoarthritis expert		1 (9)	10 (91)		1 (20)	4 (80)			6 (100)

	Pooled (panels A + B)			**Panel A**			**Panel B**		
	Frequency n (%)								
Stakeholder group	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**
Capacity to self-manage									
Consumer		3 (19)	13 (81)		1 (11)	8 (89)		2 (29)	5 (71)
Health care provider	1 (4)	3 (13)	19 (83)	1 (11)	1 (11)	7 (78)		2 (14)	12 (86)
Health policy maker		3 (30)	7 (70)		2 (33)	4 (67)		1 (25)	3 (75)
Māori advocacy			5 (100)			5 (100)			
Osteoarthritis expert			11 (100)			5 (100)			6 (100)
Treatment goals									
Consumer		2 (13)	14 (88)			9 (100)		2 (29)	5 (71)
Health care provider		3 (13)	20 (87)		1 (11)	8 (89)		2 (14)	12 (86)
Health policy maker		1 (10)	9 (90)			6 (100)		1 (25)	3 (75)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert		1 (9)	10 (91)		1 (20)	4 (80)			6 (100)
Desperation									
Consumer	1 (6)	8 (50)	7 (44)	1 (11)	3 (33)	5 (56)		5 (71)	2 (29)
Health care provider	4 (17)	8 (35)	11 (48)	1 (11)	2 (22)	6 (67)	3 (21)	6 (43)	5 (36)
Health policy maker	1 (10)	6 (60)	3 (30)		5 (83)	1 (17)	1 (25)	1 (25)	2 (50)
Māori advocacy		3 (60)	2 (40)		3 (60)	2 (40)			
Osteoarthritis expert	5 (45)	5 (45)	1 (9)	2 (40)	2 (40)	1 (20)	3 (50)	3 (50)	
Access to treatment									
Consumer	1 (6)	6 (38)	9 (56)		4 (44)	5 (56)	1 (14)	2 (29)	4 (57)
Health care provider	2 (9)	4 (17)	17 (74)	1 (11)	2 (22)	6 (67)	1 (7)	2 (14)	11 (79)
Health policy maker	1 (10)	2 (20)	7 (70)		2 (33)	4 (67)	1 (25)		3 (75)

	Pooled (panels A + B)			Panel A			Panel B		
	Frequency n (%)								
Stakeholder group	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**
Māori advocacy			5 (100)			5 (100)			
Osteoarthritis expert		2 (18)	9 (82)			5 (100)		2 (33)	4 (67)
Affordability for the consumer									
Consumer	1 (6)	2 (13)	13 (81)			9 (100)	1 (14)	2 (29)	4 (57)
Health care provider	1 (4)	4 (17)	18 (78)		1 (11)	8 (89)	1 (7)	3 (21)	10 (71)
Health policy maker		4 (40)	6 (60)		2 (33)	4 (67)		2 (50)	2 (50)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert		2 (18)	9 (82)		2 (40)	3 (60)			6 (100)
Social support factors									
Consumer	3 (19)	3 (19)	10 (63)	1 (11)	2 (22)	6 (67)	2 (29)	1 (14)	4 (57)
Health care provider	1 (4)	6 (26)	16 (70)		4 (44)	5 (56)	1 (7)	2 (14)	11 (79)
Health policy maker		6 (60)	4 (40)		3 (50)	3 (50)		3 (75)	1 (25)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert		5 (45)	6 (55)		4 (80)	1 (20)		1 (17)	5 (83)
Access to the treatment									
Consumer	1 (6)	5 (31)	10 (63)		3 (33)	6 (67)	1 (14)	2 (29)	4 (57)
Health care provider		6 (26)	17 (74)		3 (33)	6 (67)		3 (21)	11 (79)
Health policy maker		4 (40)	6 (60)		2 (33)	4 (67)		2 (50)	2 (50)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert		1 (9)	10 (91)			5 (100)		1 (17)	5 (83)

	Pooled (panels A + B)			**Panel A**			**Panel B**		
	Frequency n (%)								
Stakeholder group	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**
Active vs passive treatments									
Consumer	1 (6)	7 (44)	8 (50)	1 (11)	4 (44)	4 (44)		3 (43)	4 (57)
Health care provider		7 (30)	16 (70)		2 (22)	7 (78)		5 (36)	9 (64)
Health policy maker	1 (10)	4 (40)	5 (50)		3 (50)	3 (50)	1 (25)	1 (25)	2 (50)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert		3 (27)	8 (73)		1 (20)	4 (80)		2 (33)	4 (67)
Cost of the treatment									
Consumer		3 (19)	13 (81)		2 (22)	7 (78)		1 (14)	6 (86)
Health care provider	1 (4)	7 (30)	15 (65)		4 (44)	5 (56)	1 (7)	3 (21)	10 (71)
Health policy maker		5 (50)	5 (50)		2 (33)	4 (67)		3 (75)	1 (25)
Māori advocacy			5 (100)			5 (100)			
Osteoarthritis expert		3 (27)	8 (73)		2 (40)	3 (60)		1 (17)	5 (83)
Duration of the treatment effect									
Consumer		2 (13)	14 (88)		1 (11)	8 (89)		1 (14)	6 (86)
Health care provider	2 (9)	5 (22)	16 (70)	1 (11)	1 (11)	7 (78)	1 (7)	4 (29)	9 (64)
Health policy maker		3 (30)	7 (70)		3 (50)	3 (50)			4 (100)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert		3 (27)	8 (73)		1 (20)	4 (80)		2 (33)	4 (67)
Treatment appropriateness									
Consumer		3 (19)	13 (81)		1 (11)	8 (89)		2 (29)	5 (71)

	Pooled (panels A + B)			Panel A			Panel B		
	Frequency n (%)								
Stakeholder group	Disagree	Partially agree	Completely agree	Disagree	Partially agree	Completely agree	Disagree	Partially agree	Completely agree
Health care provider		5 (22)	18 (78)		1 (11)	8 (89)		2 (14)	12 (86)
Health policy maker		1 (10)	9 (90)		1 (17)	5 (83)			4 (100)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert			11 (100)			5 (100)			6 (100)
Treatment effectiveness									
Consumer		2 (13)	14 (88)		1 (11)	8 (89)		1 (14)	6 (86)
Health care provider		3 (13)	20 (87)		3 (33)	6 (67)		2 (14)	12 (86)
Health policy maker		1 (10)	9 (90)		1 (17)	5 (83)			4 (100)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert			11 (100)			5 (100)			6 (100)
Treatment harms									
Consumer		2 (13)	14 (88)		1 (11)	8 (89)		1 (14)	6 (86)
Health care provider		5 (22)	18 (78)		1 (11)	8 (89)		4 (29)	10 (71)
Health policy maker		1 (10)	9 (90)			6 (100)		1 (25)	3 (75)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert			11 (100)			5 (100)			6 (100)
Immediacy of the treatment effect									
Consumer	1 (6)	2 (13)	13 (81)		1 (11)	8 (89)	1 (14)	1 (14)	5 (71)
Health care provider	1 (4)	10 (43)	12 (52)	1 (11)	3 (33)	5 (56)		7 (50)	7 (50)
Health policy maker	1 (10)	4 (40)	5 (50)		4 (67)	2 (33)	1 (25)		3 (75)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			

	Pooled (panels A + B)			**Panel A**			**Panel B**		
	Frequency n (%)								
Stakeholder group	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**
Osteoarthritis expert		4 (36)	7 (64)		3 (60)	2 (40)		1 (17)	5 (83)
Quality of the evidence									
Consumer		1 (6)	15 (94)		1 (11)	8 (89)			7 (100)
Health care provider		5 (22)	18 (78)		2 (22)	7 (78)		3 (21)	11 (79)
Health policy maker		3 (30)	7 (70)		3 (50)	3 (50)			4 (100)
Māori advocacy	1 (20)	2 (40)	2 (40)	1 (20)	2 (40)	2 (40)			
Osteoarthritis expert		1 (9)	10 (91)			5 (100)		1 (17)	5 (83)
Views relating to the characteristics of the treatment									
Consumer	1 (6)	3 (19)	12 (75)		1 (11)	8 (89)	1 (14)	2 (29)	4 (57)
Health care provider	4 (17)	6 (26)	13 (57)	1 (11)	1 (11)	7 (78)	3 (21)	5 (36)	6 (43)
Health policy maker	3 (30)	5 (50)	2 (20)	3 (50)	1 (17)	2 (33)		4 (100)	
Māori advocacy			5 (100)			5 (100)			
Osteoarthritis expert	1 (9)	4 (36)	6 (55)		2 (40)	3 (60)	1 (17)	2 (33)	3 (50)
Political landscape									
Consumer	1 (6)	4 (25)	11 (69)	1 (11)	3 (33)	5 (56)		1 (14)	6 (86)
Health care provider	3 (13)	7 (30)	13 (57)		3 (33)	6 (67)	3 (21)	4 (29)	7 (50)
Health policy maker	1 (10)	2 (20)	7 (70)	1 (17)	2 (33)	3 (50)			4 (100)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert	2 (18)	2 (18)	7 (64)	1 (20)	1 (20)	3 (60)	1 (17)	1 (17)	4 (67)

	Pooled (panels A + B)			**Panel A**			**Panel B**		
				Frequency n (%)					
Stakeholder group	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**	**Disagree**	**Partially agree**	**Completely agree**
Costs and benefits									
Consumer		4 (25)	12 (75)		3 (33)	6 (67)		1 (14)	6 (86)
Health care provider	1 (4)	2 (9)	20 (87)		1 (11)	8 (89)	1 (7)	1 (7)	12 (86)
Health policy maker		2 (20)	8 (80)		1 (17)	5 (83)		1 (25)	3 (75)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert	1 (9)	2 (18)	8 (73)		2 (40)	3 (60)	1 (17)		5 (83)
National health system capacity									
Consumer	1 (6)	4 (25)	11 (69)	1 (11)	2 (22)	6 (67)		2 (29)	5 (71)
Health care provider	3 (13)	3 (13)	17 (74)		1 (11)	8 (89)	3 (21)	2 (14)	9 (64)
Health policy maker		2 (20)	8 (80)		1 (17)	5 (83)		1 (25)	3 (75)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert	1 (9)	2 (18)	8 (73)		2 (40)	3 (60)	1 (17)		5 (83)
Local health system capacity									
Consumer		4 (25)	12 (75)		2 (22)	7 (78)		2 (29)	5 (71)
Health care provider	5 (22)	3 (13)	15 (65)	2 (22)	2 (22)	5 (56)	3 (21)	1 (7)	10 (71)
Health policy maker		4 (40)	6 (60)		3 (50)	3 (50)		1 (25)	3 (75)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert	1 (9)	2 (18)	8 (73)		2 (40)	3 (60)	1 (17)		5 (83)
Access and equity									
Consumer	1 (6)	2 (13)	13 (81)		1 (11)	8 (89)	1 (14)	1 (14)	5 (71)
Health care provider	3 (13)	3 (13)	17 (74)		2 (22)	7 (78)	3 (21)	1 (7)	10 (71)

	Pooled (panels A + B)			Panel A			Panel B		
	Frequency n (%)								
Stakeholder group	Disagree	Partially agree	Completely agree	Disagree	Partially agree	Completely agree	Disagree	Partially agree	Completely agree
Health policy maker	1 (10)		9 (90)	1 (17)		5 (83)			4 (100)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert	1 (9)	1 (9)	9 (82)		1 (20)	4 (80)	1 (17)		5 (83)
Culturally appropriate services									
Consumer	1 (6)	4 (25)	11 (69)	1 (11)	1 (11)	7 (78)		3 (43)	4 (57)
Health care provider	2 (9)	3 (13)	18 (78)		1 (11)	8 (89)	2 (14)	2 (14)	10 (71)
Health policy maker	1 (10)		9 (90)	1 (17)		5 (83)			4 (100)
Māori advocacy		2 (40)	3 (60)		2 (40)	3 (60)			
Osteoarthritis expert	1 (9)	3 (27)	7 (64)		3 (60)	2 (40)	1 (17)		5 (83)
Characteristics of the health service provider									
Consumer	1 (6)	2 (13)	13 (81)		1 (11)	8 (89)	1 (14)	1 (14)	5 (71)
Health care provider	2 (9)	4 (17)	17 (74)		1 (11)	8 (89)	2 (14)	3 (21)	9 (64)
Health policy maker		3 (30)	7 (70)		2 (33)	4 (67)		1 (25)	3 (75)
Māori advocacy		1 (20)	4 (80)		1 (20)	4 (80)			
Osteoarthritis expert	1 (9)	1 (9)	9 (82)		1 (20)	4 (80)	1 (17)		5 (83)

CHAPTER 5: METHODS (PART B)

5.1 CHAPTER OVERVIEW

The aim of this chapter is to describe the methods used to model decision-makers' preferences for knee OA interventions in a CPG. The chapter begins with a description of the key properties that criteria must fulfil so they can be structured into performance levels and then weighted, according to decision-makers' preferences (see Table 4 for a description of key terms). This is followed by a description of how the criteria were weighted and how performance ratings were assigned to the criteria. The chapter ends with a description of how weights and performance ratings are combined to evaluate and prioritise OA interventions.

5.2 SPECIFYING THE TRADE-OFFS IN AN MCDA FRAMEWORK

Making decisions involves trade-offs. In a multi-criteria decision analysis (MCDA) framework, trade-offs are explicitly specified using criteria relevant to the decision problem. Recall that trade-offs represent value judgements – how much a decision-maker is willing to give up value in one criterion to achieve more on other criteria (Table 4). Modelling the preferences of decision-makers involves measuring their trade-offs between criteria and trade-offs for changing levels within criteria. Poorly constructed trade-offs may lead to spurious results with respect to stakeholders' preferences.

A trade-off should be *explicit* so that unintended value judgements are not made by decision-makers. Implicit trade-offs can lead to biased or spurious results because of unintended value judgements. For example, if a decision-maker is presented with two interventions described only by their accessibility, but considers other factors, such as intervention benefits and harms, then the trade-off captures value judgements other than intervention accessibility – namely the decision-makers' preferences for the benefits and harms of intervention. This may lead to an incorrect estimate of how much the decision-maker prefers the intervention accessibility criterion. Therefore, trade-offs should represent value judgements accurately and precisely, so that they represent what the decision maker has in mind, i.e. their preferences.[361] The following sections will describe the properties of criteria that help to minimise unintended value judgements.

5.2.1 Identifying appropriate criteria

Criteria should be valid, reliable, and concisely and unambiguously defined, such that each criterion represents a singular preference or value-judgement.

Firstly, the criteria should capture all critical values underlying the decision (**completeness**).[290] For example, it would not make sense to survey intervention safety without also considering therapeutic benefits at the same time.

Secondly, the criteria should be limited to only the essential or salient number in order to strive for "simplicity and parsimony, rather than complexity" (p.443).[362] Marsh et al. (2016)[290] explain that reducing the number of criteria can help to improve survey efficiency by easing respondents' cognitive burden that may lead to inconsistent responses. However this efficiency gain must be traded with potentially raising stakeholders' concerns about omitting criteria they consider relevant. If redundant criteria must be included, this may be offset by performing an analysis to see if the redundant criteria have any effect on the overarching decision objective. On average, 8 criteria (range 3 to 19) are included in healthcare-related MCDA studies.[268]

To achieve simplicity and parsimony two conditions must be satisfied: **non-redundancy and preference independence**. Non-redundancy is achieved by including only the most salient, relevant criteria, such that's there is no overlap or double counting between the criteria. For example, if the performance of all alternatives are the same on a criterion, or a criterion dominates all other criterion, then the criterion could be considered redundant or irrelevant, and therefore excluded. Another example is to include cost-effectiveness with cost and/or effectiveness. This double-counts cost and/or effectiveness because each criterion is captured in cost-effectiveness. Thus, each criterion should focus only on one aspect of benefit to avoid double-counting.

Preference independence requires that preferences for one criterion should not interact with the preferences for another criterion, i.e. preference for a criterion should be indifferent of preferences for all other criteria. An example is meal choice with red or white wine. Choice of red or white wine is usually dependent on whether one eats red meat or fish; and vice versa. Or in the context of health care, Marsh et al. (2016)[290] describe two criteria, the frequency and mode of drug administration. Presumably higher frequencies of administration would be acceptable if the mode of administration was oral compared to injection; thus, the criteria are not independent. Preference independence should not be confused with statistical correlation; i.e., two different criteria can be statistically dependent but preference independent, and vice versa.[362]

Thirdly, the criteria must be **understandable and explicit** such that they are meaningful and not misinterpreted by stakeholders in different ways. For example, the criterion 'effectiveness' could refer to different things such as the effect on pain, function, or even cost-effectiveness, because it is not clearly defined. Similarly, expressing the same criterion as something more complex, such as 'change in the standardised mean difference on pain' may not be meaningful and overly complex to interpret for decision-makers.

Lastly, the criteria should be **operational**, so that alternatives' performance on the criteria can be measured. If no data are available, then the alternatives cannot be equally compared and contrasted on the criteria. For example, stakeholders might consider the criterion 'effect on joint space narrowing' important when choosing or recommending OA interventions, however, if the data are not available, then alternatives performance on this criterion cannot be measured, rendering comparison against other alternatives on this criterion impossible. Measuring the performance of alternatives on the criteria should ideally use an evidence-based approach.[35, 290] Expert opinion, research evidence (including rapid literature reviews, systematic reviews and modelling exercises[297]) or both are commonly used to measure alternative performance in healthcare-related MCDA studies.[268]

5.2.1.1Key points

- In the previous chapter (Chapter 4), nine characteristics of OA interventions were considered by cross-sectoral stakeholders in their choice of OA interventions: 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), Appropriateness (Recommendation for using the intervention now), Risk of harm, Quality (quality of the evidence), Treatment Passivity and Immediacy of Treatment Effect.
- Considering the properties of criteria outlined in this section, two modifications were made to select the criteria. First, the last two criteria were excluded because they were considered to be the least important to stakeholders (Table 11). Second, Risk of harm was stratified into Risk-Mild (risk of mild adverse effects) and Risk-Serious (risk of serious adverse effects).
- Thus, eight criteria were shortlisted to include in the MCDA model which were deemed to be acceptable with respect to the participant burden imposed on participants, and the mean number of eight criteria typically used in MCDA studies.[268]

5.2.2 Structuring the criteria into performance levels

In addition to selecting appropriate criteria, the criteria must also be structured into levels and intervals describing within-criterion performance categories. Levels allow for the generation of weights, representing stakeholders' preferences, between *and* within criteria.

Levels are defined by categorising the 'best' to 'worst' plausible range of a criterion, effectively scaling the degree of achievement possible on a given criterion. Levels may be qualitatively or quantitatively described. For example, consider the criterion 'effectiveness,' where Cohen's *d* (quantitative) of <0.2, 0.2-0.5 and >0.5 specifies the intervals used to create performance levels corresponding to ‘low’, ‘medium’ and ‘high’ levels of effectiveness, respectively. A qualitative example may include subjective assessments, such as an interview candidate’s qualification level, with corresponding levels of good, mediocre or bad. Angelis et al. (2016)[362] describe the following five properties which inform well-defined performance levels.

Firstly, the levels should be **unambiguously defined**, such that there is a clear relationship between the consequences of the levels. For example, the performance of the criterion ‘effectiveness’ is ambiguous if its performance was described by ‘high’, ‘average’, and ‘weak’. The decision maker is left wondering where average sits between ‘high’ and ‘weak’. Or is the ‘average’ level ‘weak’? The performance levels are not well defined.

Secondly, the performance levels should be **comprehensive**, in that they should reflect the full range of consequences of the criteria. For example, capturing the full range of intervention performance for the criterion ‘harms’ would be restricted if it was only described using the levels ‘very high’, ‘high’ and ‘medium’. Decision-makers would be forced to prefer ‘medium’ levels of harm, even if their preferences were for lower levels of harm, and so the full range of preferences for the criterion would be constricted.

Thirdly, the performance levels should be **direct**, such that the performance levels are directly related to the overarching decision problem or objective. As Keeney[361] explains, in a decision about air pollution, two objectives might be to minimise the cost of air pollution and minimise health effects attributable to poor air quality. Air pollution concentration is often used as a proxy in place of minimising health effects. This leaves the decision-maker choosing between air pollutant concentrations and cost, which renders the trade-off between cost and health effects unclear.

Fourthly, the performance levels should be **operational**, such that data informing alternatives’ performance are available and/or measurable, and trade-offs between criteria (and performance levels) are meaningful.

Lastly, the performance levels should be **understandable** or **generalisable**, such that there is no miscommunication about the trade-offs being considered between performance-levels, or misinterpretation about the meaning of the criteria and performance-levels across different decision-makers or perspectives. For example if the objective is to minimise the cost of an intervention, is this in relation to the health system (e.g. policy-maker) or individual (e.g. consumer)? Such criteria should be defined so that it is sufficiently generalisable for different types of decision-makers to interpret.

5.2.2.1Key points

- Trade-offs are at the heart of MCDA and represent value judgements of decision-makers, i.e. their preferences.
- To measure the preferences of decision-makers precisely, criteria should be selected and structured explicitly to minimise unintended value judgements.
- In this thesis, a literature search was conducted to identify commonly used systems for evaluating intervention performance on the criteria and their levels (reported in Table 12).
- Published data were used to inform the performance levels and their intervals, or the judgement of OA experts or the research team if no standardised system for structuring intervention performance into levels was identified (reported in Table 12).

Table 12. Data supporting the criteria performance levels and their intervals

Criteria† (most to least important) Performance levels (worst to best)	Justifications for the defined performance levels and their intervals
Recommendation to use the intervention now Strong against Conditional against Neutral* Conditional for Strong for	Defined according to the GRADE approach.[363]
Quality of the evidence about the intervention Very low Low Moderate High	Defined according to the GRADE approach.[363]
Effectiveness of the intervention Low ($d<0.2$) Moderate ($d<0.5$) High ($d\geq0.5$)	The Standardised Mean Difference (SMD) described by Cohen's *d*.[364]
Duration of the intervention effect Short (up to 6hrs) Short-medium* (<3 months) Medium (3-12 months) Long (>12 months)	Total joint replacement (TJR) was considered to have the greatest treatment duration of effect. Thus, the 2017 NZ Joint Registry[365] was used to estimate the survival rate of TJR in NZ for the best level. For the medium level, we considered the volume and quality of evidence supporting land-based exercise versus the evidence for pharmaceutical over the counter products. On balance, we concluded that a medium-level for duration of treatment effect was justified by the evidence for land-based exercise, while the worst level of duration was defined by the duration of over the counter pharmaceutical interventions, such as paracetamol or non-steroidal anti-inflammatory drugs.

Criteria† (most to least important) Performance levels (worst to best)	Justifications for the defined performance levels and their intervals
Risk of serious harm (**Risk-Serious**) High (1 in 50 chance; >0.5%) Medium (1 in 200 chance; 0.2%-0.5%) Low (1 in 500 chance; <0.2%)	Levels for risk of mild-to-moderate side-effects and serious harm were informed by a conjoint analysis study of patient preferences for the characteristics of OA interventions in the UK[366] and NZ hospital data for post-operative rates of harm for total joint replacement (e.g. deep infection, cardiac arrhythmias, acute respiratory distress syndrome, and emergency readmission within 28 days due to complications of hip replacement). The 2017 NZ Joint Registry[365] was also reviewed to contextualise the highest worst level of risk of serious harm.
Risk of mild to moderate side effects (**Risk-Mild**) High (3 in 4 chance; >50%) Medium (2 in 4 chance; 25-50%) Low (1 in 4 chance; <25%)	See above.
Cost of the intervention High (>$1000 per month or >$15,000 one-off) Medium ($100-$1000 per month or $1500-$15,000 one-off) Low (<$100 per month or $1500 one-off)	We referred to costs published in a review of OA interventions[367] to judge the attribute levels for Cost. The worst level was considered to be incurred by total joint replacement, so we accounted for 2017 private[245] and public cost[367] in our estimation of the upper bound (worst level). In contrast, the best level was considered as over the counter pharmaceutical interventions, such as paracetamol or NSAIDs.
Accessibility to the intervention Inconvenient travel, or wait time (>3 months) Neither convenient or inconvenient travel, or wait time* Convenient travel, or wait time (<1 week)	The worst level is consistent with wait times for access to elective services in the NZ public health system.[368] The best level was considered by the research team as the typical wait time in NZ metropolitan and rural areas.

* These criterion levels were interpolated using a Bézier curve after the weights were collected, via a function of the 1000minds software. This was so the criteria could match the performance levels used in the Royal Australian College of General Practitioners Guideline for hip and knee osteoarthritis (RACGP CPG)[13] for the Quality criterion. An additional level for Accessibility was interpolated after a nationally representative panel was convened to reach consensus about the accessibility of interventions in an eDelphi. For the Duration criterion, an additional level was added to increase the sensitivity of the Duration criterion to match the data extracted from the RACGP CPG.

5.3 IDENTIFYING THE WEIGHTS FOR THE CRITERIA AND THEIR LEVELS

After the criteria have been selected and structured for MCDA, the next step is to quantify decision-makers' preferences for them: their weight. Weights represent the relative importance decision-makers place between and within the criteria. Weighting the criteria is an important step because the weights and intervention performance on each of the criteria are combined to generate a total score for each alternative, using what is called a 'weighted sum model'. Section 5.8 discusses the weighted-sum model in detail. Many methods are available for weighting and scoring.[294] These methods can be categorised into direct and indirect methods.

As the name implies, direct weighting and scoring involves respondents directly expressing how they feel (i.e. their preferences) about the relative importance of the criteria or alternatives from which weights or scores are derived.[265, 358] Usually the relative importance of the weights or scores is expressed in terms of an interval scale (e.g. criteria are rated on a 0 to 100 scale) or ratio scale (e.g. "criterion *A* is three times as important as criterion *B*"). Direct methods include: direct rating, points allocation, Simple Multi-attribute Rating Technique (SMART), Analytical Hierarchy Process (AHP), SMART-Exploiting Ranks (SMARTER), swing weighting, bisection method and difference methods. For a summary of these methods and key references, the reader is directed to the article by Hansen and Devlin (2019).[297]

In contrast to direct methods, indirect methods (also known as decompositional, choice-based surveys[290]) elicit decision-makers' preferences by asking them to choose between alternatives (either real or hypothetical) on some or all of the criteria (full or partial-profile). From these choices, weights and scores are (indirectly) calculated. Put another way, weights are derived from the total value of an alternative's profile consisting (conjointly) of two or more criteria from which weights are derived (indirectly) using regression-based techniques or other quantitative methods.[297, 358]

Indirect methods include conjoint analysis (or discrete choice experiments), PAPRIKA[369] and best-worst scaling. Marsh et al. (2016)[290] define conjoint analysis and best-worst scaling as full-profile methods, in contrast to PAPRIKA which involves making trade-offs between two hypothetical alternatives contrasted on only two criteria at-a-time. Another key difference between PAPRIKA and conjoint analysis is that the latter typically involves making trade-offs between two or more real or hypothetical alternatives at a time involving two or more criteria, which is cognitively more challenging or less intuitive than PAPRIKA.[369] However, considering greater than two alternatives at a time may also result in less survey questions, and therefore less respondent burden.

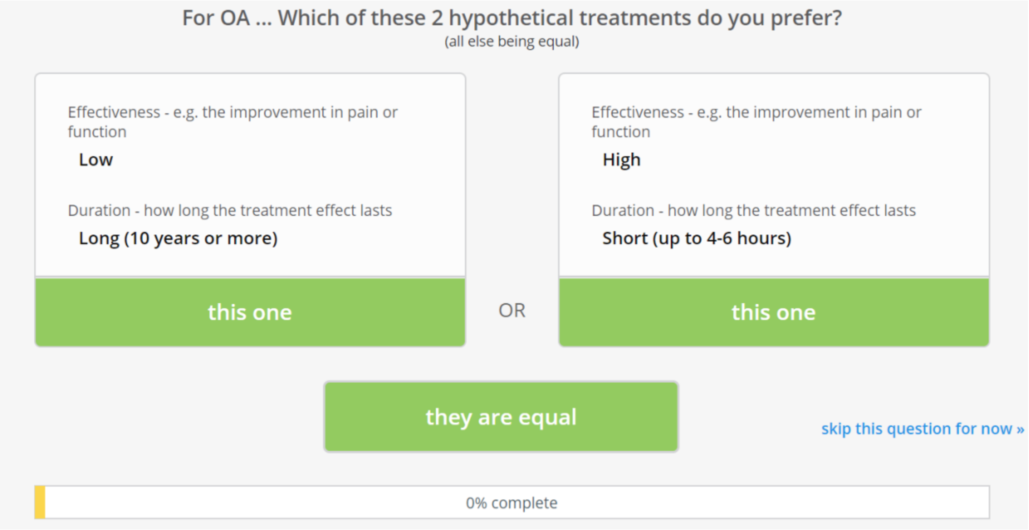
A strength of using partial profiles is that it entails less cognitive load. This is thought to reduce the likelihood of decision-makers using simplifying heuristics ('mental shortcuts') caused by examining too many criteria at once, which requires greater cognitive effort.[266] Such 'mental shortcuts' can lead to systematic errors because trade-offs are not fully considered.[370] However, they may also be criticised for not representing 'real-world' decision-making. A key assumption of indirect methods are that within-criterion preferences are monotonic (i.e. always increasing or decreasing over the range of performance levels for a given criterion). Direct methods do not assume that within-criterion preferences are monotonic.[265]

5.3.1 The PAPRIKA method

5.3.1.1How does it work?

This thesis uses an indirect method with partial profiles called **P**otentially **A**ll **P**airwise **R**an**k**ings of all possible **A**lternatives (PAPRIKA)[369] to generate weights at the individual level, which are aggregated to calculate overall sample mean preference weights. This is in contrast to conjoint analysis whereby weights for the criteria are calculated from the aggregated rankings across *all participants* using regression techniques, such as multinomial logit analysis or hierarchical Bayes estimation.[297] Crucially, the PAPRIKA method allows for cluster analysis (discussed in Section 5.6.3.4) because it generates weights for every respondent. Apart from SWING/SMART and outranking methods, PAPRIKA is the only other method which produces individualised weights.[371]

The PAPRIKA method exploits the idea that choosing between two hypothetical interventions defined on just two criteria at a time and involving a trade-off can be used to implicitly rank 2 or more alternatives. For example, in Figure 16, using the survey platform 1000minds.com (discussed on Section 5.3.1.2), a decision-maker chooses between two hypothetical interventions described on the two criteria, 'effectiveness' and 'duration'.


For OA ... Which of these 2 hypothetical treatments do you prefer?
(all else being equal)

Effectiveness - e.g. the improvement in pain or function
Low

Duration - how long the treatment effect lasts
Long (10 years or more)

Effectiveness - e.g. the improvement in pain or function
High

Duration - how long the treatment effect lasts
Short (up to 4-6 hours)

Figure 16. Example of a pairwise ranking question implemented through 1000minds.com

PAPRIKA is based on four key components (1) ‘undominated pairs’ (of alternatives), (2) ‘dominated pairs’, (3) the principle of transitivity and (4) mathematical linear programming.

Undominated pairs refer to a pair of alternatives where one alternative is characterised by at least one criterion with a higher performance-rating (e.g., categorial high, medium, low) and at least one criterion with a performance rating lower than the other alternative.

Dominated pairs are characterised by one alternative performing better on at least one criterion, and no worse on any other criteria, i.e., they perform better on at least one criterion, all else being equal.

Transitivity refers to the principle whereby if A, B and C are three hypothetical alternatives, and A>B, and B>C, then A must be greater than C. This is a principal feature of the weighted-sum model, which is used to calculate total scores and requires that the criteria must be independent and non-overlapping. This is discussed in Section 5.8.

Based on the three components above, PAPRIKA uses an algorithm based on mathematical linear programming to derive ‘part worth utilities’ for the criteria and their levels. As the decision-maker answers more questions by making successive trade-offs between pairwise alternatives (e.g. Figure 16), the algorithm identifies all unique undominated pairs and all implicitly ranked pairs based on explicitly ranked undominated pairs until all undominated pairs have been either explicitly or implicitly ranked.

The algorithm is constrained to be strictly additive (no interaction between criteria is allowed), non-negative and monotonically increasing in the levels of each criterion (always increasing or remaining constant). The part worth utilities are normalised across the criteria into ‘point values’ or weights (%) for ease of interpretation.

Point values for each category represent both the relative importance and the degree of achievement for a given criterion's level of performance.[369] The lowest performance level represent the minimum/worst performance on a criterion and is assigned zero points. The highest performance level represents the maximum/best performance of a criterion and sum across the criteria to 1 (or equivalently, 100%). In practice, the four key components are applies in real-time through an online survey-platform, 1000minds.com. Section 5.3.1.2 describes the software in more detail. Hence, PAPRIKA is also known as an 'adaptive conjoint survey' method.

Although this may sound trivial, as Ombler and Hansen (2008)[369] explain, a model defined by 3 criteria described on 2 performance-levels each results in 9 undominated pairs (i.e. decisions). However, simply doubling the number of criteria and performance levels to 6 and 4, respectively, increases the number of undominated pairwise comparisons (i.e. decisions) to 7,390,656. Clearly, it would be impractical to make this number of pairwise comparisons, but by utilising PAPRIKA's real-time software (1000minds.com) the number of decisions required to solve all undominated pairs is reduced to approximately 65 questions, without sacrificing precision in estimated criteria weights. It has been shown that decision-makers can comfortably make 50 pairwise choices, and up to at least 100[369] using this method.

Experimental design is important to minimise confidence intervals around parameter estimates in a choice model for a given sample size (i.e. statistical efficiency). Perfectly efficient designs are described as balanced, meaning that each level appears equally often within an criterion, and orthogonal, meaning that each pair of levels appears equally often across all pairs of criterion within the design.[358] The beauty of PAPRIKA is that the experimental design does not require complex planning of survey questions (or choice sets) to achieve statistical efficiency (E.g. how many alternatives should be presented to the respondent at once and how many criteria should be used to describe them?). Indeed, PAPRIKA can elicit preferences involving more than two criteria at a time however modelling has shown that overall rankings are highly correlated with the true rankings – strengthening the argument that it is an efficient and intuitive design[369] (the reader is invited to read the seminal review by Clark et al. (2014)[37] for more information on experimental design with respect to conjoint analysis). In this respect, the efficiency of PAPIRKA is also a strength relative to conjoint analysis, where the experimental design must be carefully considered to generate acceptable parameter estimates for a given sample size.

Another strength of PAPRIKA is its accessibility. In contrast to direct methods such as direct rating or SMART for example, it does not require the decision-maker to be informed or an 'expert' about the performance of the alternatives being considered (because a partial profile, rather than a full-profile considered). In turn, PAPRIKA is more accessible to respondents due to less prior knowledge requirement. Moreover, because PAPRIKA involves ordinal decisions (choosing one hypothetical intervention over another), it is more natural intuitive than making decisions between scale data (e.g. "Is alternative *A* is better than alternative *B?*" is arguably easier to evaluate than "How many times is alternative *A* better than alternative *B?*"). Moshkovich et al. (2002)[372] assert that ordinal decisions result in more stable and reliable responses.

5.3.1.2How is PAPRIKA operationalised?

The PAPRIKA method is implemented in real-time using a cloud-based software package called 1000minds which is accessed via a web-browser interface (1000minds.com). Through the survey platform, the developer/researcher is able to configure and disseminate a survey (e.g. Figure 16) to generate weights, track decision-makers' progress, generate results and generate figures. Specifically, surveys are created to quantify respondents' preferences for: the alternatives, criteria and their levels. During this stage of development, the developer can also remove 'impossible combinations' from the survey. For example, it may not be realistic for an intervention for a particular condition to be highly toxic and also highly effective. Two other useful features of

1000minds is that the developer can choose to include repeat questions and generate a report showing the weights within and between the criteria generated by their choices. These features are useful to assess the reliability of the survey responses and potentially verify the respondents overall agreement with the survey results.

The 1000minds platform also implements the weighted-sum model (described in Section 5.8) so the alternatives can be ranked in order of importance. Thus, intervention performance matrices (e.g. Table 5) can be uploaded to the platform to facilitate this process. The software also generates preliminary results (e.g. total sample mean preference weights) and figures to graphically communicate the results, such as radar and tornado charts (e.g. for a sensitivity analysis, described in Section 5.9), or the user can choose to generate a data file instead to perform additional analyses. For example, do stakeholders' preferences differ by group?

To date, 1000minds has been applied to many health-related research problems, for example to prioritise patients, prioritise new and emerging health technologies (such is the application in this thesis; examples include with similar analyses include prioritising non-communicable diseases,[373] and publicly funded health technologies in NZ[371, 374]), classify and diagnose diseases, and prioritise research agendas. However, it has also been applied in other sectors such as government, to explore retirement income policy (e.g. Au et al. (2015)[375] who apply a similar analytical approach to that in this thesis), land use and urban planning, and police work. An up-to-date list of applications is reported on the 1000minds website, https://www.1000minds.com/sectors/. Given its application across different sectors and for health-related research, 1000minds has demonstrated its ability to be configured for a wide range of decision-making problems, making it an ideal choice in this thesis.

5.3.1.2.1 What other software packages are available?

Methods for generating weights and scores follow traditional paper-surveys, however, the advent of modern computing has given rise to greater utilisation software-based solutions. Indeed, a broad review of MCDA software in 2016 by Weistroffer and Li (2016)[376] identified 69 different software for MCDA. A more focussed systematic review of MCDA software in healthcare priority setting identified seven software tools (Table 13): 1000minds, M-MACBETH, Socio Technical Allocation of Resources (STAR), Strategic Multi-Attribute Ranking Tool (SMART), Visual PROMETHEE, EVIDEM and the Prioritisation Framework. Their review highlighted that 1000minds is a robust, accessible software package that was, arguably, the only software able to address the range of software characteristics which were collected. For example, 1000minds is the only software package that is not limited by operating system (e.g. windows vs mac) or software package (e.g. the requirement to have a Microsoft Excel license). Also, unlike SMART Vaccines or Visual PROMETHEE, 1000minds is not limited by the user's operating system. Moreover, out of the software reviewed, 1000minds was one of only two options that could produce value for money charts and perform cluster analysis. In terms of licensing cost, there were four other software packages available that were free of charge. In this thesis however, 1000minds was made freely available for use, making it a compelling option, given its broad set of features.

5.3.1.3Summary

There are many different methods for generating weights and alternative scores which may be broadly categorised into direct and indirect methods. In this thesis, the PAPRIKA method and 1000minds software platform was chosen primarily due to its simplicity, efficiency and accessibility, making it an ideal choice for the broad range of stakeholders in the NZ health system.

Table 13. Overview of seven multi-criteria decision analysis software packages identified by Moreno-Calderón et al. (2020).[377]

Characteristic	**1000minds**	**SMART vaccines**	**Visual PROMETHEE**	**START toolkit (Health Foundation)**	**M-MACBETH**	**EVIDEM**	**PHE Prioritisation Framework**
Creation date	2003	2012-2015	2010	2013	Early 1990s	2006	2018
Country	New Zealand	USA	Belgium	UK	Portugal	Canada	England
Operating System compatibility	Any (cloud-based software)	Windows (XP or above)	Windows (XP or above)	Excel 97-2007	Windows (XP or above)	Excel 2003 and later versions	Excel 2003 and later versions
MCDA method	PAPRIKA	MAUT	PROMETHEE	PBMA/SAW	MACBETH	SAW	SAW
Inputs	Configuration form criterion relevance and performance matrix; pairwise trade-offs	Combination boxes and scroll bars; swing weighting	A table where alternatives, criteria, weights and their values are entered; 1000 points are distributed across all criteria	Excel table with a list of interventions, benefit in numbers, feasibility of success, cost, intervention priority, etc.	Value tree where the criteria to be evaluated are entered, table of alternatives, performance matrix, scrollbars for scoring; pairwise trade-offs	Excel table where the score and weighting values given by the evaluators are entered	Spreadsheets: define scope, define criteria, weighting, gather evidence, scoring

Outputs	Preference values bar chart, radar chart of criterion weights, criterion value functions, rankings table, disaggregated tornado chart, bubble chart (including efficiency frontier)	Ranking bar graph, rankings table	Ranking table; diamond, network and rainbow graphs, Gaia plan, etc.	Chart benefit per person, value for money graph, efficiency frontier graph	Sensitivity analysis XY graphic, robustness analysis table, ranking table, among others	Dispersion and radar graphs referring to criteria weights score per criterion and criterion contribution	Summary tables of the ranking and scenarios, final recommendation dashboard
Value for money	Yes	No	No	Yes	No	No	No
Sensitivity analysis	Tornado charts; 1-way sensitivity analysis	No	Modules where it is possible to change the criteria weighting and evaluate the final score of the alternatives	No	Graphically, the weight of each criterion can be changed 1 at a time and the others maintain their distribution proportionally to determine the impact on the final score	No	Scenario modelling to compare budget based on the potential scores and proportional investment across all programme areas

Cluster analysis	Yes	No	Yes	No	No	No	No
License/versions	Free 21-day trial; paid version typically US$20,000 but is proportional to the application	Full version with no cost	Free for academics; business edition £1250	Full version with no cost	Demo; full license from £175-£17,500 depending on context (e.g. academic vs corporate)	Full version with no cost	Full version with no cost
Supporting documentation	Templates provided, help section on website, contact developers	Software self-guides the user, or they can contact the developer	PDF manual, help section on website	Guidebook and training tools available online	Each section is supported by a guide	Tutorials and instructions available on developer's website	Each section is supported by a guide

MAUT Multi-Attribute Utility Method, MACBETH Measuring Attractiveness by a Categorical Based Evaluation Technique, MCDA multi-criteria decision analysis, PAPRIKA Potentially All Pairwise RanKings of all possible Alternatives, PBMA Program Budgeting and Marginal Analysis, PHE Public Health England, PROMETHEE Preference Ranking Organization Method for Enrichment Evaluations, SAW Simple Additive Weighting. Reproduced with permission from Springer Nature (License No.5045160426237)

5.4 INSTRUMENT DESIGN

It is important to ensure that differences in survey responses across respondents and over time is consistent for observational studies (i.e. among individuals and across assessments). If variation in survey responses is due to survey implementation among observers or time (systematic error), then the collected survey data may lead to inaccurate conclusions. Establishing the quality of survey instruments is therefore an important step in survey design. Fundamental to this is the assessment of face validity.

5.4.1 Assessing the face validity of the survey instrument

Validity refers to how well a survey, study or instrument measures what it intends to measure. Face validity is central to good health surveys and is most commonly evaluated and relied on for assessing validity in single-use surveys.[378] Face validity refers to what a respondent thinks a question means; improving face validity can help to reduce sampling errors (thereby increasing survey data quality) and increase response rate. In the context of completing a choice-based survey, respondents may misinterpret the overarching decision objective, or misinterpret the trade-offs between criteria; this may lead to incorrect weightings, and therefore conclusions about participants' preferences. To achieve greater face validity, a survey needs to accurately communicate the intended meaning of survey items to the survey respondent, so they can respond to the questions as intended by the researcher.

Interpreting survey questions involves a process of cognitive processing described by:[379] (1) comprehension (question interpretation), (2) retrieval of information from memory (information retrieval from memory), (3) judgement/estimation processes (decision-making processes) and (4) response processes (selection of an answer from the survey response categories that closely matches their decision). There are several things which may go wrong during each of these steps.

Firstly, complicated use of language can cause the survey questions to be misinterpreted. Two well-known examples include professional jargon and language translations.[380] For example, a 'myocardial infarction,' though highly specific in medical terms, is not as accessible to a wider audience than say, 'heart attack'. Willis (2004)[381] explains that researchers often overestimate the literacy of the survey respondents and employ survey methods that do not allow for immediate clarification.

Another issue relates to retrieving information from memory, such as reporting events or experiences that happened in the past. Two problems may occur such that the question under consideration may not be meaningful to the respondent (e.g. if what they experienced was so rare or long ago in the past, it is no longer meaningful to them). In this case, the respondent might guess or take no interest in the question. Secondly, respondents may be tempted to self-censor topics to more socially-desirable responses such as age, income or lifestyle habits.[382]

Once the respondents have interpreted the question, drawn on available memory and considered a response, the next step is matching the internal response to the available answers. When options do not correspond to the internalised answer, respondents may become frustrated or confused[381] and therefore less likely to complete the survey.

5.4.1.1.1 Cognitive interviewing

One way to assess if these problems are present in a respondent's thought processes is to access what a respondent is thinking when they answer survey questions. Cognitive interviewing (also known as 'think-aloud interviewing', 'think out-loud' and 'verbal protocols') is commonly used during 'pre-testing' or 'pilot testing' to identify problems in a survey prior to broad distribution and main data-collection.

Cognitive interviewing is fundamentally a form of semi-structured, in-depth interviewing. It involves the interviewer reading survey questions to a respondent and then collecting verbal information about their survey responses, by asking them to "tell me what you are thinking" by thinking aloud or reading aloud their thought processes. This helps the researcher understand how they perceive and interpret survey questions, identify potential survey problems (e.g. unexpected interpretations) and assess whether or not the survey question is generating the intended research information.[383] Subsequently, prospective survey questions can be modified to achieve their intended purpose. Through this process, the validity of a survey instrument can be improved.

Two cognitive interviewing paradigms exist. The first paradigm, 'pure' cognitive interviewing originated from laboratory-based psychology experiments which relied heavily, if not exclusively upon survey participants thinking-aloud as they read and answered survey questions. Later in the development of this method, probing was introduced to guide the interview process. The probing paradigm involves direct, additional questions about the content of survey responses.[383]

Cognitive interviewing with probing has several key advantages. First, some interviewees may struggle to verbalise their thought process; probing can help to overcome this possible barrier. Secondly, probing can help to focus the participant's behaviour (i.e. stop them from going on tangents). Thirdly, probing may help to reduce the cognitive load associated with verbalising thought processes, while still accessing information stored in short-term information. Theoretically, this may be less intrusive than performing 'pure' cognitive interviewing.[383] Finally, probing generates information that is relevant to improving the questionnaire, that may not necessarily have emerged otherwise.[383] For example, pure cognitive interviewing might suggest an issue with an survey item, but without probing, insufficient information may emerge to actually diagnose and resolve the problem.

Interviewers who use probing can choose between asking verbal probes during the think-aloud/read-aloud process (known as concurrent probing) or after the think aloud/read aloud process (retrospective probing). Concurrent probing is the most common approach used because it accesses information 'fresh' in the respondents mind, whereas if probing was conducted retrospectively, there is increased risk that the respondent will forget what they were thinking about and instead fabricate their response. However, retrospective probing is preferable for the following circumstances: (1) for testing self-administered questionnaires, to investigate if the respondent can complete the survey tasks unaided and (2) later in survey development, when researchers seek to simulate the questionnaire in the field.[381]

In practice, a combination of these approaches are and may be used during the interview process to identify response errors.[379, 381] To enhance the validity of the cognitive interview process it is important to plan the interview format, probing questions, and analysis.

5.4.1.1.2 Interview planning

As with any data collection, it is important to standardise the process with a protocol. Willis (1994)[384] proposes that the interview process should begin with a standard introduction about the purpose and format of the interview, followed by a simple warm-up exercise, *"Picture in your mind where you live. How many windows are there? As you count up the windows, tell me what you are thinking about and seeing"* (p;7). However Garcia (2011)[385] noted that the participants in her study could have been better prepared for the cognitive interview if the practice exercise matched the type and content of the questionnaire. In the context of this thesis, the following question might be appropriate: *"Picture in your mind having to choose between a number of treatment options for hip/knee OA.*

How many treatment options there? As you count up the treatments, tell me what you are thinking about and seeing."

Standard probing questions should also be prepared prior to the interview. Understanding where errors typically surface in questionnaire design can help with developing probing questions. Understanding the analytic objectives of the cognitive interviewing can help to focus probing questions. For example:

- Can the survey instructions and questions be read?
- Is there any missing information?
- Does the participant understand the descriptions used to define the criteria and performance levels?
- Is there any jargon?
- What does the participant think about the overall usability and length of the survey?

Willis (2004)[381] proposes systematically applying a checklist of common survey problem areas[386] to help anticipate where issues may arise and inform which probing questions may help to elucidate these problem areas. Areas of concern include:[381]

- reading (listen for problems with reading)
- instructions (look for issues with instructions from the respondents perspective)
- clarity (are there any issues with the intent or meaning of the question to the respondent such as vagueness, jargon, wording)
- assumptions
- knowledge/memory (are there any memory retrieval issues associated with answering a question?)
- sensitivity/bias (are the questions culturally sensitive or biased?)

- response categories (are the range of response categories reasonable and comprehensible?)
- other problems (e.g. question ordering).

In this thesis, think out loud cognitive interviewing with probing will be used to inform modifications to the survey instructions and the descriptions of the criteria in order to establish the face validity of the survey instrument in Chapter 6.

5.4.2 Sample size recommendations for pilot testing

Willis (2004)[381] recommends conducting two to four sets of interviews with 5 to 15 people in each set, interviewing as many people as possible with similar sociodemographic characteristics to the intended survey population (i.e. purposeful sampling).[381] Falkner (2003)[387] reported that a sample size of 10 can detect between 82-94% of usability problems compared with 15 detecting 90-97% of possible issues. With a sample size of 20, 95-98% of usability issues may be detected. Roughly twice the amount of time estimated to complete a survey should be set aside for cognitive interviewing.[381] Stopping criteria for the number of sets of interviews to conduct can also be based on 'saturation,'[388] which is triggered when successive interviews yield relatively few new insights (i.e. diminishing returns).[383]

In this thesis, 5 rounds of pilot testing were conducted with 1-5 people in each round. Convenience sampling was used to overcome barriers to accessing participants and time barriers. This is discussed in Chapter 6.

5.4.2.1Analysing interview data

Interview data may consist of audio transcriptions and field notes which may be qualitatively analysed. Drennan (2003)[379] notes that analysing cognitive interviews is largely subjective in an otherwise comprehensive method of survey pretesting. Beatty et al. (2007)[383] suggest that analysis can be based on whether problems can be logically attributed to question characteristics, which is the approach used in this thesis.

5.5 SAMPLING AND RECRUITMENT

Calculating an appropriate sample size is ill-defined for MCDA studies. It has been suggested that the appropriate number depends on a number of factors, including the complexity of the choice survey, the question format, the degree of heterogeneity in the target population, the availability of respondents and the need to conduct subgroup analyses.[310, 358, 389] This is, in part, due to the advent of technology which has given rise to more efficient survey designs (as seen with, for example, computer-adaptive methods, such as PAPRIKA); indeed, some researchers suggest that focussing on the representativeness of the sampling frame is more important than the absolute sample size.[390]

In this thesis, convenience and snowball sampling was used to recruit cross-sectoral participants consistent with those recruited in the study described in Chapter 4: OA consumers, health care providers, policy-makers, Māori advocates and OA experts. An extensive effort was made to invite participants, for example through: self-selection on social media websites (Facebook, twitter, webpage invitation); government organisations at the national level (Ministry of Health) and jurisdictional level (District Health Boards and Public Health Organisations); health care professional organisations (e.g. NZ Rheumatology Association, Physiotherapy NZ, Royal NZ college of GPs, NZ Orthopaedic Association); consumer and Māori advocacy organisations (e.g. Arthritis NZ); general practices through an online health service database (https://healthpages.co.nz/) and OA experts through an online database ((http://expertscape.com/). Lastly, to address the relatively small pool of health policy-makers and content area experts in NZ, reciprocal ethics approval was also sought from an Australian university to send invitations to stakeholders based in Australia.

5.6 ANALYSING PREFERENCE WEIGHTS

5.6.1 Assessing the quality of responses

Inconsistent survey responses may lead to unreliable survey data and therefore study conclusions. Reliability refers to the consistency and reproducibility of survey instrument responses. Responses to a question consist of true (systematic) variance and random error. Reducing random error is the purpose of assessing reliability. Testing whether a respondent produces the same response (i.e. agreement) to a question over a short period of time is called test re-test reliability[378] and can be used to assess the internal validity of the data.[358] Eliciting value judgements is subject to bias and random errors and so in this respect test-retest reliability is useful to assess. Other ways to assess the quality of the data include screening for consistent responses, e.g. always choosing the left (or right) alternative presented in the questionnaire.

The PAPRIKA method includes the ability to randomly generate implicitly or explicitly answered questions to test stakeholders' consistency. Keeney (2002)[361] recommends that if errors are detected, all value judgements (i.e. trade-offs) should be reviewed and adjusted as necessary. However, there is empirical evidence to suggest that decision-makers may make inconsistent choices through rational thought processes[391] and removing 'irrational' responses could in fact introduce selection bias and reduce statistical efficiency.[392] Nevertheless, including a reliability component offers the benefit of assessing the data quality of a choice survey through the consistency of the trade-offs. In this thesis, respondents will be asked to repeat three survey questions to assess the reliability of their responses using a built-in function of the 1000minds survey platform.

5.6.2 Statistical significance and probability values

A question relevant to this thesis is whether or not the preferences of stakeholders for OA interventions differ by group. Statistical significance and probability values can help to answer this question. Central to statistical testing is whether or not an observation or measure (e.g. the mean difference between preference weights) is probable or not probable at some level of confidence. For example, depending on the size of the difference and the level of significance, that can be interpreted as 'almost certainly due to chance alone' or 'probably due to chance' or that you 'cannot state with confidence that it is not due to chance'. Probability is tested under two hypotheses, the null hypothesis (H_0; there is no difference) and the alternative hypothesis (H_1; there is a difference). The calculated probability value, or *p*-value, is the probability of finding the measure (or more extreme value) when the null hypothesis is true. The significance level, or alpha (α) is the likelihood that the null hypothesis is rejected when it is true (known as a Type I error). For example, if the *p*-value is set at 0.05 and 0.01, there is a 1 in 20 and 1 in 100 chance, respectively, that a Type I error will occur. The opposite of a Type I error is the likelihood that the null hypothesis is accepted when it is in fact false. This is called a Type II error.[393]

5.6.2.1Familywise error rate

The familywise error rate (FWER) is the probability of making Type I errors with increasing numbers of statistical hypothesis tests. The FWER is calculated by the following equation, where *n* is equal to the number of comparisons conducted (assuming a 0.05 level of significance):

$$FWER = 1 - (0.95)^n$$

For example, if 14 comparisons were made, the FWER is 0.51, which means that there is at least a 51% chance of making a type I error. Post hoc tests adjust the level of significance of individual tests so that the overall rate of type I errors remains at the predetermined significance level.

5.6.3 Do preference weights differ by group?

It is important to establish if decision-makers' preferences for the criteria differ by group. If they do, then it follows that their preferences for alternatives are likely to be different. This is important, because it has implications about inferences that can be made from the data, such as the generalisability of the results. Different quantitative methods can be used to analyse preference weight data.

Quantitative methods can be grouped into parametric and non-parametric statistical tests. Parametric tests assume that the data are normally distributed. Non-parametric tests are suited to data that does not conform to a normal distribution. The distribution of data refers to the spread of data about the mean. A normal distribution describes data that are symmetric about the mean, whereby values tend to cluster around the centre and reduce in frequency to the left and right of the mean (the standard deviation is small relative to the mean). Normal distributions are characterised by: (1) sharing the same value for the mean, median and mode; (2) exhibiting symmetry (i.e. no skewness or kurtosis); and (3) distributions of half of the population above the mean, and half of the population below the mean.

The standard deviation of a distribution can be used to determine where a value, or observation sits within the distribution. For example, 68% of values fall within ±1 standard deviation of the mean, and 95% of values sit within 2 standard deviations of the mean. When the mean is zero and the standard deviation is 1, this is a special case of the normal distribution known as the Z-distribution. Observations can be transformed into standardised scores representing their relative position on the Z-distribution. These are known as z-scores, and allow comparisons to be made between observations (e.g. mean difference) and probabilities to be calculated across populations (e.g. the probability of observing a mean score).

Many parametric tests rely on the ‘Assumption of Normality’. This assumption is that the sampling distribution of the mean is normal, or that the distribution of means across samples is normal.[393] For example, say a random sample of patients with knee OA were surveyed for their level of knee pain and the mean pain level of the sample is reported. If this process was repeated over and over again, in another random sample, the distribution of sample mean pain levels would be normal. It is often assumed that the sampling distribution of the mean is normal because of the Central Limit Theorem, whereby the sampling distribution will take the shape of a normal distribution regardless of the shape of the population from which the sample was drawn.[393] This is a key assumption underpinning many parametric statistical approaches, such as the analysis of variance test (ANOVA). Due to the Central Limit Theorem, the assumption of normality can generally be imposed when sample sizes of >30 are achieved.[394]

5.6.3.1Parametric tests

ANOVA is a parametric test used to test if there is a statistical difference between the means of three or more groups. For example, do consumers have different preferences to health care policy-makers? ANOVA makes two assumptions: (1) the dependent variable (or outcome variable) is normally distributed (e.g. group-wise mean preference weight); (2) there is homogeneity of variances (i.e. the population variances in each group are equal). Bartlett’s test of equal variances can be used to detect if there is homogeneity of variances. Interestingly, although ANOVA is a parametric test, it is considered robust, such that it can also be used with Likert data, unequal sample variances, small sample sizes and non-normal distributions.[395] ANOVA is also an omnibus test, which means that it only detects that there was at least one significant difference between two groups.

To determine where the significant differences were detected between more than two groups, multiple pairwise comparisons of group means must be performed. These are called post hoc tests. Student’s t-test is typically used to test for significant differences between two group means. However, because multiple comparisons are made, a correction for Type I error must be applied to control for the FWER.

The Šidák[396] and Bonferroni corrections are two adjustments for the FWER. The main difference between the two is that the former assumes that the tests are statistically independent, whilst the latter does not. Bonferroni correction is considered more conservative (it over corrects for Type I error) than the Šidák correction.

5.6.3.2 Non-parametric equivalents

The non-parametric Kruskal-Wallis test[397] is analogous to the ANOVA test, and the appropriate post hoc test is Dunn’s test.[398] Rather than compute differences in continuous variables and means, these tests compute differences in rank ordered variables and medians. The Holm-Šidák correction[398] can be used to correct the FWER with Dunn’s test.

5.6.3.3 Fractional multinomial regression

Regression analysis is used to determine if there are associations between independent variables (also called explanatory, control, or predictor variable e.g. socioeconomic characteristics) and dependent variables (also called outcome, response, predicted, or explained variables, e.g. preference weights). There are different types of regression analysis that can be employed. For example, the simple linear regression model explains one dependent variable in terms of another independent variable, while multiple regression uses more than one independent variable to explain another dependent variable. Another type of regression, logistic regression, is used when the dependent variable is binary, e.g. 0 or 1. Multinomial regression (or multinomial logistic regression) is used where the dependent variable is nominal with more than two possible responses.[399] Finally, the fractional multinomial logistic regression model is an extension of the multinomial regression model,[400, 401] whereby the dependent variable is in fractions that sum to 1.[402] The PAPRIKA method produces preference weights between 0 and 1 that add to 1, which makes this an appropriate model to select for exploring associations between sociodemographic characteristics and preference weights.

5.6.3.4 Cluster analysis

Clustering is a group of methods that aim to group data by similarity, thereby revealing underlying structures, or patterns, in data. Referred to as *person-oriented* approaches, cluster analysis identifies and describes groups of individual cases defined by similarities across multiple dimensions of interest. This contrasts with traditional *variable-oriented* methods, which posit that any object can be investigated by reducing it to its primary elements and analysing each element in isolation.[403]

For example, people living with OA (object) are often characterised by their sociodemographic characteristics (e.g. age), and these characteristics are often examined for associations with an outcome of interest (e.g. preference for risk of harm). This approach results in the assumptions about the relationships between variables across the entire population, such as ‘lower age is associated with lower risk of harm’. There are limitations of this approach which are linked to the generalisability of results that Mandara[403] raises. These are now discussed in relation to the example above.

First, the variable-centred approach assumes that results are the same for every person. Within-group variation demonstrates that this is not the case, and therefore limits the external validity of the results.

Second, in variable-centred studies, the unit of analyses is the variables and not groups. If the unit of analysis is the variable, then conclusions on the case level are not justified. Knowing about the correlation between consumer preferences for intervention harm and age only tells us about this relationship across all consumers on the dimension age. Age does not explain everything about an individual consumer, or subgroup of consumers (e.g. those with early, mild or advanced OA).

Third, measuring one element or dimension of a multidimensional system only gives insight into one part of the whole. Without studying all the dimensions in a system, it is unclear what and how relationships with other elements can affect results. Continuing with the previous example, other demographic characteristics such as professional occupation and work experience, may interact with age to change consumers' preferences for intervention harm.

As mentioned earlier, cluster analysis, is a *person-oriented approach* which focus on groups of individual cases defined by similarities across one or more dimensions. Fundamentally, it is a group of methods which classify data into groups that maximise within-group homogeneity, and minimise between-group heterogeneity, where the number of groups, as well as their forms, may be unknown.[404]

Cluster analysis differs from variable-oriented approaches because it investigates the whole system, not just the major dimensions of a system. In the example above, cluster analysis could be used to determine if consumer preferences for harm can be clustered into groups; and then all sociodemographic characteristics regressed against the clusters to explore if they could explain the clusters observed.

Two major clustering methods include hierarchical and non-hierarchical clustering. There are differences. First, non-hierarchical clustering requires the researcher to determine the number of clusters a priori in the final solution. Second, while both methods create mutually exclusive clusters, non-hierarchical clustering does not represent any theoretically nested structure. This is explained further in (Henry et al., 2015; p.122[405]). In this thesis, a non-hierarchical theoretical framework is appropriate because nested relationships within intervention preferences are not being examined; therefore, either method is appropriate.

Like all clustering, hierarchical clustering is a sequential process, with the ultimate goal of finding two clusters that are nearest to each other to merge. Once clusters are formed, they are linked with other cases to create larger clusters, until all cases are linked into a single cluster. The first step in this process involves measuring the distance between cases in a space defined by the variables used in the analysis.

To link observations together equally in a multivariate space, the distance between individual cases need to be calculated. The most common distance metric is the Euclidean distance, which is calculated by summing the squared differences between cases on each variable and using the square root of the sum. This allows for the distance between two cases to be calculated across all variables and reflected in a single distance value. Euclidean distances are recommended for cluster analysis in psychology.[405]

Because Euclidean distance can only be calculated for two cases at a time, a problem arises where more than two cases need to be compared at once, as is the case with clusters. *Linkage* solves this problem by comparing more than two cases simultaneously.

The second step in cluster analysis is to determine the linkage measure. Linkage refers to the point in a cluster where a distance is measured to determine the similarity between clusters. This can be calculated in several ways, the most common are briefly discussed below.

Single linkage, also known as nearest neighbour linkage, measures the distance between cases in two clusters (two pairs of cases) with the smallest distance between them. Complete method, also known as farthest neighbour linkage, is the opposite, and measures distance between clusters by using the cases with greatest distance between them.

Average linkage takes the average of the distance values between pairs of observations; the distances between each case in the first cluster and all cases in the second cluster are calculated and then averaged.[406]

Ward's linkage links clusters together based on the similarity between observations in the same cluster. It minimises the within-cluster sum of squares of each cluster when clusters are joined together.

Henry et. al. (2005)[405] recommendations for cluster analysis are: (1) use single linkage if no clusters are detected; (2) if the data are multivariate normally distributed, then use Ward's linkage and, (3) if the data are not multivariate normally distributed, use complete linkage or centroid linkage.

5.7 RATING ALTERNATIVES' PERFORMANCE ON THE CRITERIA

The previous sections focussed on the measurement and analysis of decision-makers' preference weights. This section shifts from decision-makers' preferences to OA interventions' performance on the criteria and their levels described in the previous sections. Hence, in this section, the alternatives will be referred to as 'interventions' when concerning their performance.

Once the criteria and their levels have been specified (seen in Table 12), the next step is to measure the performance of the alternatives on each of the criteria and assign performance ratings, according the performance level associated with alternatives' level of achievement on the criteria.

In this thesis, data for 75 OA interventions and evidence about their performance on the criteria established from Objective 1 were extracted from the 2018 Royal Australian College of General Practitioners guideline for hip and knee OA (RACGP CPG).[13] At the time of study conception, this information provided the most complete, rigorous, NZ-relevant and up-to-date evidence at the time to measure intervention performance and rate the interventions on six of the criteria: Duration (duration of the intervention effect), Effectiveness (magnitude of treatment effect), Recommendation (previously appropriateness; for using the intervention now) to use the intervention now, Risk-Mild (risk of mild adverse effects), Risk-Serious (risk of serious adverse effects) and Quality (quality of the evidence). For the two remaining criteria not covered in the RACGP CPG, Accessibility (travel or wait time to access the intervention) and Cost (total financial costs relevant to the use or provision of healthcare for OA), NZ-specific data were sourced to inform the performance ratings for these criteria.

The remainder of this section will describe in detail the sources used to measure interventions' performance and assign performance ratings to the criteria for each intervention, shown in Table 12.

5.7.1 Accessibility

The Accessibility criterion was considered specific to the NZ context and not considered in the RACGP CPG. To rate the accessibility of the interventions, a nationally representative and multidisciplinary panel of 19 experts in the management of OA were invited to comment on the accessibility of interventions in a 2 round eDelphi survey. First, the interventions were clustered into like groups (by me) and then reviewed for consistency by two members of the research team (supervisors). To reduce the participant burden with reviewing all 75 interventions in the RACGP CPG, the interventions were clustered into 12 groups (Table 14). An *a priori* threshold for consensus was adopted as >70% agreement either convenient or inconvenient accessibility, the same threshold used in the RACG CPG.

In round 1, the expert panel rated the accessibility of interventions in the context of providing, funding, planning or delivering interventions within the NZ public health system on a 5-point scale (5= very inconvenient travel, or long wait time (>3 months); 4= inconvenient travel, or long wait time; 3= neither convenient or inconvenient travel, or wait time; 2= convenient travel, or wait time; 1= very convenient travel, or wait time (<1 week)). A response option, 'I don't know/unsure' was also included for each intervention group.

In round 2, the response scale was collapsed into three categories (convenient travel or short wait time (<1 week); neither convenient nor inconvenient travel or wait time, and inconvenient travel or long wait time (>3 months)). Only the intervention groups that did not reach consensus were brought into round 2.

Following round 2, the intervention groups with <70% agreement were assigned a mid-level rating for accessibility. After all the interventions groups were rated by the eDelphi panel, the individual OA interventions were assigned the accessibility ratings associated with their intervention group.

Two additional concessions were made to rate the accessibility of the OA interventions: if an intervention was not available, approved or indicated for OA in the NZ Formulary (www.pharmac.govt.nz), or NZ Pharmaceutical (PHARMAC) Schedule (www.nzformulary.org), the lowest level of accessibility was selected.

5.7.1.1Results of the eDelphi survey

5.7.1.1.1 Round 1

The results of the eDelphi survey are shown in Table 15. Out of the 19 OA management experts identified, 10 (53%) participated in round 1. Following round 1, five intervention groups reached consensus. The intervention groups rated as convenient were alternative medicines and pharmacological agents. The inconvenient groups are mechanical aids, psychological interventions and surgical interventions. Seven intervention groups were carried into round 2.

5.7.1.1.2 Round 2

Nine participants completed round 2, where the panellists reached consensus about one intervention group that was convenient to access: pharmacological interventions - prescription only medicine. Four other intervention groups did not reach consensus (electrotherapies, exercise, injectable agents, other physical therapies, self-management & education, weight management). The level of consensus achieved following round 1 and 2 are summarised in Table 15.

5.7.1.2Concluding remarks about the accessibility of interventions

Alternative medicines, pharmacological agents and prescription-only medicines were considered by the eDelphi panellists to be convenient to access, while mechanical aids, psychological interventions and surgical interventions were considered by the panellists to be inconvenient to access. The panellists did not reach consensus about the following intervention classes, therefore, they were rated as neither convenient or inconvenient to access: electrotherapies, exercise, injectable agents, other physical therapies, self-management and education and weight management.

Table 14. Interventions clustered into 12 groups for rating their accessibility in New Zealand by the Delphi panel

Intervention Group/ Class	Intervention
Alternative medicines	Avocado-soybean unsaponifiables, boswellia serrata extract, curcuma/curcuminoid, pycnogenol, glucosamine, chondroitin, glucosamine and chondroitin in compound form, vitamin D, omega-3 fatty acids, collagen, methylsulfonylmethane
Electrotherapies	Pulsed electromagnetic/ shortwave therapy, other electrotherapy - (laser, shock wave, interferential), transcutaneous electrical nerve stimulation (TENS), therapeutic ultrasound, acupuncture (traditional, electro acupuncture and laser acupuncture)
Exercise interventions	Land-based exercise including: muscle strengthening, range of motion, aerobic conditioning, neuromuscular balance, cycling, Tai Chi, yoga, aquatic exercise or hydrotherapy
Injectable agents	Viscosupplementation injection, platelet-rich plasma (PRP) injection, stem cell therapy, dextrose prolotherapy, fibroblast growth factor (FGF), corticosteroid injection
Mechanical aids and devices	Knee braces (varus unloading/re-alignment braces, valgus unloading/re-alignment braces, re-aligning patellofemoral braces), shoe orthotics (shock absorbing insoles, wedged insoles, or arch supports), footwear (unloading shoes, minimalist footwear and rocker soled shoes), patellar taping, assistive walking devices (e.g. cane)
Pharmacological interventions (over the counter)	Paracetamol/acetaminophen, topical capsaicin, topical NSAIDs

Pharmacological interventions (prescription medicine only)	Interleukin-1 (IL-1) inhibitors, methotrexate, oral opioids, transdermal opioids, colchicine, anti-nerve growth factor (NGF), calcitonin, bisphosphonates, doxycycline, oral non-steroidal anti-inflammatory drugs (NSAIDs) including COX-2 inhibitors, (diacerein, duloxetine, strontium ranelate - not available in the New Zealand Universal List of Medicines)
Psychological interventions	Cognitive behavioural therapy
Other physical therapies	Manual therapy (massage, mobilisation and manipulation)
Self-management and education interventions	self-management and education, heat and cold therapy
Surgical interventions	Arthroscopic lavage and debridement, arthroscopic meniscectomy, arthroscopic cartilage repair, total joint replacement
Weight management	Weight management

Table 15. Level of consensus reached on the 12 intervention groups after rounds 1 and 2 in the context of providing, funding, planning or delivering interventions within the NZ public health system

	Median response	**% Response***				
Round 1 (n=10)		**Inconvenient†**	**Neither convenient or inconvenient**	**Convenient†**	**Do not know**	**Total**
Electrotherapies	3	20%	40%	30%	10%	100%
Exercise interventions	4	20%	30%	50%	0%	100%
Injectable agents	3	50%	10%	0%	40%	100%
Pharmacological interventions - prescription only medicine	4	20%	30%	50%	0%	100%
Other physical therapies	4	30%	20%	40%	10%	100%
Self-management & education interventions	3	40%	20%	40%	0%	100%
Weight management	2	60%	20%	20%	0%	100%
Alternative medicines	4	0%	30%	**70%**	0%	100%
Mechanical aids and devices	2	**70%**	10%	10%	10%	100%
Pharmacological interventions - over the counter	5	0%	10%	**80%**	10%	100%
Psychological interventions	2	**100%**	0%	0%	0%	100%
Surgical interventions	2	**100%**	0%	0%	0%	100%
Round 2(n=9)						
Electrotherapies	3	22%	33%	22%	22%	100%
Exercise interventions	4	22%	22%	44%	11%	100%
Injectable agents	6	33%	11%	0%	56%	100%
Pharmacological interventions - prescription only medicine	4	11%	11%	**78%**	0%	100%
Other physical therapies	3	44%	22%	22%	11%	100%

Self-management & education interventions	3	22%	44%	33%	0%	100%
Weight management	2	67%	22%	11%	0%	100%

*consensus was defined as ≥70% convenient or inconvenient travel, or wait time; †very inconvenient or inconvenient, and very convenient or convenient have been combined for brevity. 5= very inconvenient travel, or long wait time (>3 months); 4= inconvenient travel, or long wait time; 3= neither convenient or inconvenient travel, or wait time; 2= convenient travel, or wait time; 1= very convenient travel, or wait time (<1 week). 6= I don't know

5.7.2 Cost

The total cost of delivering interventions was not reported in the RACGP CPG.[13] Cost data for the interventions were sourced, where possible, from NZ sources, using methods previously described.[407] Assumptions and resource costs are presented in Table 16. Prescription and therapeutic interventions were cross-checked with the New Zealand Pharmaceutical Schedule (www.pharmac.govt.nz) for government pricing. The New Zealand Formulary (www.nzformulary.org) informed the average monthly (30-day) dosing costs, or product recommended retail prices of the pharmacological interventions. Cost data for other interventions were extracted from NZ health authority sources, and were adjusted for inflation and/or currency using recommendations by Welte et al. (2004)[408] for NZ$2017 prices. If an intervention was unavailable in NZ, a 'medium' performance rating was assigned, accounting for off-label costs and the potential for additional travel costs (e.g. flights). We assumed 'soft consumables' such as heat pads had a lifespan of 6 months, and 'hard consumables' such as assistive walking devices had a lifespan of 5 years.

Table 16. Assumptions used to estimate the total cost of interventions in New Zealand

Assumptions
Soft consumables last 6 months
Hard consumables last 5 years
Resource costs from Pinto et al. (2013)[367] in NZ$2017:
Physiotherapy secondary care $84
Physiotherapy visit $72 ($145/hr)
Physiotherapy group $20/hr
GP visit $70 ($280/hr)
GP group $140/hr
Physiotherapy pool visit $2.83
Massage visit $59 ($118/hr)
Practice nurse visit $17 ($68/hr)
Practice nurse group $34/hr
Dietician, psychologist, other visit $74
Dispensing cost = $5.00 for 3 months
Drug dosing schedules were informed by the NZ Formulary or studies reported in the RACGP CPG
Drug prices are RRP or informed by the NZ PHARMAC schedule

5.7.3 Duration

Duration of the intervention effect was calculated as the mean length of follow-up in trials demonstrating a meaningful positive effect in the RACGP CPG.[13] The point estimate of the standardised mean difference (SMD) effect size was considered irrespective of the confidence interval to estimate the duration of intervention effect. If there was an effect, individual studies were reviewed to extract the follow up interval in the studies which reported an effect. These follow-up times were averaged; the mean follow-up duration for each intervention was used to inform the performance ratings for the Duration criterion.

5.7.4 Effectiveness

The intervention effectiveness criterion was measured using the Cohen's *d*,[364] a widely used statistic to measure the standardised difference between two means. Effect sizes of 0.2, 0.5 and 0.8 are considered low, medium and high levels of effect, respectively. The SMDs were extracted for effect on pain in the RACGP CPG.[13] If mean differences (MD) were reported, the statistical methods adopted by the Cochrane collaboration were used to calculate SMDs.[409] Furthermore, if an intervention was rated very low quality of evidence and the sum of the quality of the evidence criteria (risk of bias, inconsistency, indirectness and imprecision) exceeded the maximum 3 points needed to downgrade RCT evidence from high to very low (GRADE quality of evidence ranges from high, moderate, low and very low, and starts at high for randomised controlled trials) then the effectiveness rating of the intervention was downgraded an additional level to account for a 'floor effect' when rating the quality of evidence using GRADE.

5.7.5 Quality

The quality of evidence criterion was informed by the GRADE quality of evidence tables in the 2018 RACGP CPG.[13] GRADE rates the quality of evidence on 4 points ranging from high to very low (high, moderate and low and very low). These ratings were assigned to the equivalent levels of the Quality criterion. Quality of evidence is judged by: (1) risk of bias, (2) inconsistency of results, (3) indirectness of evidence, (4) imprecision and (5) other considerations.[359]

5.7.6 Recommendation

In this thesis, the recommendation for using an OA intervention was considered at first-, second- and third-line OA care which required that 3 evidence summaries, or performance matrices, were generated for each 'line' of OA care. A performance matrix is a summary table where the performance ratings on each of the criteria are shown for each of the interventions. The recommendation rating was mapped to the recommendation for each intervention reported RACGP CPG,[13] such that an intervention could receive one of 5 possible recommendations (in the preceding sentences). However, the RACGP CPG recommendations did not explicitly consider the timing of intervention in their recommendations (i.e., for first-, second- and third-line OA care), which this criterion concerns. Therefore, a 'timing rubric' (SUPPLEMENT 5) for first-, second- and third-line OA care was developed. The rubric systematically adjusted the level of recommendation given to the interventions. First, a score ranging from 5 to 1 is assigned to each of the 5 levels "strong for", "conditional for", "conditional (neutral)", "conditional against" and" strong against" depending on if they were explicitly mentioned as the following in the RACGP CPG knee algorithm (p.65[13]): 1 core interventions, 2) adjunct optional and advanced interventions or 3) severe/end-stage recommended surgical interventions. Next, the following scores were applied to the interventions in the knee algorithm: (i) core interventions were not penalised at any stage of OA care; a "strong for" recommendation was assigned a score of 5, to the "strong against" recommendation which was assigned a score of 1; (ii) optional adjunct and advanced interventions were penalised for first-line care and third-line care: a "strong for" recommendation received a score of 4, whilst "conditional against" and "strong against" were assigned a combined score of 1; no adjustments were made for second-line care (a "strong for" recommendation receives a score of 5) and (iii) surgical interventions were not penalised for third-line care, but penalised at first- and second-line care; "strong for" and "conditional for" recommendations were combined and assigned a score of 3, while "conditional against" and "strong against" were combined and assigned a score of 1. Thus, 3 intervention matrices were produced relative to the intervention recommendation for first-, second- and third-line care.

5.7.7 Risk-Mild and Risk-Serious

Harms in the RACGP CPG[13] were reported using the classification by Aronson and Ferner (2005).[410] To score the "risk of mild to moderate side-effects" and "risk of serious harm" criteria, the RACGP expert panel statements "no adverse events reported," "very low [or] low risk of harm," and "very low [or] low likelihood of serious adverse effects" were transposed to low levels of mild or moderate side-effects and serious harm. Where these statements were absent in the CPG, or "few adverse events" was mentioned in the CPG, we referred to the absolute risk ratings reported in the CPG technical document. Absolute risk refers to the likelihood of the outcome occurring.[411] For each intervention, if there were no harms specifically reported for mild or moderate side effects, or serious harms, the intervention was assigned the lowest score for mild or moderate side-effects and serious harms.

5.7.8 Total knee replacement

The performance of total knee replacement (TKR) was not assessed in the RACGP CPG.[13] Given the intention to create a generalisable MCDA framework, and high utilisation rates of TKR in NZ and globally[21, 152, 412] the performance of TKR was assessed on each of the criteria. The following sections describe how the GRADE approach[363] was used to assess the performance of TKR.

5.7.8.1 Accessibility of TKR

The performance of TKR on the Accessibility criterion was consistent with the approach described in Section 5.7.1.

5.7.8.2 Cost of TKR

The cost of the TKR was informed by the NZ private[245] and public[367] estimated cost of intervention.

5.7.8.3 Duration of the TKR effect

Survival rates of joint replacement reported in the 2017 New Zealand National Joint Registry were extracted to estimate the duration of the intervention effect.[365]

5.7.8.4 Effectiveness of TKR

Effectiveness was scored using the SMD reported in the systematic review by Shan et al. (2015).[413]

5.7.8.5 Quality of evidence GRADE evaluation for TKR

TKR was not evaluated in the RACGP CPG. Therefore, the quality of evidence was judged by myself and thesis supervisors for first-, second- and third-line OA care using GRADE[363] to derive "low" quality of evidence scores for first- and second-line care, and "high" quality of evidence score for third-line care, based on the systematic reviews for hip and knee replacement by Shan et al. (2014 & 2015).[413, 414] Justification for the quality of evidence scores we selected for first-, second- and third-line care are summarised according to the GRADE criteria in Table 17 and Table 18.

Table 17. Quality of evidence assessment for third-line OA care

Paper(s)	Hip: Shan, L., et al. (2014). "Total hip replacement: a systematic review and meta-analysis on mid-term quality of life." Osteoarthritis and Cartilage **22**(3): 389-406.	Knee: Shan, L., et al. (2015). "Intermediate and long-term quality of life after total knee replacement: a systematic review and meta-analysis." J Bone Joint Surg Am **97**(2): 156-168.
Study design/ Risk of Bias	Low – starts at low according to the guidance by GRADE[363]	Low - starts at low according to the guidance by GRADE[363]
Inconsistency	N/A – judged as low in the paper	N/A – judged as low in the paper
Indirectness	N/A – studies only considered THR	N/A – studies only considered TKR
Imprecision	N/A – confidence intervals suitably narrow and above the null	N/A– confidence intervals suitably narrow and above the null
Publication bias	N/A – stated as low levels in the paper	Unclear – not reported
Effect size	**+2** as the effect size was greater than Cohen's *d* of 0.8	**+2** as the effect size was greater than Cohen's *d* of 0.8
Overall score for quality of the evidence	**Low + 2 points = High**	**Low + 2 points = High**

Table 18. Quality of evidence assessment for first- and second-line OA care

	Hip: Shan, L., et al. (2014). "Total hip replacement: a systematic review and meta-analysis on mid-term quality of life." Osteoarthritis and Cartilage **22**(3): 389-406.	Knee: Shan, L., et al. (2015). "Intermediate and long-term quality of life after total knee replacement: a systematic review and meta-analysis." J Bone Joint Surg Am **97**(2): 156-168.
Study design/ Risk of bias	Low – starts at low according to the guidance by GRADE[363]	Low - starts at low according to the guidance by GRADE[363]
Inconsistency	N/A – judged as low in the paper	N/A – judged as low in the paper
Indirectness	N/A – studies only considered THR	N/A – studies only considered TKR

Imprecision	**-1** as there is poor evidence available for first-line OA care	**-1** as there is poor evidence available for first-line OA care
Publication bias	N/A – stated as low levels in the paper	Unclear – not reported
Effect size	N/A on the basis of imprecision	N/A on the basis of imprecision
Overall score for quality of the evidence	**Low -1 point = Very low**	**Low -1 point = Very low**

5.7.8.6Recommendation GRADE evaluation for TKR

The RACGP CPG treatment algorithm[13] specifically referred to total joint replacement for third-line care. Following the GRADE approach[359] the research team (myself and thesis supervisors) calculated that the criteria level for recommending the intervention for first-line care should be "conditional against", for second-line OA care "neutral" and "strong for" for third-line care, based on the two systematic reviews available for hip and knee joint replacement.[413, 415] The GRADE evaluation is summarised in Table 19.

Table 19. Total joint replacement GRADE recommendation assessment for first-, second- and third-line OA care

OA care phase	**Recommendation for total knee replacement†**
First-line	Conditional against – 1) There's no evidence to warrant a strong for or a strong against rating, thus leaving neutral, and conditional for/against recommendation; 2) with lack of expert panel judgement, we are left with a neutral rating; 3) reviewing the literature either conditional for/against recommendation, and with agreement among HA, AB & JC, it made more sense to rate as conditional against than conditional for.
Second-line	Neutral – Radiographically patients with KL scores of 3 who have joint replacement early have better outcomes, however those who have joint replacement at the very late/severe time of OA have greater overall improvement (because their level of deterioration is relatively greater than those who had an earlier joint replacement). Agreement between HA, AB & JC is that the decision for, or against, is neutral given no expert panel judgement for a for or against (either strongly or conditionally) recommendation.
Third-line	Strong for – Total joint replacement is the only intervention explicitly stated in the RACGP for third-line OA care.

†Same for total hip replacement

5.7.8.7 Risk of mild to moderate side effects and risk of serious harm

To rate the risk of harms, the 2017 NZ National Joint Registry report[365] was referred to for total knee replacement revision rates. We also referred to unpublished public hospital total hip and knee replacement rates of: infection, emergency readmission (e.g. complications of internal orthopaedic prosthetic devices, implants and grafts, infection and inflammatory reaction due to internal joint prosthesis, mechanical complications of internal joint prosthesis and wound infection), phlebitis and thrombophlebitis of other deep vessels of lower extremities; and other complications including cardiac abnormalities, acute respiratory distress syndrome (ARDS), respiratory failure and pulmonary collapse and gastrointestinal bleeding.

5.7.9 Section Summary

After measuring the performance of the alternatives on each of the criteria, the criteria are assigned performance ratings, according to the performance level associated with a given alternative's level of achievement on each of the criteria. To measure the performance and assign ratings to the OA interventions, evidence and recommendations were extracted from the 2018 RACGP CPG for OA.[13] The Accessibility and Cost criteria were not included in the RACGP CPG, such that a nationally representative panel of OA experts were convened to evaluate intervention accessibility in a 2-round eDelphi; local cost data were gathered to inform the Cost criterion. The performance of TKR on each of the criteria was evaluated using the GRADE approach.

5.8 SCORING THE INTERVENTIONS AND RANKING THEM

After generating intervention scores and preference weights, prioritising the interventions involves calculating their ‘total scores’,[297]‘total value’[290] or ‘value index’.[35] This enables scores and weights to be combined that are consistent with stakeholders’ preferences. The most commonly used approach to aggregate scores and weights is to use the weighted-sum model[290, 297] (also referred to in the literature as additive, linear, scoring, point-count and points models (or systems), or in MCDA literature, additive multi-attribute value models).[297] The formula for the weighted-sum model is seen in the formula below where V_j is the overall value of intervention J, S_{ij} is the score for intervention *j* on criterion *i*, and W_i is the weight attached to criterion *i*:

$$\mathrm{V}_j = \sum_{i=1}^{n} \mathrm{S}_{ij} \cdot \mathrm{W}_i$$

In essence, interventions’ scores (the level of achievement) on the criteria are multiplied by the preference weights and then these weighed scores are summed across the criteria to get a total score for each intervention (ranging from 0 to 100), thus revealing the relative importance of interventions.

Another, arguably more intuitive way to represent the weighted-sum function above is to use what is called a ‘points system’. In the points system each criterion is categorised into mutually-exclusive and exhaustive performance categories (e.g. high, medium, low) which are assigned ‘point values’ that represent the combined effect of the criterion’s relative importance (between-criteria preference weights) and its degree of achievement as reflected by the category (within-criterion preferences).[297] This is the approach that the previous sections in this thesis has followed.

Introduced in Section 5.3.1, weights for the criteria sum to 100 points (or per cent) at their 'best' or maximum performance rating. Using the points system, each alternative is scored on the point values for each criterion, which are then summed to produce a total score for each intervention (in contrast to multiplying the interventions' performance by the weights to produce a weighted score which is then summed to produce a total score). This thesis uses the more intuitive points system to express the weighted-sum model.

A key requirement of the weighted-sum model is that the criteria are preferentially independent (for example, preference for one criterion should not depend on another). If preference-independence is present, other aggregation functions can be used to model the interaction, such as multiplicative models (e.g. outranking and goal programming). However, these models are adopted less often due their complexity, making them hard to interpret and difficult to populate with data, which may explain their poor adoption.[290, 297] Furthermore, incorrectly specifying such models can have considerably greater errors than additive models as evidenced by empirical data.[416]

5.9 EVALUATING UNCERTAINTY

Evaluating uncertainty is useful for assessing the confidence in decision and the impact of collecting more information to further inform the decision, if possible. For decision-makers, evaluating uncertainty is a key aspect of the decision-making process[417] and is considered best-practice reporting the results of MCDA.[290, 418]

One-way deterministic sensitivity analysis is often used to explore uncertainty in the point estimate of alternatives' total scores.[268] This type of analysis involves exploring the effect of 'manually' changing the performance-rating of the criterion ±1-level on the total score (point estimate) of an alternative. Tornado diagrams are useful tools for communicating uncertainty using one-way deterministic sensitivity analysis. Other sensitivity analyses are possible which allow more than one level to be altered at a time, such as probabilistic sensitivity analyses.[418] However, this approach requires information about the distribution of intervention performance on all the criteria, which is not available. Hence, the one-way deterministic sensitivity analysis was used in this thesis. Interested readers are directed to Briggs et al. (2012)[418] for a more information about uncertainty analyses.

5.10 CHAPTER SUMMARY

This chapter described the theory around planning and pilot-testing used to develop and model stakeholders' preferences for alternatives using MCDA. The key properties of criteria, along with the method used to weight them was discussed. A description of statistical methods for analysing preference weights was introduced. Next, the weighted-sum model was explained to demonstrate how interventions are evaluated and prioritised. The chapter concluded by describing how uncertainty in the estimates of intervention total scores can be assessed using one-way deterministic sensitivity analysis.

SUPPLEMENT 5

The following 'timing rubric' (full view) adjusts the level of recommendation for each intervention to first-, second- and third-line interventions. The images on the following pages show the zoomed in view of the first- and second-line algorithm, followed by the third-line algorithm.

1st line interventions

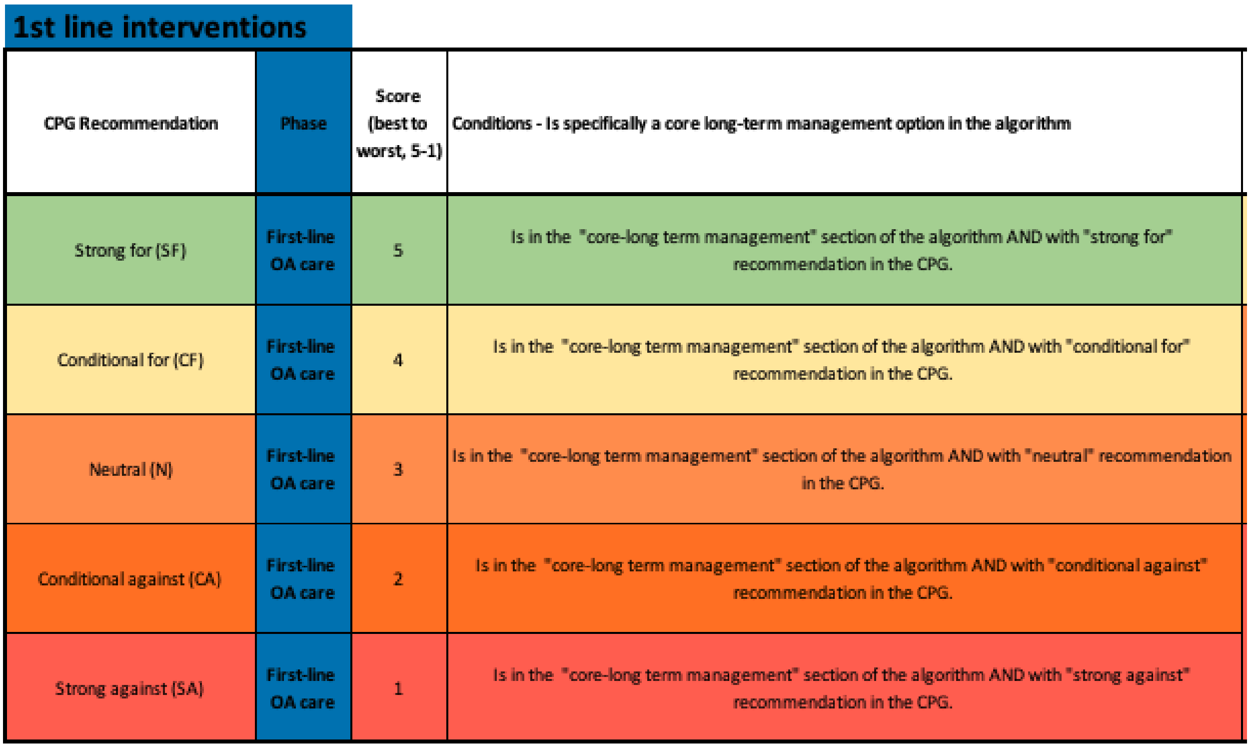
CPG Recommendation	Phase	Score (best to worst, 5-1)	Conditions - Is specifically a core long-term management option in the algorithm	Score (best to worst, 5-1)	Conditions - is specifically an optional adjunctive management option in the algorithm	Score (best to worst, 5-1)	Conditions - NOT explicit in the algorithm AND NOT a surgical option	Score (best to worst, 5-1)	Conditions - must be a surgical or end-stage intervention
Strong for (SF)	First-line OA care	5	Is in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.	4	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "strong for" recommendation in the CPG.	3	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "strong for" recommendation in the CPG.	3	Is in the "Surgical management of knee OA" section of the algorithm AND considered an "end-stage OA" intervention, with a "strong for" OR "conditional for" recommendation.
Conditional for (CF)	First-line OA care	4	Is in the "core-long term management" section of the algorithm AND with "conditional for" recommendation in the CPG.	3	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional for" recommendation in the CPG.		Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional for" recommendation in the CPG.		
Neutral (N)	First-line OA care	3	Is in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.	2	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "neutral" recommendation in the CPG.	2	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "neutral" recommendation in the CPG.	2	Is in the "Surgical management of knee OA" section of the algorithm AND considered an "end-stage OA" intervention, with a "neutral" recommendation.
Conditional against (CA)	First-line OA care	2	Is in the "core-long term management" section of the algorithm AND with "conditional against" recommendation in the CPG.	1	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional against" OR "strong against" recommendation in the CPG.	1	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional against" OR "strong against" recommendation in the CPG.	1	Is in the "Surgical management of knee OA" section of the algorithm AND considered an "end-stage OA" intervention, with a "conditional against" OR "strong against" recommendation.
Strong against (SA)	First-line OA care	1	Is in the "core-long term management" section of the algorithm AND with "strong against" recommendation in the CPG.						

2nd line interventions

CPG Recommendation	Phase	Score (best to worst, 5-1)	Conditions - Is an optional adjunctive management option in the algorithm	Score (best to worst, 5-1)	Conditions - is a core long-term management option in the algorithm	Score (best to worst, 5-1)	Conditions - NOT explicit in the algorithm as an optional adjunctive option, OR a core long-term management option	Score (best to worst, 5-1)	Conditions - must be a surgical or end-stage intervention
SF	Second-line OA care	5	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "strong for" recommendation in the CPG.	5	Is in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.	3	Is NOT in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.	3	Is in the "Surgical management of knee OA" section of the algorithm AND considered an "end-stage OA" intervention, with a "strong for" OR "conditional for" recommendation.
CF	Second-line OA care	4	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with conditional for" recommendation in the CPG.	4	Is in the "core-long term management" section of the algorithm AND with "conditional for" recommendation in the CPG.				
N	Second-line OA care	3	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "nautral" recommendation in the CPG.	3	Is in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.	2	Is NOT in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.	2	Is in the "Surgical management of knee OA" section of the algorithm AND considered an "end-stage OA" intervention, with a "neutral" recommendation.
CA	Second-line OA care	2	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "conditional against" recommendation in the CPG.	2	Is in the "core-long term management" section of the algorithm AND with "conditional against" recommendation in the CPG.	1	Is NOT in the "core-long term management" section of the algorithm AND with "conditional against" OR "strong against" recommendation in the CPG.	1	Is in the "Surgical management of knee OA" section of the algorithm AND considered an "end-stage OA" intervention, with a "conditional against" OR "strong against" recommendation.
SA	Second-line OA care	1	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "strong against" recommendation in the CPG.	1	Is in the "core-long term management" section of the algorithm AND with "strong against" recommendation in the CPG.				

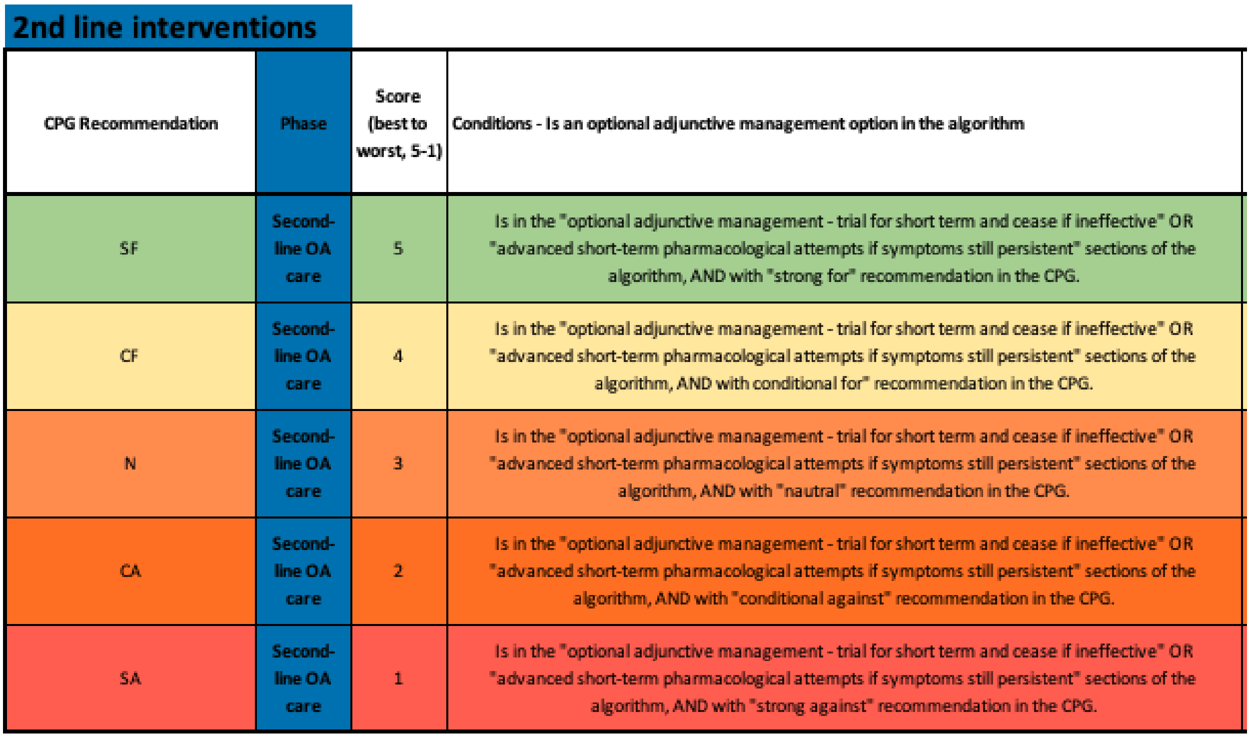
3rd line interventions

CPG Recommendation	Phase	Score (best to worst, 5-1)	Conditions - is total knee or hip replacement OR a surgical intervention	Score (best to worst, 5-1)	Conditions - Is explicitly stated as a core intervention in the algorithm	Score (best to worst, 5-1)	Conditions - is an adjunctive intervention explicitly stated in the algorithm	Score (best to worst, 5-1)	Conditions - is not explicitly stated in the algorithm as a core intervention option or optional adjunctive intervention option.
SF	Third-line OA care	5	Is in the "end-stage OA" section of the algorithm AND with a "strong for" recommendation.	5	Is in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.	4	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "strong for" recommendation in the CPG.	3	Is NOT specifically in the algorithm as a core or optional adjunct intervention AND is a "strong for" or "conditional for" recommendation in the CPG.
CF	Third-line OA care	4	Is in the "end-stage OA" section of the algorithm AND with a "conditional for" recommendation.	4	Is in the "core-long term management" section of the algorithm AND with "conditional for" recommendation in the CPG.	3	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "conditional for" recommendation in the CPG.		
N	Third-line OA care	3	Is in the "end-stage OA" section of the algorithm AND with a "neutral" recommendation.	3	Is in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.	2	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "neutral" recommendation in the CPG.	2	Is NOT specifically in the algorithm as a core or optional adjunct intervention AND is a "neutral" recommendation in the CPG.
CA	Third-line OA care	2	Is in the "end-stage OA" section in the algorithm AND with a "conditional against" recommendation.	2	Is in the "core-long term management" section of the algorithm AND with "conditional against" recommendation in the CPG.	1	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "conditional against" OR "strong against" recommendation in the CPG.	1	Is NOT specifically in the algorithm as a core or optional adjunct intervention AND is a "strong against" or "conditional against" recommendation in the CPG.
SA	Third-line OA care	1	Is in the "end-stage OA" section of the algorithm AND with a "strong against" recommendation.	1	Is in the "core-long term management" section of the algorithm AND with "strong against" recommendation in the CPG.				



1st line interventions

CPG Recommendation	Phase	Score (best to worst, 5-1)	Conditions - Is specifically a core long-term management option in the algorithm
Strong for (SF)	First-line OA care	5	Is in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.
Conditional for (CF)	First-line OA care	4	Is in the "core-long term management" section of the algorithm AND with "conditional for" recommendation in the CPG.
Neutral (N)	First-line OA care	3	Is in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.
Conditional against (CA)	First-line OA care	2	Is in the "core-long term management" section of the algorithm AND with "conditional against" recommendation in the CPG.
Strong against (SA)	First-line OA care	1	Is in the "core-long term management" section of the algorithm AND with "strong against" recommendation in the CPG.



2nd line interventions

CPG Recommendation	Phase	Score (best to worst, 5-1)	Conditions - Is an optional adjunctive management option in the algorithm
SF	Second-line OA care	5	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "strong for" recommendation in the CPG.
CF	Second-line OA care	4	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with conditional for" recommendation in the CPG.
N	Second-line OA care	3	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "nautral" recommendation in the CPG.
CA	Second-line OA care	2	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "conditional against" recommendation in the CPG.
SA	Second-line OA care	1	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" sections of the algorithm, AND with "strong against" recommendation in the CPG.

Score (best to worst, 5-1)	**Conditions - is specifically an optional adjunctive management option in the algorithm**
4	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "strong for" recommendation in the CPG.
3	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional for" recommendation in the CPG.
2	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "neutral" recommendation in the CPG.
1	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional against" OR "strong against" recommendation in the CPG.

Score (best to worst, 5-1)	**Conditions - is a core long-term management option in the algorithm**
5	Is in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.
4	Is in the "core-long term management" section of the algorithm AND with "conditional for" recommendation in the CPG.
3	Is in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.
2	Is in the "core-long term management" section of the algorithm AND with "conditional against" recommendation in the CPG.
1	Is in the "core-long term management" section of the algorithm AND with "strong against" recommendation in the CPG.

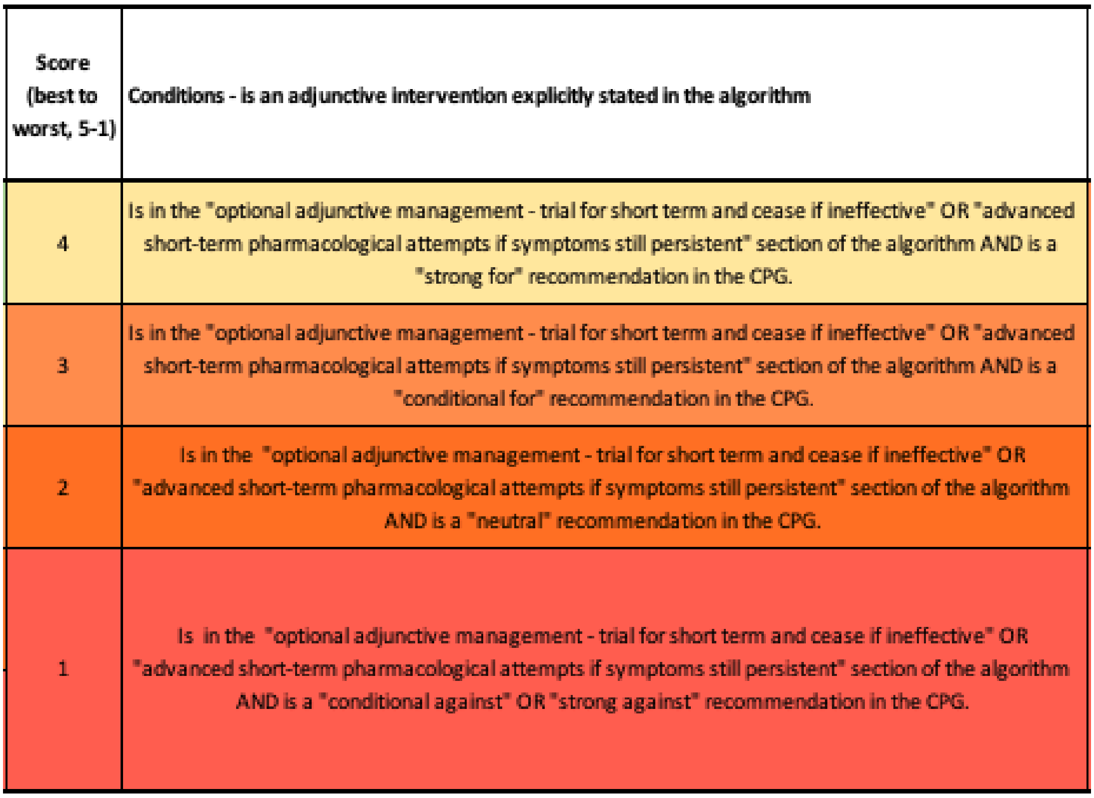
Score (best to worst, 5-1)	Conditions - NOT explicit in the algorithm AND NOT a surgical option
3	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "strong for" recommendation in the CPG.
	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional for" recommendation in the CPG.
2	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "neutral" recommendation in the CPG.
1	Is NOT in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm, AND with "conditional against" OR "strong against" recommendation in the CPG.

Score (best to worst, 5-1)	Conditions - NOT explicit in the algorithm as an optional adjunctive option, OR a core long-term management option
3	Is NOT in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.
2	Is NOT in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.
1	Is NOT in the "core-long term management" section of the algorithm AND with "conditional against" OR "strong against" recommendation in the CPG.

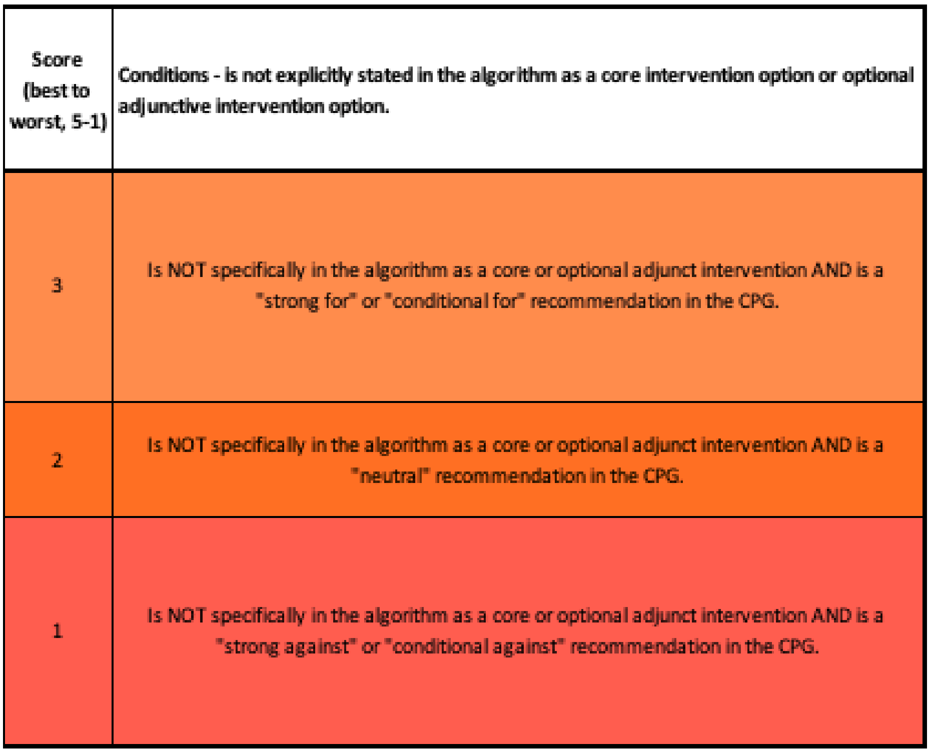
3rd line interventions

CPG Recommendation	Phase	Score (best to worst, 5-1)	Conditions - is total knee or hip replacement OR a surgical intervention
SF	Third-line OA care	5	Is in the "end-stage OA" section of the algorithm AND with a "strong for" recommendation.
CF	Third-line OA care	4	Is in the "end-stage OA" section of the algorithm AND with a "conditional for" recommendation.
N	Third-line OA care	3	Is in the "end-stage OA" section of the algorithm AND with a "neutral" recommendation.
CA	Third-line OA care	2	Is in the "end-stage OA" section in the algorithm AND with a "conditional against" recommendation.
SA	Third-line OA care	1	Is in the "end-stage OA" section of the algorithm AND with a "strong against" recommendation.

Score (best to worst, 5-1)	Conditions - Is explicitly stated as a core intervention in the algorithm
5	Is in the "core-long term management" section of the algorithm AND with "strong for" recommendation in the CPG.
4	Is in the "core-long term management" section of the algorithm AND with "conditional for" recommendation in the CPG.
3	Is in the "core-long term management" section of the algorithm AND with "neutral" recommendation in the CPG.
2	Is in the "core-long term management" section of the algorithm AND with "conditional against" recommendation in the CPG.
1	Is in the "core-long term management" section of the algorithm AND with "strong against" recommendation in the CPG.



Score (best to worst, 5-1)	Conditions - is an adjunctive intervention explicitly stated in the algorithm
4	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "strong for" recommendation in the CPG.
3	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "conditional for" recommendation in the CPG.
2	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "neutral" recommendation in the CPG.
1	Is in the "optional adjunctive management - trial for short term and cease if ineffective" OR "advanced short-term pharmacological attempts if symptoms still persistent" section of the algorithm AND is a "conditional against" OR "strong against" recommendation in the CPG.



Score (best to worst, 5-1)	Conditions - is not explicitly stated in the algorithm as a core intervention option or optional adjunctive intervention option.
3	Is NOT specifically in the algorithm as a core or optional adjunct intervention AND is a "strong for" or "conditional for" recommendation in the CPG.
2	Is NOT specifically in the algorithm as a core or optional adjunct intervention AND is a "neutral" recommendation in the CPG.
1	Is NOT specifically in the algorithm as a core or optional adjunct intervention AND is a "strong against" or "conditional against" recommendation in the CPG.

CHAPTER 6: STAKEHOLDERS' PREFERENCES FOR OSTEOARTHRITIS INTERVENTIONS IN HEALTH SERVICES: A CROSS-SECTIONAL STUDY USING MULTI-CRITERIA DECISION ANALYSIS

The original research in this chapter is the application of methods described in the previous chapter and has been published in Osteoarthritis and Cartilage Open[42] and is included as SUPPLEMENT 6 with permission from Elsevier as a thesis work. As the primary author for this article I am responsible for the study design, acquisition of data, analysis, interpretation, manuscript drafting, response to reviewers and approved the submitted version of the manuscript. Supervisor and co-author contributions are acknowledged in the initial study conceptualisation, planning and overall editorial aspects of manuscript preparation prior to the final article submission.

6.1 ABSTRACT

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.

6.2 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.[14, 202] However, the recommended 'core' interventions are not systematically delivered to or taken up by patients,[25, 258, 419-421] resulting in missed opportunities for potential health gains, a tendency to deliver low-value care and increased downstream health system costs without health gains.[22, 23] 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.[29, 260]

Stakeholders' preferences for health interventions,[422] especially in primary care settings,[27] play an important role in determining their uptake, highlighting the importance of widely engaging stakeholders in service co-design and care delivery recommendations.[29] 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.[31, 34, 261, 423-425] 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.[22, 23]

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.[426] 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.[36, 37] MCDA has been used to explore OA patients' preferences for physical activity,[38] patients' drug preferences,[274] and healthcare providers' treatment choices for people with OA.[282] 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.

6.3 METHODS

6.3.1 Design

This cross-sectional study followed six stages for conducting MCDA (Figure 17), aligned with MCDA good practice guidelines.[290] 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[427] (SUPPLEMENT 7).

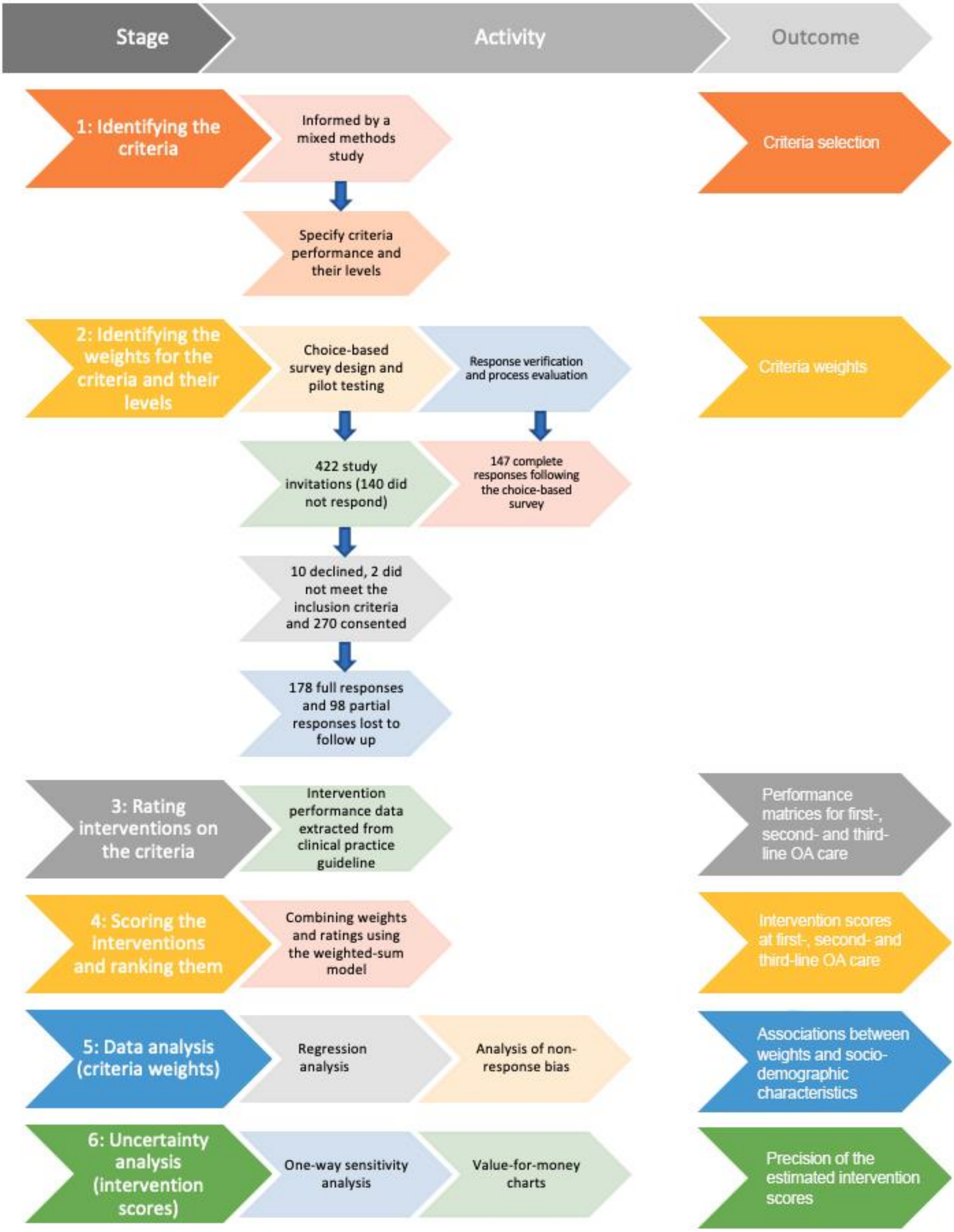


Figure 17. Flow diagram of the study by stage, primary activity and outcomes for each stage

6.3.2 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[56]), 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), Māori health (with an active interest in advocacy or consumer representation for Māori health) and OA researchers (having published at least one academic article related to OA). A characterisation of these groups is reported elsewhere[428] (Chapter 4 in this thesis).

Māori, the indigenous peoples of NZ, are recognised as a priority group with respect to enhancing healthcare equity and equality of health outcomes.[429] 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[146]) and socio-cultural characteristics.[260] The Qualtrics platform (Provo, USA) was used to screen and collect participants' demographic data.

6.3.2.1 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 Māori 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.

6.3.3 Stage 1: Identifying the criteria and their levels for selecting OA interventions

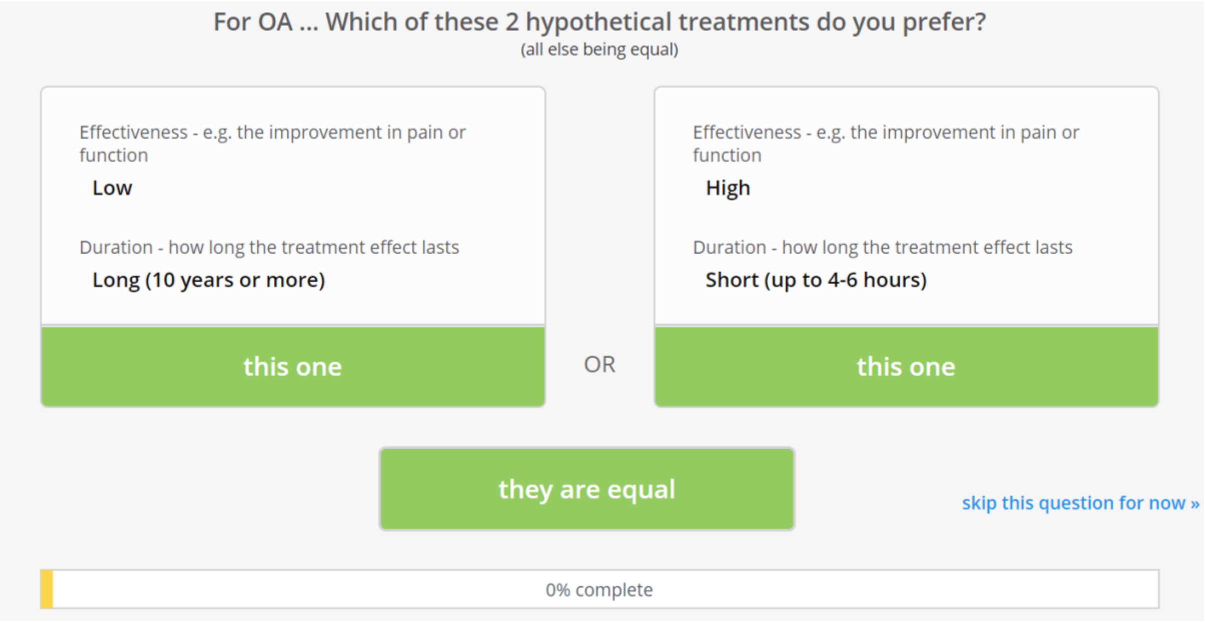
Stage 1 was informed by our earlier mixed-methods study[41] whereby multi-disciplinary and cross-sectoral stakeholders identified nine criteria influencing their choice or recommendation of OA interventions in the NZ health system. These criteria were: Accessibility (travel or wait time to access the 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.[41] 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 MCDAs, the mean number of criteria is eight[268]).

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 Chapter 5, Table 12.

6.3.4 Stage 2: Identifying the weights for the criteria and their levels

6.3.4.1 Choice-based survey

A choice-based survey administered by 1000minds software (www.1000minds.com) and implementing the PAPRIKA method[369] – 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 (Figure 18).


For OA ... Which of these 2 hypothetical treatments do you prefer?
(all else being equal)

Effectiveness - e.g. the improvement in pain or function
Low
Duration - how long the treatment effect lasts
Long (10 years or more)

Effectiveness - e.g. the improvement in pain or function
High
Duration - how long the treatment effect lasts
Short (up to 4-6 hours)

Figure 18. Example of the 1000minds pairwise-ranking question

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).[369]

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 both the relative importance of the criterion overall and the level's degree of achievement or performance on the criterion.[369] 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-second YouTube instructional video and a definition sheet for the criteria, were included with the survey (SUPPLEMENT 8). Participants were asked to complete the survey within two weeks, and reminders were sent to encourage completion.

6.3.4.2Pilot-testing

Before being launched, the survey and accompanying supplementary materials were pilot-tested with a convenience sample of 17 interviewees. The pilot-testing approach and amendments are included in SUPPLEMENT 9. The pilot-testing revealed that the survey instructions and criteria descriptions were required to improve the clarity of the survey instructions and the trade-offs between the hypothetical interventions.

6.3.4.3 Response verification and process evaluation

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. Immediately after completing the choice-based survey, each participant reviewed their own results and evaluated the extent to which their criteria weights aligned with their overall intuitive and holistic evaluations of the relative importance of the criteria, expressed on a single 7-point Likert scale (strongly agree (1), agree, somewhat agree, neither agree or disagree, somewhat disagree, disagree, strongly disagree (7)). Participants' overall perceived expectations about the relative importance of the criteria were evaluated as the group median response and percent agreement across the response categories 'strongly agree', 'agree' and 'somewhat agree'. We also examined if participant response consistency meaningfully changed the criteria weights if all three or none of the repeated questions were answered consistently.

A process evaluation was undertaken by asking participants to what extent: (1) the pairwise-ranking questions could be answered, and (2) the survey format was user-friendly (5-point Likert scale: very easy, easy, neither easy nor difficult, difficult, very difficult). Free text fields were available for participants to explain their reasons for their ratings. Participants' perceptions of the survey task and structure were evaluated as the median response and pooled proportion across the response categories 'very easy' and 'easy'.

6.3.5 Stage 3: Rating interventions on the criteria

6.3.5.1 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).[13] 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.[41] Cost was estimated using data and methods described in a systematic review.[408] 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.

6.3.5.2 Rating performances

Each intervention was rated on the criteria and summarised into three 'performance matrices' for first-, second- and third-line OA care.[430] 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).

6.3.6 Stage 4: Scoring the OA interventions and ranking them

A 'total score' for each intervention was calculated using a weighted-sum model[297]: 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,[13] our analysis hereinafter will focus on the 15 "recommended" interventions (p. 65): 3 first-line (core) interventions, 11 second-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.

6.3.7 Stage 5: Data analysis (criteria weights)

6.3.7.1 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.

6.3.7.2 Association with stakeholder group

To investigate if participants' weights on the criteria differed by stakeholder group, fractional multinomial logistic regression (FMNL)[400, 401, 431] 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.

6.3.7.3 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.

6.3.8 Stage 6: Uncertainty analysis (intervention scores)

6.3.8.1 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 11.

6.4 RESULTS

6.4.1 Stage 2: Identifying the criteria weights and process evaluation

6.4.1.1 Participants

Invitations were sent to 422 people, of whom 272 consented to participate; 178 (42.2%) completed the choice-based survey, and 147 of those completed the data verification and process evaluation. Their socio-demographic characteristics are summarised in Table 20.

6.4.1.2 Choice-based survey

The weights for the criteria and levels are reported in Table 21. 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 seconds per question and answered a mean of 45 questions (range 20-92) each in total (median 15 minutes in total).

6.4.1.2.1 Response verification and process evaluation

Of the 147 (82.6%) who completed the response verification, 83.0% strongly agreed, agreed, or somewhat agreed, that their weights reflected their overall assessment of the importance of the criteria (median=agree) (Table S11). The maximum difference between the mean weights for the overall sample and the mean weights of participants who answered all or none of the consistency questions correctly was at most 2.4%, suggesting no bias introduced by response inconsistency (Table S12).

For the process evaluation, 49.7% found it very easy or easy to use (median=neutral) and 48.9% found the survey task difficult or very difficult to complete (median=neutral) (Table S13). Comparing sub-group median responses to the group median response (Table S14) revealed that: (i) healthcare providers and OA researchers were less certain that their weights reflected their overall assessment of the importance of the criteria (median response=somewhat agree); (ii) policy-informants and OA researchers found the survey user friendly (median response=easy), and (iii) providers and policy-informants found the survey task more difficult (median response=difficult). Content analyses of the free-text responses did not reveal any over-arching themes.

Table 20. Socio-demographic characteristics of the choice-based survey participants (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	70 (39)	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 (13)	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

Table 21. Criteria definitions and their sample mean criteria weights produced from the choice-based survey (N=178), in decreasing order of relative importance. Criterion weights at their best performance level sum to 1 (or equivalently, 100%)

Criteria (most to least important) Performance levels (worst to best)	Definition	Full sample mean weight^ (n=178)	Mean weight^ by group			
			Consumers (n=63)	Providers (n=79)	Policy-informants (n=24)	OA Researchers (n=12)
Recommendation to use the intervention now	Recommendation for using the intervention at first-line OA care.					
Strong against		0.0	0.0	0.0	0.0	0.0
Conditional against		0.0647	0.0645	0.0618	0.0720	0.0697
Neutral*		0.1108	0.1073	0.1118	0.1116	0.1213
Conditional for		0.1529	0.1462	0.1581	0.1462	0.1678
Strong for		0.1904	0.1848	0.1947	0.1851	0.2038
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.					
Very low		0.0	0.0	0.0	0.0	0.0
Low		0.0587	0.0526	0.0607	0.0704	0.0546
Moderate		0.1319	0.1145	0.1377	0.1549	0.1389
High		0.1765	0.1560	0.1835	0.2100	0.1709
Effectiveness of the intervention	The clinical effect of the intervention on pain.					
Low ($d<0.2$)		0.0	0.0	0.0	0.0	0.0
Moderate ($d<0.5$)		0.0983	0.0865	0.1021	0.1205	0.0911
High ($d \geq 0.5$)		0.1501	0.1376	0.1559	0.1720	0.1335
Duration of the intervention effect	The duration of follow up demonstrating a meaningful effect on pain.					
Short (up to 6hrs)		0.0	0.0	0.0	0.0	0.0
Short-medium* (<3 months)		0.0392	0.0421	0.0371	0.0300	0.0561
Medium (3-12 months)		0.0748	0.0817	0.0704	0.0592	0.0993
Long (>12 months)		0.1318	0.1506	0.1218	0.1145	0.1339

Criteria (most to least important) Performance levels (worst to best)	Definition	Full sample mean weight^ (n=178)	Mean weight^ by group			
			Consumers (n=63)	Providers (n=79)	Policy-informants (n=24)	OA Researchers (n=12)
Risk of serious harm (**Risk-Serious**)	Treatment side-effects that have significant medical consequences, e.g., lead to death, permanent disability or prolonged hospitalisation.					
High (1 in 50 chance; >0.5%)		0.0	0.0	0.0	0.0	0.0
Medium (1 in 200 chance; 0.2%-0.5%)		0.0795	0.0864	0.0763	0.0701	0.0826
Low (1 in 500 chance; <0.2%)		0.1282	0.1325	0.1223	0.1179	0.1651
Risk of mild to moderate side effects (**Risk-Mild**)	Treatment side-effects that are not serious (see risk of serious harm).					
High (3 in 4 chance; >50%)		0.0	0.0	0.0	0.0	0.0
Medium (2 in 4 chance; 25-50%)		0.0527	0.0513	0.0610	0.0396	0.0319
Low (1 in 4 chance; <25%)		0.0941	0.0980	0.1016	0.0720	0.0686
Cost of the intervention	Total financial costs relevant to the use or provision of healthcare for OA.					
High (>$1000 per month or >$15,000 one-off)		0.0	0.0	0.0	0.0	0.0
Medium ($100-$1000 per month or $1500-$15,000 one-off)		0.0407	0.0450	0.0354	0.0513	0.0326
Low (<$100 per month or $1500 one-off)		0.0661	0.0735	0.0584	0.0760	0.0582
Accessibility to the intervention	The extent to which the intervention can be accessed by people with OA.					
Inconvenient travel, or wait time (>3 months)		0.0	0.0	0.0	0.0	0.0
Neither convenient or inconvenient travel, or wait time*		0.0313	0.0335	0.0309	0.0263	0.0331
Convenient travel, or wait time (<1 week)		0.0627	0.0670	0.0618	0.0526	0.0661

*Interpolated criterion level using a Bézier curve; *d*= Cohen's *d* for effect size; ^the weights, multiplied by 100, are equivalent to per cent points and at their best level sum to 1 (100%).

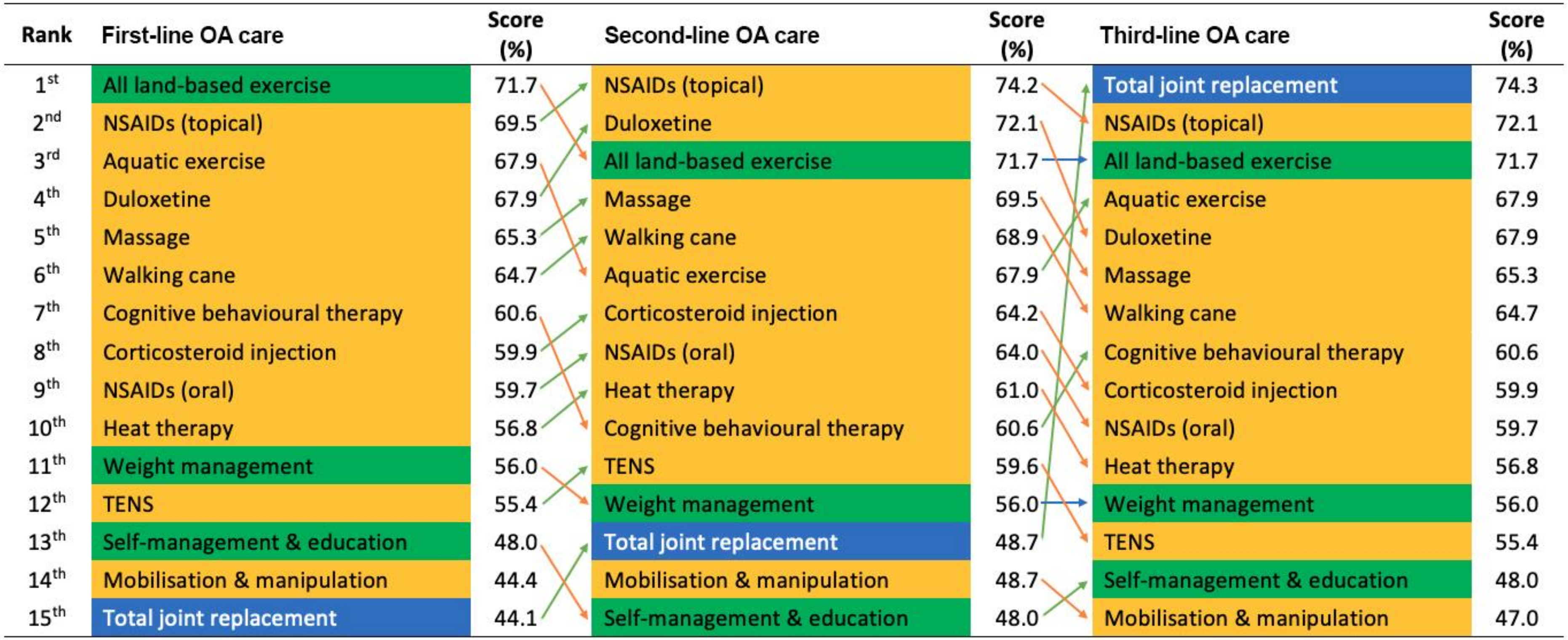
6.4.2 Stage 3: Rating interventions on the criteria

The results of the Delphi exercise and the GRADE evaluation are summarised in Chapter 5, Section 5.7.1. The assigned performance ratings across the criteria at each OA care phase are reported in the performance matrix in SUPPLEMENT 12.

6.4.3 Stage 4: Intervention scores and rankings

The total scores of the 15 guideline-recommended interventions are reported in Figure 19, 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 (Table S16), 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%), 'pycnogenol' (69.9%) and 'curcuma' (66.4%).



Rank	First-line OA care	Score (%)	Second-line OA care	Score (%)	Third-line OA care	Score (%)
1st	All land-based exercise	71.7	NSAIDs (topical)	74.2	Total joint replacement	74.3
2nd	NSAIDs (topical)	69.5	Duloxetine	72.1	NSAIDs (topical)	72.1
3rd	Aquatic exercise	67.9	All land-based exercise	71.7	All land-based exercise	71.7
4th	Duloxetine	67.9	Massage	69.5	Aquatic exercise	67.9
5th	Massage	65.3	Walking cane	68.9	Duloxetine	67.9
6th	Walking cane	64.7	Aquatic exercise	67.9	Massage	65.3
7th	Cognitive behavioural therapy	60.6	Corticosteroid injection	64.2	Walking cane	64.7
8th	Corticosteroid injection	59.9	NSAIDs (oral)	64.0	Cognitive behavioural therapy	60.6
9th	NSAIDs (oral)	59.7	Heat therapy	61.0	Corticosteroid injection	59.9
10th	Heat therapy	56.8	Cognitive behavioural therapy	60.6	NSAIDs (oral)	59.7
11th	Weight management	56.0	TENS	59.6	Heat therapy	56.8
12th	TENS	55.4	Weight management	56.0	Weight management	56.0
13th	Self-management & education	48.0	Total joint replacement	48.7	TENS	55.4
14th	Mobilisation & manipulation	44.4	Mobilisation & manipulation	48.7	Self-management & education	48.0
15th	Total joint replacement	44.1	Self-management & education	48.0	Mobilisation & manipulation	47.0

Key: Green = first-line (core) interventions; orange = second line (optional or advanced pharmacological) interventions and blue = third-line interventions (total joint replacement).

Figure 19. RACGP guideline recommended OA interventions (N=15) ranked by the full sample mean preference weights at first-, second- and third-line OA care.

6.4.4 Stage 5: Relationships between weights and stakeholder groups

6.4.4.1Regression 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.[41] Average partial effects (APEs) of the FMNL regression revealed weak evidence of associations between weights and stakeholder groups (Table 22). The APEs were relatively small after accounting for other socio-demographic characteristics (no more than 5.7%, aligning with the robustness check, SUPPLEMENT 12, Table S17), suggesting that weights did not differ meaningfully by stakeholder group (or within consumer or healthcare provider groups – see SUPPLEMENT 12, Tables S17-S19).

The level of agreement across groups by ranked interventions was very strong (N=75, W=0.990, p<0.000; (SUPPLEMENT 12, Table S21).

Table 22. 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	Recommendation 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	0.002	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	0.000	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)	(0.013)	(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

6.4.4.2 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 12, Table S22). 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 ($r_s = 1.00$, 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.

6.4.5 Stage 6: Uncertainty analysis

6.4.5.1 One-way sensitivity analysis

The uncertainty analyses at each phase of OA care (first-line care shown in Figure 20) 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 (disaggregated intervention scores are shown in SUPPLEMENT 12, Disaggregated intervention scores).

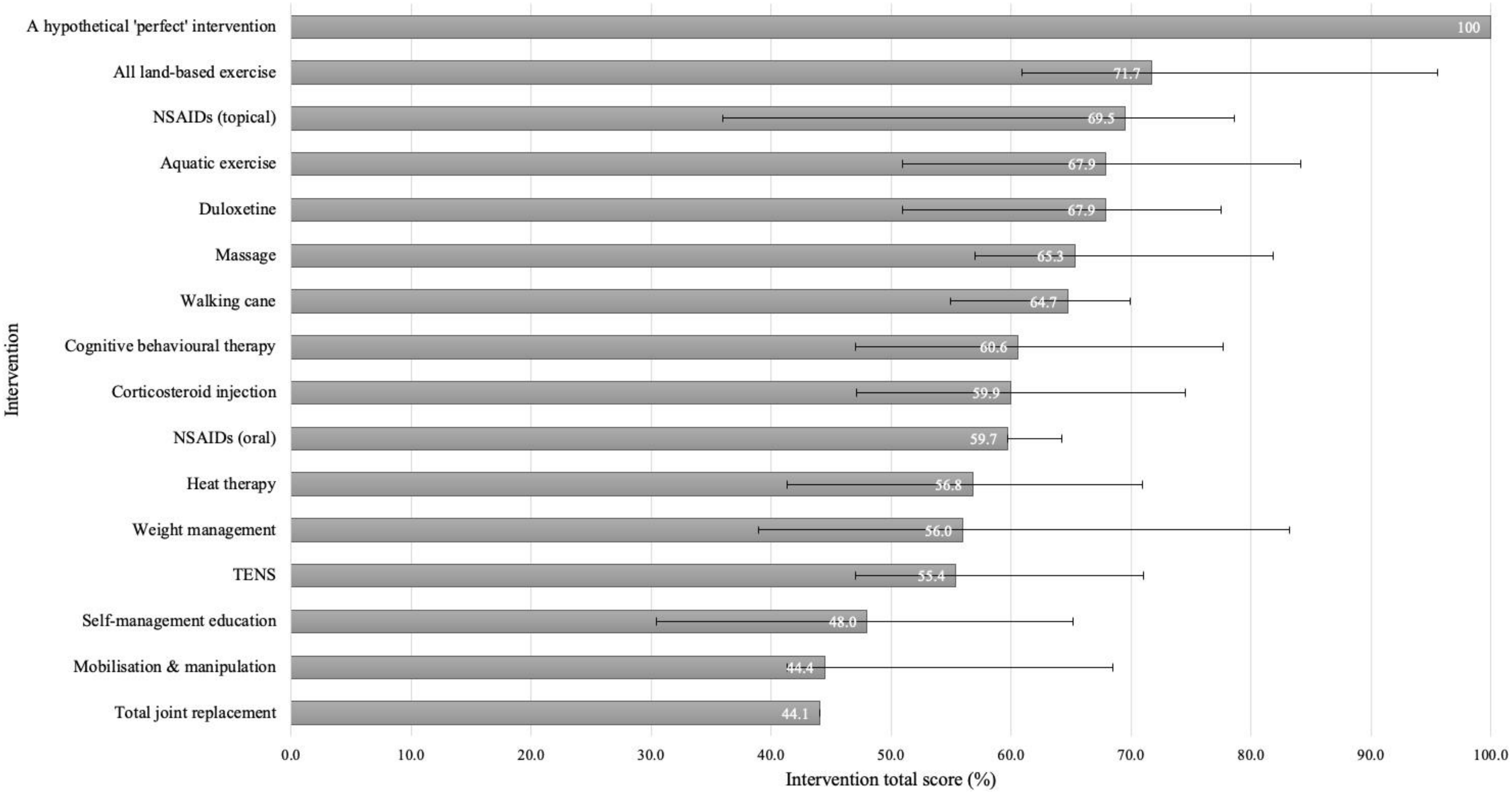


Figure 20. Error bars representing the aggregate uncertainty in the 15 guideline-recommended interventions' total scores across all performance ratings for first-line OA care.

6.5 DISCUSSION

This study has systematically combined the preferences of stakeholders for OA interventions with CPG recommendations[13] 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.[432] Participants valued Recommendation, Quality, Effectiveness, Duration and Risk-Serious approximately 2.5-3 times more than the two least important criteria, Cost and Accessibility. These findings suggest that stakeholders' are willing to forego intervention Cost and Accessibility in favour of superior performance on the other criteria. Ultimately, the weights show that the choice of OA interventions is influenced by some criteria more than others, yet these differences may not accurately reflect the complexity of their real-world implementation.

Weights on the criteria did not differ meaningfully by stakeholder group (or by subgroup, SUPPLEMENT 12). The regression analysis detected only small associations (≤5.7%; Table 22) between weights and stakeholder group. Accounting for variance in group size made virtually no difference to the relative importance of the scored interventions (r_s=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 (Figure 19) due to poor performance on the Quality and Effectiveness criteria (SUPPLEMENT 12 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.[433] Weight management and the application of active self-management strategies for OA require substantial behaviour change for patients, which is often challenging to sustain.[434] 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.[14, 31, 32] A number of studies have investigated consumer or provider preferences for the characteristics of OA interventions using MCDA methods.[38, 274, 277, 281, 282, 284] However, none has incorporated stakeholders' preferences across a health system. Broader engagement may lead to more effective implementation strategies,[170, 261, 435] particularly in primary care settings and in relation to policy change.[27, 436] Yet, only about 2% of CPGs tailor guidelines to local health system user preferences.[437] 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. [29, 171] For developers of health strategies, for example Models of Care[29] (currently absent in NZ[8]) or Models of Service Delivery,[171] 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 and provider experiences, and use of healthcare resources – the quadruple aim of value-based health care.[438]

Strengths of our study include the mixed-methods design[41] used to inform the criteria selection and the independent source of evidence[13] used to inform the performance ratings. A limitation of our study is that 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.

6.6 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.

6.7 ADDITIONAL ANALYSES

Two additional analyses were performed that were not reported in the published manuscript: cluster analysis (SUPPLEMENT 12; Cluster analysis, p.409) and a sensitivity analysis involving value-for-money charts (SUPPLEMENT 12; Sensitivity analysis, p.412). Because these analyses had no impact on the main findings of the study, they will not be discussed here any further.

SUPPLEMENT 6

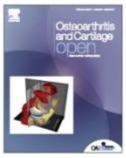
ARTICLE IN PRESS

Osteoarthritis and Cartilage Open xxx (xxxx) xxx

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131



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

Jason Chua [a], Paul Hansen [b], Andrew M. Briggs [c], Ross Wilson [a], David Gwynne-Jones [a], J. Haxby Abbott [a,*]

[a] *Centre for Musculoskeletal Outcomes Research, University of Otago, New Zealand*
[b] *Department of Economics, University of Otago, New Zealand*
[c] *School of Physiotherapy and Exercise Science, Curtin University, Australia*

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 ($\leq$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

---

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

https://doi.org/10.1016/j.ocarto.2020.100110
Received 1 August 2020; Accepted 5 October 2020
2665-9131/© 2020 Osteoarthritis Research Society International (OARSI). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: J. Chua et al., Stakeholders' preferences for osteoarthritis interventions in health services: A cross-sectional study using multi-criteria decision analysis, Osteoarthritis and Cartilage Open, https://doi.org/10.1016/j.ocarto.2020.100110

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

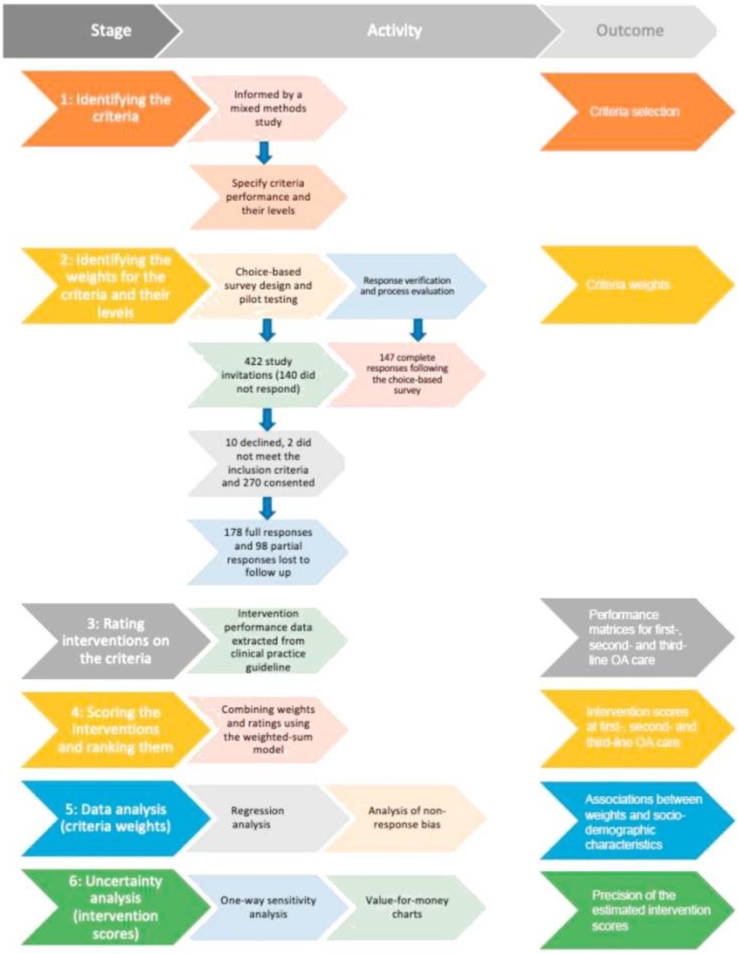


Fig. 1. Flow diagram of the study by stage, primary activity and outcomes for each stage.

(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), Māori health (with an active interest in advocacy or consumer representation for Māori health) and OA researchers (having published at least one academic article related to OA). A characterisation of these groups is reported elsewhere [29].

Māori, 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 Māori 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 MCDAs, 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 both 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

For OA ... Which of these 2 hypothetical treatments do you prefer?
(all else being equal)

Effectiveness - e.g. the improvement in pain or function
Low

Duration - how long the treatment effect lasts
Long (10 years or more)

Effectiveness - e.g. the improvement in pain or function
High

Duration - how long the treatment effect lasts
Short (up to 4-6 hours)

Fig. 2. Example of the 1000minds pairwise-ranking question.

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 hereinafter 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	70 (39)	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)
Maori 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 (13)	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.

completed the data verification and process evaluation. Their socio-demographic characteristics are summarised in Table 1.

3.1.2. *Choice-based survey*

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%), 'pycnogenol' (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 ($r_s = 1.00$, $n = 75$, $p < 0.01$). 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 (disaggregated 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 at their best performance level sum to 1 (or equivalently, 100%).

Criteria[b] (most to least important) Performance levels (worst to best)	Definition	Full sample mean weight[c] (n = 178)	Mean weight[c] by group			
			Consumers (n = 63)	Providers (n = 79)	Policy-informants (n = 24)	OA Researchers (n = 12)
Recommendation to use the intervention now	Recommendation for using the intervention at first-line OA care.					
Strong against		0.0	0.0	0.0	0.0	0.0
Conditional against		0.0647	0.0645	0.0618	0.0720	0.0697
Neutral[a]		0.1108	0.1073	0.1118	0.1116	0.1213
Conditional for		0.1529	0.1462	0.1581	0.1462	0.1678
Strong for		0.1904	0.1848	0.1947	0.1851	0.2038
Quality of the evidence about the intervention	The extent to which one can be confident that the effects of the reatment or service described are real.					
Very low		0.0	0.0	0.0	0.0	0.0
Low		0.0587	0.0526	0.0607	0.0704	0.0546
Moderate		0.1319	0.1145	0.1377	0.1549	0.1389
High		0.1765	0.1560	0.1835	0.2100	0.1709
Effectiveness of the intervention	The clinical effect of the intervention on pain.					
Low ($d < 0.2$)		0.0	0.0	0.0	0.0	0.0
Moderate ($d < 0.5$)		0.0983	0.0865	0.1021	0.1205	0.0911
High ($d > 0.5$)		0.1501	0.1376	0.1559	0.1720	0.1335
Duration of the intervention effect	The duration of follow up demonstrating a meaningful effect on pain.					
Short (up to 6hrs)		0.0	0.0	0.0	0.0	0.0
Short-medium[a] (<3 months)		0.0392	0.0421	0.0371	0.0300	0.0561
Medium (3 12 months)		0.0748	0.0817	0.0704	0.0592	0.0993
Long (>12 months)		0.1318	0.1506	0.1218	0.1145	0.1339
Risk of serious harm (**Risk-Serious**)	Treatment side-effects that have significant medical consequences, e.g. lead to death, permanent disability or prolonged hospitalisation.					
High (1 in 50 chance; >0.5%)		0.0	0.0	0.0	0.0	0.0
Medium (1 in 200 chance; 0.2% 0.5%)		0.0795	0.0864	0.0763	0.0701	0.0826
Low (1 in 500 chance; <0.2%)		0.1282	0.1325	0.1223	0.1179	0.1651
Risk of mild to moderate side effects (**Risk-Mild**)	Treatment side-effects that are not serious (see risk of serious harm).					
High (3 in 4 chance; >50%)		0.0	0.0	0.0	0.0	0.0
Medium (2 in 4 chance; 25 50%)		0.0527	0.0513	0.0610	0.0396	0.0319
Low (1 in 4 chance; <25%)		0.0941	0.0980	0.1016	0.0720	0.0686
Cost of the intervention	Total financial costs relevant to he use or provision of healthcare for OA.					
High (>$1000 per month or >$15,000 one-off)		0.0	0.0	0.0	0.0	0.0
Medium ($100-$1000 per month or $1500-$15,000 one-off)		0.0407	0.0450	0.0354	0.0513	0.0326
Low (<$100 per month or $1500 one-off)		0.0661	0.0735	0.0584	0.0760	0.0582
Accessibility to the intervention	The extent to which the intervention can be accessed by people with OA.					
Inconvenient travel, or wait time (>3 months)		0.0	0.0	0.0	0.0	0.0
Neither convenient or inconvenient travel, or wait time[a]		0.0313	0.0335	0.0309	0.0263	0.0331
Convenient travel, or wait time (<1 week)		0.0627	0.0670	0.0618	0.0526	0.0661

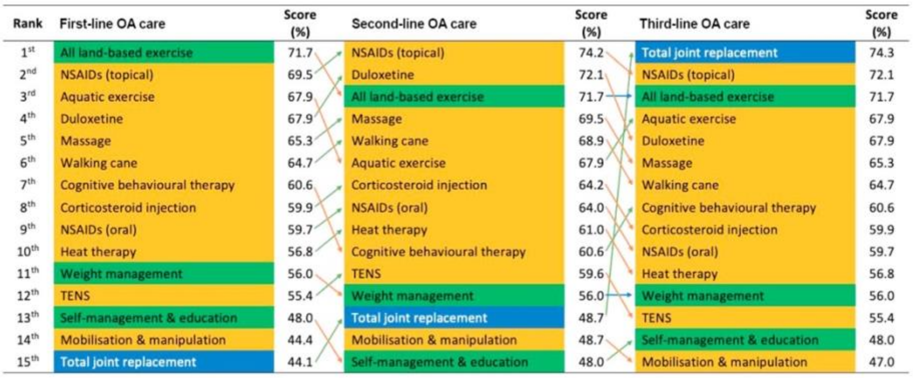
[a] Interpolated criterion level using a Bézier curve; *d* Cohen's *d* for effect size.
[b] Refer to Supplement 2 for a complete description of the criteria, including how interventions' were rated on the criteria.
[c] The weights, multiplied by 100, are equivalent to per cent points and at their best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

Recommendation, Quality, Effectiveness, Duration and Risk-Serious approximately 2.5–3 times more than the two least important criteria, Cost and Accessibility. These findings suggest that stakeholders' are willing to forego intervention Cost and Accessibility in favour of superior performance on the other criteria. Ultimately, the weights show that the choice of OA interventions is influenced by some criteria more than others, yet these differences may not accurately reflect the complexity of their real-world implementation.

Weights on the criteria did not differ meaningfully by stakeholder group (or by subgroup, Supplement 7). The regression analysis detected only small associations ($\leq$5.7%; Table 3) between weights and stakeholder group. Accounting for variance in group size made virtually no difference to the relative importance of the scored interventions ($r_s = 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



Rank	First-line OA care	Score (%)	Second-line OA care	Score (%)	Third-line OA care	Score (%)
1st	All land-based exercise	71.7	NSAIDs (topical)	74.2	Total joint replacement	74.3
2nd	NSAIDs (topical)	69.5	Duloxetine	72.1	NSAIDs (topical)	72.1
3rd	Aquatic exercise	67.9	All land-based exercise	71.7	All land-based exercise	71.7
4th	Duloxetine	67.9	Massage	69.5	Aquatic exercise	67.9
5th	Massage	65.3	Walking cane	68.9	Duloxetine	67.9
6th	Walking cane	64.7	Aquatic exercise	67.9	Massage	65.3
7th	Cognitive behavioural therapy	60.6	Corticosteroid injection	64.2	Walking cane	64.7
8th	Corticosteroid injection	59.9	NSAIDs (oral)	64.0	Cognitive behavioural therapy	60.6
9th	NSAIDs (oral)	59.7	Heat therapy	61.0	Corticosteroid injection	59.9
10th	Heat therapy	56.8	Cognitive behavioural therapy	60.6	NSAIDs (oral)	59.7
11th	Weight management	56.0	TENS	59.6	Heat therapy	56.8
12th	TENS	55.4	Weight management	56.0	Weight management	56.0
13th	Self-management & education	48.0	Total joint replacement	48.7	TENS	55.4
14th	Mobilisation & manipulation	44.4	Mobilisation & manipulation	48.7	Self-management & education	48.0
15th	Total joint replacement	44.1	Self-management & education	48.0	Mobilisation & manipulation	47.0

Key: Green = first-line (core) interventions; orange = second line (optional or advanced pharmacological) interventions and blue = third-line interventions (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	0.002	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	0.000	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)*	*(0.013)*	*(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.

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 and provider 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] used to inform the performance ratings. A limitation of our study is that

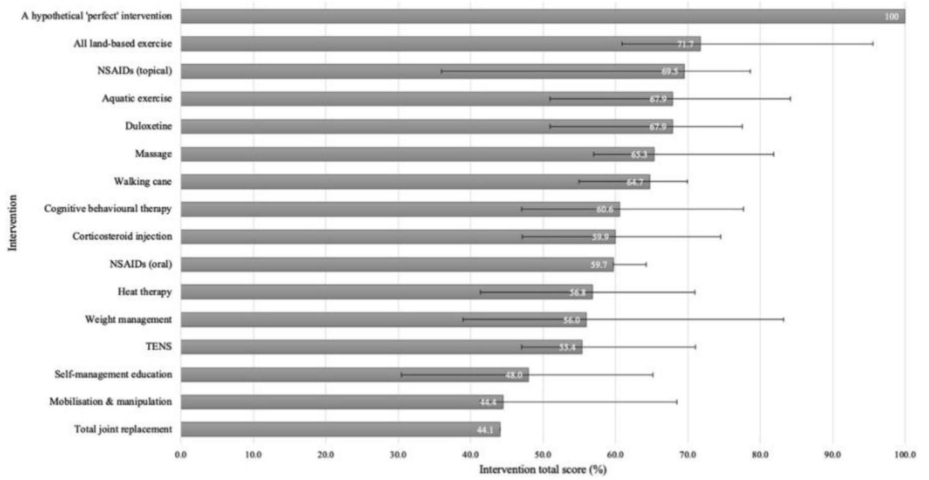


Fig. 4. Error bars representing the aggregate uncertainty in the 15 guideline-recommended interventions' total scores across all performance ratings for first-line OA care.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2020.100110.

References

[1] A.E. Nelson, K.D. Allen, Y.M. Golightly, A.P. Goode, J.M. Jordan, A systematic review of recommendations and guidelines for the management of osteoarthritis: the Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative, Semin. Arthritis Rheum. 43 (2014) 701 712.
[2] P.J. Larmer, N.D. Reay, E.R. Aubert, P. Kersten, Systematic review of guidelines for the physical management of osteoarthritis, Arch. Phys. Med. Rehabil. 95 (2014) 375 389.
[3] N. Østerås, K.P. Jordan, B. Clausen, C. Cordeiro, K. Dziedzic, J. Edwards, et al., Self-reported quality care for knee osteoarthritis: comparisons across Denmark, Norway, Portugal and the UK, RMD Open 1 (2015).
[4] L.C. Li, E.C. Sayre, J.A. Kopec, J.M. Esdaile, S. Bar, J. Cibere, Quality of nonpharmacological care in the community for people with knee and hip osteoarthritis, J. Rheumatol. 38 (2011) 2230 2237.
[5] A. Dhawan, R.C. Mather Iii, V. Karas, M.B. Ellman, B.B. Young, B.R. Bach Jr., et al., An epidemiologic analysis of clinical practice guidelines for non-arthroplasty treatment of osteoarthritis of the knee, Arthrosc. J. Arthrosc. Relat. Surg. 30 (2014) 65 71.
[6] T. Egerton, L.E. Diamond, R. Buchbinder, K.L. Bennell, S.C. Slade, A systematic review and evidence synthesis of qualitative studies to identify primary care clinicians' barriers and enablers to the management of osteoarthritis, Osteoarthritis Cartilage 25 (2017) 625 638.
[7] A.M. Kanavaki, A. Rushton, N. Efstathiou, A. Alrushud, R. Klocke, A. Abhishek, et al., Barriers and facilitators of physical activity in knee and hip osteoarthritis: a systematic review of qualitative evidence, BMJ Open 7 (2017), e017042.
[8] S. Brownlee, K. Chalkidou, J. Doust, A.G. Elshaug, P. Glasziou, I. Heath, et al., Evidence for overuse of medical services around the world, Lancet 390 (2017) 156 168.
[9] P. Glasziou, S. Straus, S. Brownlee, L. Trevena, L. Dans, G. Guyatt, et al., Evidence for underuse of effective medical services around the world, Lancet 390 (2017) 169 177.
[10] A.M. Briggs, R.S. Hinman, B. Darlow, K.L. Bennell, M. Leech, T. Pizzari, et al., Confidence and attitudes toward osteoarthritis care among the current and emerging health workforce: a multinational interprofessional study, ACR Open Rheumatol. 1 (2019) 219 235.
[11] R. Speerin, C. Needs, J. Chua, L.J. Woodhouse, M. Nordin, R. McGlasson, et al., Implementing Models of Care for musculoskeletal conditions in health systems to support value-based care, Best Pract. Res. Clin. Rheumatol. (2020), https://doi.org/10.1016/j.berh.2020.101548.
[12] S. Ferreira de Meneses, F. Rannou, D.J. Hunter, Osteoarthritis guidelines: barriers to implementation and solutions, Ann. Phys. Rehabil. Med. 59 (2016) 170 173.
[13] R. Lau, F. Stevenson, B.N. Ong, K. Dziedzic, S. Treweek, S. Eldridge, et al., Achieving change in primary care–causes of the evidence to practice gap: systematic reviews of reviews, Implement. Sci. 11 (2016) 40.
[14] A. Boivin, K. Currie, B. Fervers, J. Gracia, M. James, C. Marshall, et al., Patient and public involvement in clinical guidelines: international experiences and future perspectives, Qual. Saf. Health Care 19 (2010) 1 4.
[15] C.A. Chong, I.J. Chen, G. Naglie, M.D. Krahn, How well do guidelines incorporate evidence on patient preferences? J. Gen. Intern. Med. 24 (2009) 977 982.
[16] A.Y. Lim, M. Doherty, What of guidelines for osteoarthritis? Int. J. Rheum. Dis. 14 (2011) 136 144.
[17] C.E. Young, F.M. Boyle, K.S. Brooker, A.J. Mutch, Incorporating patient preferences in the management of multiple long-term conditions: is this a role for clinical practice guidelines? J. Comorb. 5 (2015) 122 131.
[18] S. Woolf, H.J. Schünemann, M.P. Eccles, J.M. Grimshaw, P. Shekelle, Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations, Implement. Sci. 7 (2012) 61.
[19] A.C. Tricco, W. Zarin, P. Rios, V. Nincic, P.A. Khan, M. Ghassemi, et al., Engaging policy-makers, heath system managers, and policy analysts in the knowledge synthesis process: a scoping review, Implement. Sci. 13 (2018) 31.
[20] R. Salloum, E. Shenkman, J. Louviere, D. Chambers, Application of discrete choice experiments to enhance stakeholder engagement as a strategy for advancing implementation: a systematic review, Implement. Sci. 12 (2017).
[21] E.W. de Bekker-Grob, M. Ryan, K. Gerard, Discrete choice experiments in health economics: a review of the literature, Health Econ. 21 (2012) 145 172.
[22] M.D. Clark, D. Determann, S. Petrou, D. Moro, E.W. de Bekker-Grob, Discrete choice experiments in health economics: a review of the literature, Pharmacoeconomics 32 (2014) 883 902.
[23] D. Pinto, U. Bockenholt, J. Lee, R.W. Chang, L. Sharma, D.J. Finn, et al., Preferences for physical activity: a conjoint analysis involving People with chronic knee pain, Osteoarthr. Cartil. 27 (2) (2019) 240 247.
[24] T.-L. Laba, J-a Brien, M. Fransen, S. Jan, Patient preferences for adherence to treatment for osteoarthritis: the MEdication Decisions in Osteoarthritis Study (MEDOS), BMC Muscoskel. Disord. 14 (2013) 160.
[25] N.K. Arden, A.B. Hauber, A.F. Mohamed, F.R. Johnson, P.M. Peloso, D.J. Watson, et al., How do physicians weigh benefits and risks associated with treatments in patients with osteoarthritis in the United Kingdom? J. Rheumatol. (2012) 111066, jrheum.
[26] K. Marsh, M. Ij, P. Thokala, R. Baltussen, M. Boysen, Z. Kalo, et al., Multiple criteria decision analysis for health care decision making-emerging good practices: report 2 of the ISPOR MCDA emerging good practices task force, Value Health 19 (2016) 125 137.
[27] E. von Elm, D.G. Altman, M. Egger, S.J. Pocock, P.C. Gotzsche, J.P. Vandenbroucke, The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies, Lancet 370 (2007) 1453 1457.
[28] National Clinical Guideline Centre, National institute for health and clinical excellence: guidance, in: Osteoarthritis: Care and Management in Adults, National Institute for Health and Care Excellence (UK), London, 2014.
[29] J. Chua, A.M. Briggs, P. Hansen, C. Chapple, H. Abbott, Choosing interventions for hip or knee osteoarthritis - what matters to stakeholders? A mixed-methods study, Osteoarthr. Cartil. Open 2 (3) (2020) 100062, accepted.
[30] T. Blakely, M. Tobias, B. Robson, S. Ajwani, M. Bonné, A. Woodward, Widening ethnic mortality disparities in New Zealand 1981 99, Soc. Sci. Med. 61 (2005) 2233 2251.
[31] New Zealand Government, Briefing to the incoming minister: Health, in: The Treasury, 2014. Wellington.
[32] J. Chua, A.M. Briggs, P. Hansen, C. Chapple, J.H. Abbott, Choosing interventions for hip or knee osteoarthritis - what matters to stakeholders? A mixed-methods study, Osteoarthr. Cartil. Open (2020) 100062.
[33] K. Marsh, T. Lanitis, D. Neasham, P. Orfanos, J. Caro, Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature, Pharmacoeconomics 32 (2014) 345 365.
[34] P. Hansen, F. Ombler, A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives, J. Multi-Criteria Decis. Anal. 15 (2008) 87 107.
[35] Royal Australian College of General Practitioners, Guideline for the Management of knee and hip Osteoarthritis, RACGP, East Melbourne, Vic, 2018.
[36] R. Welte, T. Feenstra, H. Jager, R. Leidl, A decision chart for assessing and improving the transferability of economic evaluation results between countries, Pharmacoeconomics 22 (2004) 857 876.
[37] E.M. Roos, C.B. Juhl, Osteoarthritis 2012 year in review: rehabilitation and outcomes, Osteoarthritis Cartilage 20 (2012) 1477 1483.
[38] P. Hansen, N. Devlin, Multi-criteria decision analysis (MCDA) in healthcare decision-making, in: Oxford Research Encyclopedia of Economics and Finance, Oxford University Press, 2019.
[39] FMLOGIT: Stata Module Fitting a Fractional Multinomial Logit Model by Quasi Maximum Likelihood [computer Program], Boston College Department of Economics, 2008. Version Revised 17 Feb 2017.
[40] D. McFadden, The Revealed Preferences of a Public Bureaucracy, Department of Economics, University of California, Berkeley, 1968.
[41] D. McFadden, The revealed preferences of a government bureaucracy: empirical evidence, Bell J. Econ. (1976) 55 72.
[42] G. Guyatt, A.D. Oxman, E.A. Akl, R. Kunz, G. Vist, J. Brozek, et al., GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables, J. Clin. Epidemiol. 64 (2011) 383 394.
[43] K.S. Dziedzic, K.D. Allen, Challenges and controversies of complex interventions in osteoarthritis management: recognizing inappropriate and discordant care, Rheumatology 57 (2018) iv88 iv98.
[44] C.A. Pellegrini, G. Ledford, R.W. Chang, K.A. Cameron, Understanding barriers and facilitators to healthy eating and physical activity from patients either before and after knee arthroplasty, Disabil. Rehabil. 40 (2018) 2004 2010.
[45] L. Walsh, S. Hill, A.E. Wluka, P. Brooks, R. Buchbinder, A. Cahill, et al., Harnessing and supporting consumer involvement in the development and implementation of Models of Care for musculoskeletal health, Best Pract. Res. Clin. Rheumatol. 30 (2016) 420 444.
[46] C. Berchi, P. Degieux, H. Halhol, B. Danel, M. Bennani, C. Philippe, Impact of falling reimbursement rates on physician preferences regarding drug therapy for osteoarthritis using a discrete choice experiment, Int. J. Pharm. Pract. 24 (2016) 114 122.
[47] J. Ratcliffe, M. Buxton, T. McGarry, R. Sheldon, J. Chancellor, Patients' preferences for characteristics associated with treatments for osteoarthritis, Rheumatology 43 (2004) 337 345.
[48] L. Fraenkel, L. Suter, C.E. Cunningham, G. Hawker, Understanding preferences for disease-modifying drugs in osteoarthritis, Arthritis Care Res. 66 (2014) 1186 1192.
[49] L.J. Damschroder, D.C. Aron, R.E. Keith, S.R. Kirsh, J.A. Alexander, J.C. Lowery, Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science, Implement. Sci. 4 (2009).
[50] A. Fretheim, H.J. Schünemann, A.D. Oxman, Improving the use of research evidence in guideline development: 3. Group composition and consultation process, Health Res. Pol. Syst. 4 (2006) 15.
[51] J. Shiffman, S. Smith, Generation of political priority for global health initiatives: a framework and case study of maternal mortality, Lancet 370 (2007) 1370 1379.
[52] A.R. Gagliardi, M.C. Brouwers, V.A. Palda, L. Lemieux-Charles, J.M. Grimshaw, How can we improve guideline use? A conceptual framework of implementability, Implement. Sci. 6 (2011) 26.
[53] A.M. Briggs, M. Chan, H. Slater, Models of Care for musculoskeletal health: moving towards meaningful implementation and evaluation across conditions and care settings, Best Pract. Res. Clin. Rheumatol. 30 (2016) 359 374.
[54] J. Baldwin, A.M. Briggs, W. Bagg, P. Larmer, An osteoarthritis model of care should be a national priority for New Zealand, N. Z. Med. J. 30 (2017) 78 86.
[55] R. Sikka, J.M. Morath, L. Leape, The Quadruple Aim: care, health, cost and meaning in work, BMJ Qual. Saf. 24 (2015) 608 610.

SUPPLEMENT 7

STROBE Statement

Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	**Recommendation**	**Page**	**Relevant text**
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	260	Title
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	261	Abstract
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	262	Introduction
Objectives	3	State specific objectives, including any prespecified hypotheses	263	Introduction, paragraph 4.
Methods				
Study design	4	Present key elements of study design early in the paper	263	Methods>Design
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	266	Methods>sampling and recruitment > recruitment sources
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	266	Methods>sampling and recruitment > recruitment sources
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	267 268 271	Stage 1 Stage 2 Stage 4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	270	Stage 3 Chapter 5
Bias	9	Describe any efforts to address potential sources of bias	271	Stage 5
Study size	10	Explain how the study size was arrived at	-	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	268	Stage 2

Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	271	Stages 5, selection bias
		(*b*) Describe any methods used to examine subgroups and interactions	271	Stage 5, response consistency & association with sociodemographic characteristics
		(*c*) Explain how missing data were addressed	-	-
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	-	-
		(*e*) Describe any sensitivity analyses	272	Stage 6, uncertainty analysis
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	265	Figure 15
		(b) Give reasons for non-participation at each stage	265	Figure 15
		(c) Consider use of a flow diagram	265	Figure 15
Descriptive data	14*	(a) Give characteristics of study participants (egg demographic, clinical, social) and information on exposures and potential confounders	274	Table 19
		(b) Indicate number of participants with missing data for each variable of interest	-	-
Outcome data	15*	Report numbers of outcome events or summary measures	-	-
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (egg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	275 279	Table 20 Figure 17
		(*b*) Report category boundaries when continuous variables were categorized	-	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-	-
Other analyses	17	Report other analyses done—egg analyses of subgroups and interactions, and sensitivity analyses		Supplements 8 & 12

Discussion				
Key results	18	Summarise key results with reference to study objectives	285	Discussion, paragraph 1
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	287	Discussion, second-last paragraph
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	288	Conclusion
Generalisability	21	Discuss the generalisability (external validity) of the study results	287	Discussion, last paragraph
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based		Declared with the journal submission requirements

SUPPLEMENT 8

Survey instructions

The purpose of this survey is to reveal your preferences for the attributes (or characteristics) of hip or knee osteoarthritis treatments - e.g., accessibility, cost, effectiveness, duration of the treatment effect, and so on.

Please download and review the attribute definitions sheet (PDF) now. A quick review will make it easier for you to complete the survey.

INSTRUCTIONS

Consider any stage of OA, for example, the early/ mild or advanced/ severe stage of OA in your decisions.

Review the two boxes below and then click on your preferred treatment option, based on **your primary area** of experience or expertise - i.e., consumer, health care provider, health policy-maker, Māori advocate or content area expert.

You'll choose from two treatment options at-a-time. As they are described using treatment attributes and attribute levels, they are hypothetical and do not have names.

Each time you choose an option, the attributes and/or attribute levels will change, but not the question.

Watch this short 1.5-minute instructional video (YouTube) which demonstrates how to complete the survey.

If you consider other attributes in your decision-making, assume they are at an average level AND equal for both treatments - e.g., if you wanted to consider effectiveness in addition to cost and duration, the average level of effectiveness between low and high is 'medium.'

Attributes	**Levels (BEST to WORST)**
Accessibility	Convenient, Inconvenient
Cost	Low, Medium, High
Duration	Long, Medium, Short
Effectiveness	High, Medium, Low
Quality of the evidence	High, Moderate, Low, Very Low
Risk of mild or moderate side-effects	Low, Moderate, High
Risk of serious harm	Low, Moderate, High
Recommendation	Very Good, Good, Bad, Very Bad

Definition sheet

Table of attribute and attribute-level definitions	
Abbreviated attribute name and levels (best to worst)	Definition (from unpublished data)
Accessibility: **Convenient travel or wait time (<1 week):** the treatment can be accessed by the person living with OA in a week or so, regardless of their travel needs. **Inconvenient travel or wait time (>3 months):** There may be a waiting time of a month or more to receive the treatment; the provider may be inconvenient to reach; or, the treatment may not be accessible at all because of health-system related factors.	The extent to which the treatment or service can be accessed by people with OA. For example, the distance to nearest provider, wait time and, the ability for culturally and linguistically groups or people from diverse sociodemographic background to equally access health care for OA (fairness).
Cost per month OR one-off total: **Low:** \$0-\$100 per month OR less than \$1500 total. **Medium:** \$100-\$1000 per month OR \$1500-\$15,000 total. **High:** \$1000 or more per month OR \$15,000+ total.	Total financial costs relevant to the use or provision of health care for OA - e.g., costs to the health system, out-of-pocket costs to the consumer and, the societal costs of providing health care for OA. Societal costs include tax revenue and lost wages due to time away from work, reduced employment or early retirement.

Duration – how long the treatment effect lasts: **Long (10 years or longer):** The effects of the treatment are experienced for 10 years or longer - e.g. joint replacement. **Medium (3-12 months or more):** The effects of the treatment are experienced for several months to a year or more - e.g. physical exercise. **Short (up to 4-6 hours):** The effects of the treatment are experienced for up to 4-6 hours - e.g. drug therapy.	The length of time the benefits of the treatment last. E.g., the beneficial effects of surgery, if appropriate, may last for 10-15 years after initial healing has occurred, with little ongoing care until 10-15 years have elapsed. In contrast, drug therapy may require frequent dosing every 4 hours to maintain its effect on pain.
Effectiveness – e.g. the improvement in pain or the ability to function. The estimated magnitude of change caused by the intervention: High Medium Low	The ability for the treatment or service to achieve the desired result - e.g. the change in pain and function, caused by the intervention. Effectiveness is different to quality of the evidence because it describes the impact, or how 'big' the change caused by the treatment is, not how likely it is to happen, or how confident you are that it'll happen – this is the 'quality of the evidence'. For example, a highly effective treatment with a very low quality of evidence means that the likelihood, or chance of it actually working is very small and, if it did work, it would have a high/large impact on pain and/or function.
Quality of the evidence – how confident you are that the treatment works: **High:** Further research is very unlikely to change our confidence in the likely effect of the intervention. **Moderate:** Further research is likely to have an important impact on our confidence in the likelihood of effect of the intervention and may change the estimate. **Low:** Further research is very likely to have an important impact on our confidence in the likelihood of effect of the intervention and is likely to change the estimate. **Very low:** Any estimate of the treatment effect is very uncertain.	The extent to which one can be confident that the effects of the treatment or service described are real. "Evidence" can mean different things to different people, however, there is an accepted hierarchy of how valid each source is. For example, anecdotal claims about the effectiveness of treatment in advertisements, from peers or family members, or from individual treatment providers may not be as convincing as independent health professional advice, rigorous research, peer-reviewed Systematic Reviews, or authoritative Clinical Practice Guidelines. See 'effectiveness' to see how it contrasts from quality of the evidence.

Risk of mild or moderate side-effects - e.g., *temporary* pain, discomfort, nausea, heartburn or stomach pain: Low: 1 in 4 chance = 25% Moderate: 2 in 4 chance = 50% High: 3 in 4 chance = 75%	Treatment side-effects associated with comfort or safety.
Risk of serious harm - e.g., implant failure, drug toxicity, stomach bleeding or ulcer: Low: 1 in 500 chance = 0.2% Moderate: 1 in 200 chance = 0.5% High: 1 in 50 chance = 2%	Treatment side-effects associated with comfort or safety.
Recommendation to use the treatment now: **Very good:** all or almost all informed people would use the treatment now. **Good:** most informed people would use the treatment now, *but not all.* **Bad**: most informed people would try another treatment first, *but not all.* **Very bad:** all or almost all informed people would try another treatment first.	Providing or using the right treatments or services for early/mild OA. For example, it would not be recommended or appropriate to use powerful drug treatments such as opioids before, say, self-management and education, physical exercise or, less-powerful drug therapies such as paracetamol. '…*but not all*' means that a substantial number of informed people would still choose (or not choose) the treatment.

SUPPLEMENT 9

Pilot testing protocol

We aimed to clarify if the participants understood the following survey elements:

a. Understand the survey instructions
b. Understand the survey question
c. Conceptually interpret the treatment attributes and attribute levels the same as our intention
d. Assess overall perception of the questionnaire understandability, ease of use and length.

Sample size: Falkner (2003) reported that a sample size of 10 can detect between 82-94% of usability problems compared with 15 detecting 90-97% of possible issues. With a sample size of 20, 95-98% of issues may be detected. Purposeful sampling is more desirable because respondents may be able to detect context-specific issues that may go otherwise unnoticed. We sampled following stakeholders: consumers, health care providers, health care policy-makers, Māori advocacy and content area experts.

Method: The think-aloud method with concurrent probing was used to learn about the participants perceptions and reactions to the survey questionnaire instructions, the question layout and the response categories; that are, the treatment attributes and attribute levels descriptions. In the later stage of the survey evaluation, retrospective probing was used to simulate the field administration of the survey where a respondent completes the instrument unaided.

Interview script

Thanks for coming here today to help us test out our survey questions. At this point, we are not collecting information about you. Instead, we're testing our questions on a few people such as yourself, so that we can improve them. I will read you the questions, and I'd like you to answer them. However, I'd also like to hear about what you're thinking. Please try to think aloud—just tell me everything that comes to mind, whether it seems important or not. I'll also be asking you about how you come up with your answers and how you're interpreting the questions, and I'll take lots of notes. If any question seems unclear, is hard to answer, or doesn't make sense, please tell me that—don't be shy. We'll just take our time and get as far as we can in an hour. Do you have any questions before we start?

Warm up exercise 1: Picture in your mind where you live. How many windows are there? As you count the windows, tell me what you are thinking about and seeing.

Warm up exercise 2: Picture in your mind having to choose between two different types of treatments for hip or knee osteoarthritis. What are the two treatment options you are thinking about? As you imagine the two treatment options, tell me about what you are thinking or imagining to help you choose between the two.

Table S8. Interview checklist

Question/ section	Probe	Comments	Processing areas assessed by probe			
INTERVIEWER TO READ OUT	“Tell me what you are thinking.”		Comprehension	Retrieval	Decision	Response processes
PART A						
Intro	Can you repeat what the instructions are telling you in your own words?		X			
	What do the words ‘treatment attribute’ mean to you?		X			
	What do the words, ‘attribute level’ mean to you?		X			
	Can you tell me in your own words what the question was asking?		X			
	What were you thinking about when you answered the question or how did you arrive at that answer?				X	
Steps	Can you repeat what the instructions are telling you in your own words?		X			
	Was it easy or hard to interpret the Steps? Any jargon?				X	X
Definitions sheet	READ OUT EACH ABBREVIATED ATTRIBUTE AND ITS LEVELS					
	Can you tell me in your own words what [attribute] and [the attribute’s levels] mean to you?		X			
	Are the attribute levels relevant to you?			X		
	Accessibility					
	Cost					
	Duration					
	Effectiveness					
	Quality of the evidence					

Question/ section	Probe	Comments	Processing areas assessed by probe			
INTERVIEWER TO READ OUT	“Tell me what you are thinking.”		Comprehension	Retrieval	Decision	Response processes
	Risk of mild/moderate harm					
	Risk of serious harm					
	Timing					
Actual survey question	ASK THE RESPONDENT TO REVIEW THE QUESTION AND THINK ALOUD					
	Can you repeat the question in your own words?		X			
	How did you arrive at your answer?				X	
	Was it easy or hard for you to answer the question?				X	
	What attribute levels were you thinking of when you arrived at that?				X	
	Was that easy or hard to answer? I noticed that you hesitated. Tell me what you were thinking.				X	
	Are the options – this one, they are equal, skip question – useful to you?					X
Questions about the overall perception of the survey	Do you think it would be hard for other people to answer the questions?					
	What do you think about the survey length?					

Table S9. Summary of feedback after each round of interviewing and survey revisions (5 rounds total, 4 revisions; N=17 participants)

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
INTRODUCTION - PART A	In this survey, we present you with hypothetical treatment options for hip or knee osteoarthritis. We want you to choose which of these treatment options is the best, to you. We won't show you the treatment names because we want you to choose between the attributes of the treatments (or the treatment characteristics) rather than the treatments themselves. (They're hypothetical	Need to revise the instructions. Rephrase as the participant is confused between interventions, intervention properties and attributes.	-	-	-	-	-	-	-	-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	treatments, remember!)									
INTRODUCTION - PART B	Can you repeat what the instructions are telling you in your own words?	The text is too complicated to understand.		Video potentially too long and difficult to understand. Revise the video down to 1-minute max. Revise and simplify the video.		Video is too long. Video is fine x2. Video is fine.		-		Introduction and video all OK.

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	What do the words 'treatment attribute' mean to you?	Properties or characteristics of treatment.		Remove the reference to numbers of levels in the table – confusing.		-		-		-
	What do the words, 'attribute level' mean to you?	The degree of something…the relative degree, big to small, long to short.		-		-		-		-
	Can you tell me in your own words what the question was asking?	Any stage of OA seems vague.		-		-		-		-
	What were you thinking about when you answered the question or how did you arrive at that answer?	-		-		-		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
SURVEY INSTRUCTIONS	Can you repeat what the instructions are telling you in your own words?	If there is no clear preference between the two then consider them equal. Clarify language - Consider choices…IN RELATION to your area of expertise. not AND YOUR AREA…professional capacity…e.g.. in regard to your professional capacity…		-		-		Instructions made sense		Steps OK
	Was it easy or hard to interpret the Steps? Any jargon?	- The tip doesn't work, because it creates confusion. Consider all other things in your calculus equal at their medium or average range/level. - Tip needs rewording, it doesn't make sense. - Tips section doesn't make sense; too wordy overall and complicated.		Tip needs rewording Reword the tips section Reorder the instructions and steps Simplify the instructions		Suggests replacing harms with side-effects because harm has negative connotations		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
ATTRIBUTE DEFINITION SHEET	Can you tell me in your own words what [attribute] and [the attribute's levels] mean to you?	-		-		-		Remove advanced/late treatment references in the performance matrix.		-
	Are the attribute levels relevant to you?	-		-		-		-		-
	Accessibility	Inconvenient travels means?		-		-		-		-
	Cost	What the financial implications of the treatment will be for me. Direct costs vs NHS costs were considered by the interviewee. What's the intention – personal, national or institutional cost. This q was answered by the definitions sheet. Cost depends on perspective - personal or system		-		Difficultly contemplating cost and duration of Tx effect with the cost-levels. Remove cost data		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	Duration	How long it'll continue to be effective. What about treatment effectiveness that decreases over time? E.g. in years 7-10 it could be less effective in years 1-6. How long will it stay at its current level and how long will it last in total. Continued QULAITY over time and overall duration or effectiveness. How long a treatment goes on for until it ceases to be a treatment. Add effect to description.- Big jump between levels. Add effect to description.- Add effect to duration		Considered duration with effectiveness. Update duration to include 'of the treatment effect'		-		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	Effectiveness	The potency of the treatment. Its restorative potential. The extent to which it can mitigate or negate the OA. Effectiveness is very much based on the personal choice; what the person wants to achieve after the operation or management… - Remove reference to Cohen's d		Difficult to discern between effectiveness and Qual of the evidencex4 Fix up example so it's not ambiguous. Add example into definitions sheet so it's easier to grasp.		No issue discerning between quality and effectiveness once the definitions sheet was reviewed		-		-
	Quality of the evidence	-How confident you are in that treatment will do what it says it will do before you engage in the treatment-How established the evidence is.-likelihood of the impact'.. Difficult to comprehend high effectiveness but low quality.		Difficult to discern between effectiveness and Qual of the evidencex4.		Tricky to contemplate quality of evidencex1		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	Risk of mild/moderate harm	Undesirable physical physiological consequences stemming directly or directly from the treatment.; Perhaps capitalise mild/moderate		-		Suggested replacing mild/moderate to mild to moderate side-effects		-		-
	Risk of serious harm	-There's an issue with leaving out long-term vs serious harms, which could be immediate as well. -Severity of the risk and also a question of whether or not its short term or long-term risk. -Asking about long-term risk -Did not consider these examples serious - e.g. death -Perhaps capitalise serious		-		Clarity needed re: joint failure vs implant failure, which are both different things		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	Appropriateness	-Uncertain is not well described…Perhaps change uncertain to ambiguous. Uncertain doesn't seem to fit between appropriate and not appropriate…unknown, unsure, ambiguous, unclear?-What are the consequences of inappropriate timing? Perhaps need to include this in the description. E.g., if delivered late, what risks are incurred? What's the impact of timing?-Perhaps timing: what are the implications of action at appropriate and inappropriate vs uncertain? What are the consequences of not appropriate timing – e.g. cost utilisation, worsening health, -Difficult to conceptualise		Difficult to understandx3 Problematic, description is too long.		Wording is better now; considered timing in terms of appropriate timing and future risk. Misinterpreted. Timing still problematic.		All other attributes interpreted OK, except for timing. Timing definition also problematic.		New definition seems to work OK.

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
		implications of inappropriate timing. Appropriateness in terms of unknown benefits or harms caused by intervention-Need to include late to the not appropriate level. Possibly dominates other criteria.-A 'no brainer'. Possibly dominates other criteria.-Add late to not appropriate								
SURVEY QUESTION										
	Can you repeat the question in your own words?	-		-		-		-		-
	How did you arrive at your answer?	-		-		-		-		-
	Was it easy or hard for you to answer the question?	-		Found it difficult to do the exercisex2.		-		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
	What attribute levels were you thinking of when you arrived at that?	It's difficult to consider duration, cost and risk without thinking of frequency. E.g., low cost, short duration but high frequency might be the same as a high cost, long duration treatment … cumulative cost over time.		-		-		-		-
	Was that easy or hard to answer? I noticed that you hesitated. Tell me what you were thinking.	-		-		-		-		-
	Are the options – this one, they are equal, skip question – useful to you?	-		-		-		-		-
QUESTIONS ABOUT THE OVERALL PERCEPTION OF THE SURVEY	Do you think it would be hard for other people to answer the questions?	-		How does the person know what is mild/ early OA? This may need		-		-		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
				more description.						
	What do you think about the survey length?	Too long. Although it did get quicker as the participant came familiar with it.		OK, but too long.		-		-		-
OTHER COMMENTS		Perhaps remove this combination?è High effectiveness and low evidence seem as a very unlikely combination.è Change timing definitions to include early AND late timing. At the moment it only includes EARLY but not late.è Does medium have to medium for the levels…e.g. With duration, 10yeras + and 4-6 hours, medium is weeks to months. Should it be more like 5 years???		Need to tidy up some of the attribute descriptions so they're unambiguous. E.g. effectiveness and duration. Timing is still problematic.		- Results aligned with participant's preferences for the attributes - remove the cost level data so it's less confusing and there's less text on the screen.		Most comments were too specific for the general nature of our survey. Length and clarity OK, the definitions sheet is vital		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
INTERVIEWER NOTES		Survey instructions and process was OK. There may be an issue with conceptualising the consequences of inappropriate timing. Participant 1 felt that long-term risk was important to consider, but wasn't in the list.Timing (not appropriate) def needs to include 'late', too.Participant 1 noted that his preferences would change with desperation - i.e. early vs advanced OA.Big jump between levels for cost and duration.Risk of Harm vs duration: per duration or the course of treatment? Where does or should frequency come into the equation?Add all else being equal to the		Participant 5 found it difficult to navigate the survey itself and missed the most important part - making trade-offs between pairs of attributes at a time.		Considered cost, duration of treatment and an additional attribute - frequency of Tx. Survey took a while to complete. The timing attribute needs further clarity. Survey takes a long time to complete.		Possibly language barrier with interpretation; things were OK apart from timing. Timing seems ambiguous.		-

Survey Section	Probe	Summary of results following interviews 1 - 4	Revisions	Summary of interviews 5 - 9	Revisions	Summary of interviews 10 - 13	Revisions	Summary if interviews 14 - 16	Revisions	Summary of final interview 17
		instructions. Need to describe quality better so that it distinguishes between effectiveness.Instructions too long and complicated.DNC survey; Responses would change depending on disease severity.completed survey in <60'. Responses would change depending on disease severity.Responses would change depending on disease severity; considered frequency with duration & costResponses would change depending on early/mild or advanced/severe OA								

Preferences for early vs. advanced OA

During pilot testing we identified that preferences might be conditional to OA severity. To test this conditionality we conducted a pilot study whereby participants completed the same preference task twice to examine if preference weights differed between early and advanced OA. Specifically, participants completed the preference task under these two conditions sequentially and within two weeks of one another:

1. In the **early/mild** stage of OA...Which of these two hypothetical treatments do you prefer?
2. In the **advanced/severe** stage of OA...Which of these two hypothetical treatments do you prefer?

We compared the mean weight differences of the eight criteria between these two conditions and found that there were no meaningful differences (Table S10), leading to our conclusion that OA disease severity did not change underlying weights for the intervention criteria, and would likely have minimal overall effect on the relative importance of the interventions, the main objective of our study. Therefore we modified our overall survey question to the generalisable format that is, "For OA…Which of these two hypothetical interventions do you prefer?"

Table S10. Mean difference between criteria weights under the two pilot study survey conditions (early vs. advanced OA). N=29 participants

	Test value = 0		95% Confidence Interval of the Difference	
Criterion	Sig. (2-tailed)	Mean Difference	Lower	Upper
Accessibility	0.025	0.02686	0.0036	0.0501
Cost	0.026	0.02392	0.0031	0.0448
Duration	0.923	0.00155	-0.0311	0.0342
Effectiveness	0.195	-0.01727	-0.0439	0.0093
Quality of Evidence	0.164	-0.02154	-0.0524	0.0093
Recommendation	0.301	-0.01823	-0.0536	0.0172
Risk-Mild	0.744	-0.00356	-0.0257	0.0186
Risk-Serious	0.505	0.00826	-0.0168	0.0333

Bonferroni adjusted alpha for 0.05 = 0.006; alpha for 0.10 = 0.004

SUPPLEMENT 10

Table S11. Extent of alignment between the revealed criteria weights and overall expectations about their importance by group (N=147; median=Agree)

	Level of agreement, N (%)							
Stakeholder group	Strongly agree	Agree	Somewhat agree	Neither agree or disagree	Somewhat disagree	Disagree	Strongly disagree	Median Response
Consumers	7 (14.9)	20 (42.6)	9 (19.1)	8 (17)	2 (4.3)	1 (2.1)	0 (0)	Agree
Māori advocate	2 (40)	2 (40)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)	Agree
Providers	4 (6.3)	27 (42.2)	25 (39.1)	3 (4.7)	4 (6.3)	0 (0)	1 (1.6)	Somewhat agree
Policy-makers	3 (15)	10 (50)	4 (20)	1 (5)	1 (5)	0 (0)	1 (5)	Agree
OA researchers	2 (18.2)	3 (27.3)	4 (36.4)	1 (9.1)	1 (9.1)	0 (0)	0 (0)	Somewhat agree
Total	18 (12.2)	62 (54.5)	42 (83)	14 (9.5)	8 (5.4)	1 (0.7)	2 (1.4)	Agree

Table S12. Results of the consistency check. The weight difference between the overall sample mean weights and the mean weights of the respondents answering all or none of the consistency checks correctly was, at most 0.024 points (or equivalently 2.4%)

Criteria	Number of correct consistency questions (N=178)			
	3 (n=67)	2 (n=78)	1 (n=27)	0 (n=6)
	Weight difference between the overall sample mean preference weights minus group consistency check mean preference weight			
Recommendation for using the treatment now	0.006	-0.005	0.003	-0.024
Quality of the evidence	0.00	0.003	-0.005	0.014
Effectiveness of the intervention	-0.003	-0.003	0.012	0.016
Duration of the intervention effect	-0.003	0.001	0.004	-0.005
Risk of serious harm	0.008	-0.002	-0.011	0
Risk of mild to moderate side-effects	-0.004	0.001	0.007	-0.001
Cost of the intervention	-0.001	0.005	-0.007	-0.005

Accessibility to the intervention	0.001	0	-0.004	0.005

Table S13. Survey user-friendliness by group (N=147; median=Neutral)

	Level of difficulty, N (%)					
Stakeholder group	Very easy	Easy	Neutral	Difficult	Very difficult	Median Response
Consumers	7 (14.9)	13 (27.7)	15 (31.9)	11 (23.4)	1 (2.1)	Neutral
Māori advocate	0 (0)	1 (20)	2 (40)	2 (40)	0 (0)	Neutral
Providers	7 (10.9)	22 (34.4)	28 (43.8)	7 (10.9)	0 (0)	Neutral
Policy-makers	5 (25)	9 (45)	4 (20)	2 (10)	0 (0)	Easy
OA researchers	2 (18.2)	7 (63.6)	1 (9.1)	1 (9.1)	0 (0)	Easy
Total	21 (14.3)	52 (35.4)	50 (34.0)	23 (15.6)	1 (0.7)	Neutral

Table S14. Difficulty completing the survey task by group (N=147; median=Neutral)

	Level of difficulty, N (%)					
Stakeholder group	Very easy	Easy	Neutral	Difficult	Very difficult	Median Response
Consumers	1 (2.1)	13 (27.7)	11 (23.4)	18 (38.3)	4 (8.5)	Neutral
Māori advocate	0 (0)	3 (60)	0 (0)	2 (40)	0 (0)	Easy
Providers	1 (1.6)	10 (15.6)	20 (31.3)	30 (46.9)	3 (4.7)	Difficult
Policy-makers	1 (5)	5 (25)	4 (20)	9 (45)	1 (5)	Neutral-Difficult
OA researchers	0 (0)	4 (36.4)	2 (18.2)	5 (45.5)	0 (0)	Neutral
Total	3 (2.0)	35 (23.8)	37 (25.2)	64 (43.5)	8 (5.4)	Neutral

SUPPLEMENT 11

Uncertainty analysis: rules and decisions

Intervention name (RACGP CPG technical document page number)

Criteria

- Ratings (worst to best)

Rule

Accessibility (current rating)

- **inconvenient travel, or wait time**
- **neither convenient or inconvenient travel, or wait time**
- **convenient travel, or wait time**

We defined consensus of >70% as absence of uncertainty. If this criterion was rated as ‘neither convenient or inconvenient travel, or wait time’ by the Delphi panellists (i.e. they could not decide) then this criterion was modified ±1 level.

Duration (current rating; mean follow-up time in months ±SD [range]) [number of studies]

- **Short: up to 4-6hrs**
- **Short-medium <3 months**
- **Medium: 3-12 months**
- **Long: >12 months**

If the range of the follow-up duration stated in the RACGP CPG technical document evidence tables exceeded the defined interval for the criterion rating, then the rating was modified to match the duration rating(s) associated with the range of follow-up duration. Only studies demonstrating a meaningful effect on pain were considered.

Cost (current rating; mean cost per month in 2017NZD ±SD [range])

- **High $1k/mo or >$15k one-off**
- **Medium $100-$1k/mo or $1500-$15k one-off**

- **Low <$100/mo or <$1500 one-off**

If the range of the total cost of delivering the studies considered in the RACGP CPG evidence tables exceeded the defined interval for the intervention rating, then the criterion was modified to match the cost rating(s) associated with the range of the intervention cost.

Effectiveness (current rating; standardised mean difference (SMD) [Lower CI – Upper CI])

- **Low: <0.2**
- **Medium: 0.2 - 0.5**
- **High: >0.5**

If the confidence interval of the estimate of intervention effect exceeded the defined interval for the intervention rating, then the rating was modified to match the effectiveness rating(s) associated with the SMD confidence intervals.

Quality of the evidence (QoE; current level, QoE lowest rating – QoE highest rating)

- **Very low**
- **Low**
- **Moderate**
- **High**

If the quality of evidence for any of the outcomes considered in the RACGP CPG technical document GRADE evidence tables differed from the quality of evidence rating for the intervention, then the criterion level was modified to match the QoE rating(s) associated with the range of outcome QoE ratings assigned by the RACGP guideline developers.

Risk of mild to moderate harm (current rating; absolute risk [CI])

- **High: 3 in 4: >50%**
- **Moderate: 2 in 4: 25% ≤50%**
- **Low: 1 in 4: <25%**

If the confidence interval of the absolute risk for mild or moderate harms stated in the RACGP CPG technical document GRADE evidence tables exceeded the defined interval for the intervention rating, then the criterion rating was modified to match the risk of mild to moderate harm rating(s) associated with the SMD confidence intervals. If no explicit mild to moderate harms were mentioned in the guideline or technical document, then the level of risk of mild to moderate harm was defaulted to 'low'.

Risk of serious harm (current rating; absolute risk [CI])

- **High: >0.5%**
- **Moderate: 0.2% - 0.5%**
- **Low: <0.2%**

If the confidence interval of the absolute risk for serious harms stated in the RACGP CPG technical document GRADE evidence tables exceeded the defined interval for the intervention rating, then the criterion rating was modified to match the risk of mild to moderate harm rating(s) associated with the SMD confidence intervals. If no explicit serious harms were mentioned in the guideline or technical document, then the level of risk of serious harm was defaulted to ‘low’.

Recommendation (recommendation for first-, second- and third-line OA care; recommendation for first-line OA care)

- strong against
- conditional against
- conditional (neutral)
- conditional for
- strong for

If the intervention received a ‘conditional (neutral)’ recommendation for first-line OA care in the RACGP CPG (i.e. denoting uncertainty over the balance of benefits) then the criterion was modified ±1 level.

Aquatic exercise (p.36)	**Notes**
Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level.	
Duration (medium 3-12 months; 7 ±4.4 (3-15) [5])	Change +1 level.
The range exceeds the defined interval for a medium duration rating.	
Cost (medium; $367.43 ±321.56 [$160-$1152] [12])	Change -1 level (high).
The cost of interventions included in the RACGP CPG guideline extends into the defined range for a high cost rating.	
Effectiveness (medium; -0.31 [-0.47, -0.15])	Change -1 level.
The lower confidence interval of the SMD extends into the interval defined for a rating of low effectiveness.	
Quality of the evidence (low; low–moderate)	Change +1 level.
The pain and function outcomes were graded moderate QoE.	

Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (for recommendation)	No change.
-	
Walking cane (p.61)	
Accessibility (inconvenient travel, or wait time)	No change.
-	
Duration (short-medium; 2 ±0 (2-2) [1])	No change.
-	
Cost (Low; RRP $200 one-off cost)	No change.

-	
Effectiveness (medium; -0.38 [-0.69, -0.07])	Change ±1 level.
The confidence interval extends into the defined intervals for low and high effectiveness ratings.	
Quality of the evidence (low; low-low)	No change.
-	
Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (neutral; conditional for)	No change.
-	

Cognitive behavioural therapy (p.25)	
Accessibility (inconvenient travel, or wait time)	No change.
-	
Duration (medium; 6.8 ±3.9 (2.5-12) [3])	Change -1 level.
The range extends into defined interval for a short-medium duration rating.	
Cost (medium; $801 ±360 [$289-$1157])	Change +1 level.
The range extends into the defined interval for a high cost rating.	
Effectiveness (low; -0.21 [-0.42, -0.01]; [4])	Change +1 level.
The CI extends into the defined interval for a medium effectiveness rating.	
Quality of the evidence (low; very low - moderate)	Change ±1 level.
Pain outcome rated moderate QoE. Others, very low.	

Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (neutral; conditional for)	No change.
-	
Corticosteroid injection (p.141)	
Accessibility (Convenient travel, or wait time)	No change.
-	
Duration (short-medium duration; 2.8 ±1.9 (0.2-6) [11])	Change +1 level.
The range extends into the defined interval for a medium duration rating.	
Cost (Low; $79 ±3 [$75-$83])	No change.

Cost is based on 4 drugs recommended in the New Zealand Formulary for OA use: 1) Betamethasone sodium phosphate with betamethasone acetate 2) Methylprednisolone acetate 3) Dexamethasone phosphate 4) Triamcinolone acetonide	
Effectiveness (medium; -0.4 [-0.58, -0.22])	Change +1 level.
The confidence interval extends into the defined interval for a high effectiveness rating.	
Quality of the evidence (very low; very low-low)	Change +1 level.
several outcomes were graded low QoE.	
Risk of mild to moderate harm (low; N/A)	
None reported.	No change.
Risk of serious harm (low; 6 fewer per 1000 (from 23 fewer to 58 more))	
The confidence interval for extends into the defined interval for high risk of serious harm rating.	Change +2 levels.

Recommendation (neutral; conditional for)	No change.
-	
Duloxetine (p.119)	
Accessibility (inconvenient travel, or wait time)	No change.
-	
Duration (medium duration, 3.3 ±0.5 (3-4) [3])	No change.
-	
Cost (Medium; RRP $166 RRP)	No change.
Effectiveness (medium; -0.43 [-0.58, -0.29])	Change +1 level.
The confidence interval extends into the defined interval for a high effectiveness rating.	

Quality of the evidence (moderate; moderate - high)	Change +1 level.
Several outcomes were rated as high quality of evidence in the RACGP CPG evidence table for this intervention.	
Risk of mild to moderate harm (low; 180 more per 1000 (from 49 more to 358 more))	
“Among the participants in the three included RCTs, treatment with duloxetine was well tolerated, with the majority of adverse events being of mild or moderate intensity (e.g. constipation, nausea, hyperhidrosis, cough, myalgia, arthralgia, palpitations).”	Change –1 level.
Risk of serious harm (low; 0 fewer per 1000 (from 8 fewer to 34 more))	
The confidence interval extends into the defined interval for a high risk of serious harm rating.	Change -2 levels.
Recommendation (neutral recommendation; conditional for)	No change.
-	
Heat therapy (p.42)	
Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.

Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level..	
Duration (short-medium duration (<3 months); 0.6 ±0.4 (0.2-1) [2])	No change.
-	
Cost (low; RRP $22 one-off cost)	Change -1 level (medium; using one-off cost)
Effectiveness (medium; -0.38 [-0.69, -0.07])	Change ±1 level.
The confidence interval extends into the defined interval for a large or small effect size.	
Quality of the evidence (very low; very low - low):	Change +1 level.
Outcomes for pain and total adverse events were graded low quality of evidence by the RACGP CPG developers.	
Risk of mild to moderate harm (low; N/A)	No change.
None reported.	

Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (neutral; conditional for)	No change.
-	
All land-based exercise (p.27)	
Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level.	
Duration (medium; 4.5 ±5.5 (2-24) [18])	Change ±1 level.
The follow up range extends into the defined intervals for short-medium and long duration of effect ratings.	
Cost (medium; \$581 ±630 [\$56-\$1728]; [54 studies])	Change ±1 level

The range of the cost of the programs considered in the RACGP CPG papers considered exceed the defined level for a medium cost rating.	
Effectiveness (medium; -0.49 [-0.59, -0.39])	Change +1 level.
The confidence interval extends into the defined range for a high effectiveness rating.	
Quality of the evidence (low; low-moderate)	Change +1 level.
The pain and quality of life outcomes were graded as moderate quality of evidence.	
Risk of mild to moderate harm (low)	No change.
The absolute risk of study withdrawals is 17 fewer per 1000 (from 5 more to 34 fewer).	
Risk of serious harm (low)	No change.
None reported explicitly for 'all land-based exercise'.	
Recommendation (strong for recommendation; strong for)	No change.

-	
Massage (p.37)	
Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of massage could not be reached such that the rating could be modified ±1 level.	
Duration (short-medium; 3 ±1.9 (1-6) [4])	Change +1 level.
The range extends into the defined interval for a medium duration rating.	
Cost (medium; $489 ±307 [$74-$970])	Change +1 level (low).
The range of cost extends into defined interval for a low cost rating.	
Effectiveness (high; -0.70 [-0.97, -0.43])	Change -1 level.
The confidence interval extends into the defined range for a medium effectiveness rating.	

Quality of the evidence (low; low-moderate)	Change +1 level.
The function outcome was graded as moderate quality of evidence by the RACGP CPG developers.	
Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (neutral; conditional for)	No change.
-	
Mobilisation & manipulation (p.39)	
Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level.	

Duration (short-medium; 0.5 ±0 (0.5) [1])	No change.
-	
Cost (medium; $145 ±0 [$145-$145])	No change.
-	
Effectiveness (low; -0.16 [-0.52, +0.21])	Change +2 levels.
The confidence interval extends into the defined interval for a high effectiveness rating.	
Quality of the evidence (very low; very low-low)	Change +1 level.
The outcomes, withdrawals due to adverse events and total adverse events, were rated as low quality of evidence by the RACGP CPG developers.	
Risk of mild to moderate harm (low; N/A)	No change.
None reported.	

Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (neutral; conditional for)	No change.
-	
NSAIDS (oral, p.84)	
Accessibility (convenient travel, or wait time)	No change.
-	
Duration (medium duration, 3.4 ±0.9 (3-6) [11])	No change.
-	
Cost ($4; low)	No change.
Based on 30d cost and $5 dispensing fee for 3 months and GP visit cost.	

Effectiveness (medium; -0.26 [-0.31, -0.22])	No change.
-	
Quality of the evidence (moderate; moderate-high)	Change +1 level.
QoE for the outcomes, pain and function, total adverse events and gastrointestinal effects, was high. The overall QoE was rated as moderate due to QoE for withdrawals due to adverse effects and serious adverse effects.	
Risk of mild to moderate harm (medium; see comment)	No change.
The RACGP refer to adverse events but do not specify the rates. We referred to Sostres, C., Gargallo, C. J., & Lanas, A. (2013) for rates of mild to moderate harm.	
Risk of serious harm (high; 36 more per 1000 (from 18 more to 56 more))	No change.
-	
Recommendation (neutral; conditional for):	No change.
-	

NSAIDS (topical, p.89)	
Accessibility (convenient travel, or wait time)	No change.
-	
Duration (medium duration, 3 ±0 (3-3) [6])	No change.
-	
Cost (medium; RRP $134)	No change.
-	
Effectiveness (medium; -0.2 [-0.29, -0.11])	Change -1 level.
The confidence interval extends into the defined interval for a low effectiveness rating.	
Quality of the evidence (moderate; moderate-high)	Change +1 level.

The QoE for the outcomes pain, function, withdrawals due to adverse events and local reactions were all rated as high by the RACGP CPG developers.	
Risk of mild to moderate harm (low; 39 more per 1000 (from 5 more to 82 more)	No change.
Local reactions (dermatitis, skin dryness, itching, dermatosis, allergic reaction). "Usually adverse events from topical NSAIDs agents are minimal, but there is mild toxicity because of local skin reactions." (p.43, RACGP CPG)	
Risk of serious harm (low; 1 more per 1000 (from 6 fewer to 38 more))	Change +2 levels (high).
The confidence interval extends into the defined interval for a high risk of serious harm rating.	
Recommendation (against recommendation for first-line OA care; neutral at first-line OA care)	Change ±1 level.
A neutral recommendation from the RACGP CPG developers indicates uncertainty in the recommendation for this intervention.	
Self-management and education programmes (p.23)	

Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level.	
Duration (medium 8.1 ±4.1 (2.5-12) [4])	Change -1 level.
The range extends into the defined interval for a short-medium duration rating.	
Cost (medium; $236 Lord (1999) in Pinto (2012)) **RACGP considered papers: $649.48 ±320 [$205-$1044]), 5 studies**	Change -1 level.
The range of studies considered in the RACGP CPG exceeds the defined interval for a medium cost rating.	
Effectiveness (low; -0.16 [-0.39, +0.06])	Change +1 level.
The confidence interval extends into the defined interval for a medium effectiveness rating.	
Quality of the evidence (very low; very low-very low)	No change.
-	

Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (neutral; neutral)	Change ±1 level.
A neutral recommendation from the RACGP CPG developers indicates uncertainty in the recommendation for this intervention.	
Transcutaneous electrical nerve stimulation (p.65)	
Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level.	
Duration (short-medium; 0.5 ±0 (0.5-1) [3])	No change.
-	

Cost (high; $1344 ±136 [$1152-$1440])	If personal portable TENS are considered, the cost rating could be low. Change -2 levels (low).
“Portable TENS units are now widely available for people to use at home as a self-management strategy. Unlike other electrotherapy devices, portable TENS may be used as a continuous therapy by individuals to modulate pain, allowing them to engage in other activities while the unit is active.” (P.38 RACGP CPG).	
Effectiveness (high; -0.76 [-1.13, -0.39])	Change -1 level.
The confidence interval extends into the defined interval for a high effectiveness rating.	
Quality of the evidence (very low; very low-low)	Change +1 level.
The pain outcome was rated low quality of evidence.	
Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.

None reported.	
Recommendation (neutral; conditional for)	No change.
-	
Total joint replacement (-)	
Accessibility (inconvenient travel, or wait time)	No change.
-	
Duration (17-year survival of 92.9% (New Zealand Joint Registry 19 Year Report))	No change.
-	
Cost (high; $24,050 [$17,500-$30,600] one-off cost)	No change.
-	
Effectiveness (high; -1.72 [-0.97, -2.46])	No change.
-	

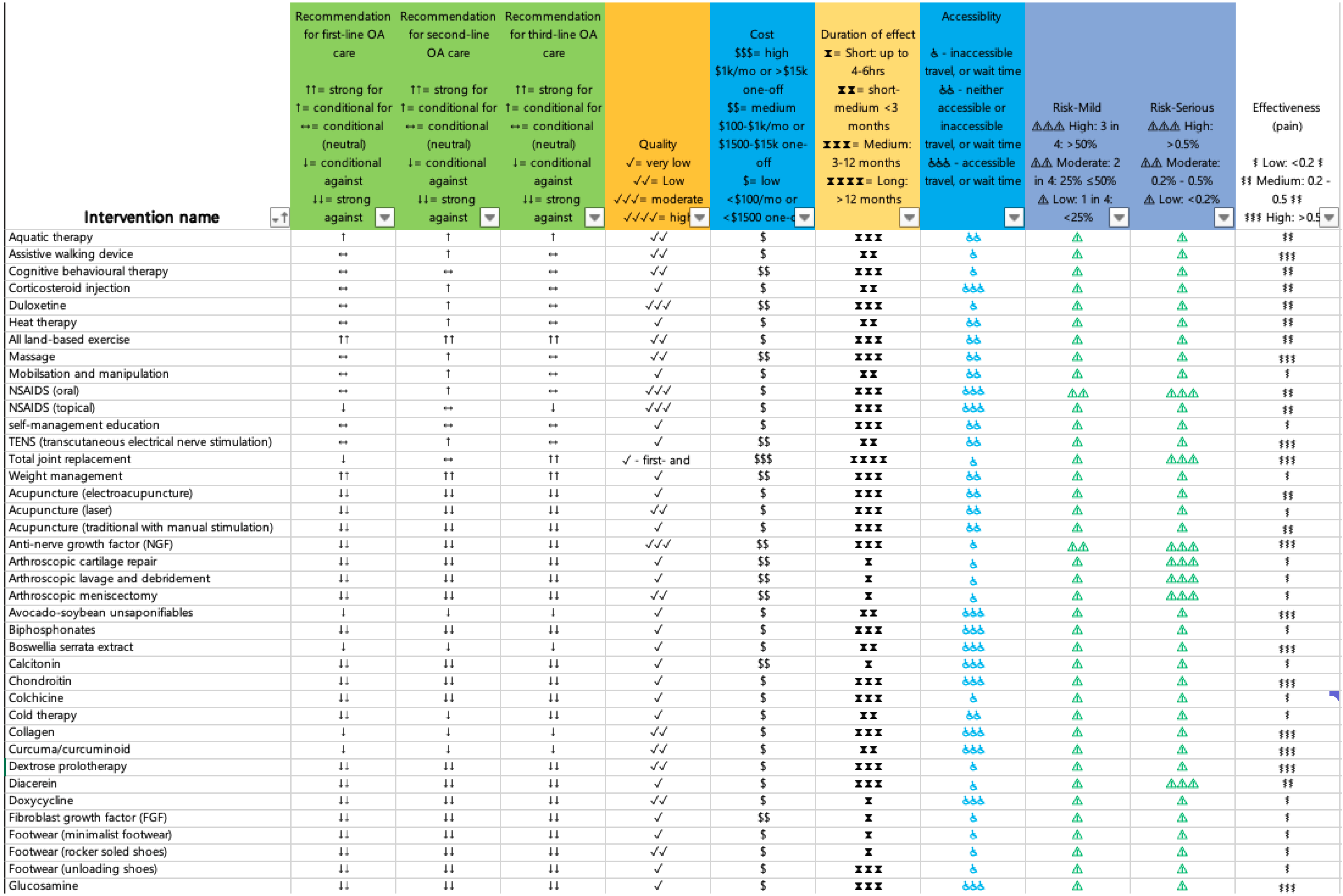
Quality of the evidence (very low for first- & second-line OA care; high for third-line OA care)	No change.
-	
Risk of mild to moderate harm (low; N/A):	No change.
-	
Risk of serious harm (high; based on confidential NZ public hospital rates of serious harm):	No change.
-	
Recommendation (first-line care – against; second-line care – neutral'; third-line care – strong for recommendation):	Change ±1 level for second-line OA care only.
For second-line care, the intervention received a neutral recommendation, denoting uncertainty.	
Weight management (p.40)	

Accessibility (neither convenient or inconvenient travel, or wait time)	Change ±1 level.
Consensus about the accessibility of this intervention could not be reached such that the rating could be modified ±1 level.	
Duration (medium duration (3-12 months); 12 ±4.9 (6-18) [3])	Change +1 level.
The range extends into the defined interval for a long duration rating.	
Cost (medium; $427.59 Sevick (2009) in Pinto (2012)) **RACGP CPG considered papers: $805 ±521 [$204-$1476]), 3 studies**	Change –1 level.
The range of the cost of studies considered in the RACGP CPG extends into the defined interval for a high cost rating.	
Effectiveness (medium; 0.38 [-0.88, +0.11])	Change ±1 level.
The range extends into the defined interval for a high effectiveness rating.	
Quality of the evidence (very low; very low-moderate):	Change +2 (moderate quality of evidence).
Most outcomes were rated as low quality of evidence by the RACGP CPG developers, however, percentage weight loss was rated moderate quality of evidence.	

Risk of mild to moderate harm (low; N/A)	No change.
None reported.	
Risk of serious harm (low; N/A)	No change.
None reported.	
Recommendation (strong for recommendation; conditional for)	No change.
-	

SUPPLEMENT 12

Table S15. Intervention performance matrix



Intervention name	Recommendation for first-line OA care ↑↑= strong for ↑= conditional for ↔= conditional (neutral) ↓= conditional against ↓↓= strong against	Recommendation for second-line OA care ↑↑= strong for ↑= conditional for ↔= conditional (neutral) ↓= conditional against ↓↓= strong against	Recommendation for third-line OA care ↑↑= strong for ↑= conditional for ↔= conditional (neutral) ↓= conditional against ↓↓= strong against	Quality ✓= very low ✓✓= Low ✓✓✓= moderate ✓✓✓✓= hig	Cost $$$= high $1k/mo or >$15k one-off $$= medium $100-$1k/mo or $1500-$15k one-off $= low <$100/mo or <$1500 one-	Duration of effect ⧗= Short: up to 4-6hrs ⧗⧗= short-medium <3 months ⧗⧗⧗= Medium: 3-12 months ⧗⧗⧗⧗= Long: >12 months	Accessiblity ♿ - inaccessible travel, or wait time ♿♿ - neither accessible or inaccessible travel, or wait time ♿♿♿ - accessible travel, or wait time	Risk-Mild ⚠⚠⚠ High: 3 in 4: >50% ⚠⚠ Moderate: 2 in 4: 25% ≤50% ⚠ Low: 1 in 4: <25%	Risk-Serious ⚠⚠⚠ High: >0.5% ⚠⚠ Moderate: 0.2% - 0.5% ⚠ Low: <0.2%	Effectiveness (pain) ⚕ Low: <0.2 ⚕ ⚕⚕ Medium: 0.2 - 0.5 ⚕⚕ ⚕⚕⚕ High: >0.5
Aquatic therapy	↑	↑	↑	✓✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕
Assistive walking device	↔	↑	↔	✓✓	$	⧗⧗	♿	⚠	⚠	⚕⚕⚕
Cognitive behavioural therapy	↔	↔	↔	✓✓	$$	⧗⧗⧗	♿	⚠	⚠	⚕⚕
Corticosteroid injection	↔	↑	↔	✓	$	⧗⧗	♿♿♿	⚠	⚠	⚕⚕
Duloxetine	↔	↑	↔	✓✓✓	$$	⧗⧗⧗	♿	⚠	⚠	⚕⚕
Heat therapy	↔	↑	↔	✓	$	⧗⧗	♿♿	⚠	⚠	⚕⚕
All land-based exercise	↑↑	↑↑	↑↑	✓✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕
Massage	↔	↑	↔	✓✓	$$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Mobilsation and manipulation	↔	↑	↔	✓	$	⧗⧗	♿♿	⚠	⚠	⚕
NSAIDS (oral)	↔	↑	↔	✓✓✓	$	⧗⧗⧗	♿♿♿	⚠⚠	⚠⚠⚠	⚕⚕
NSAIDS (topical)	↓	↔	↓	✓✓✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕
self-management education	↔	↔	↔	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕
TENS (transcutaneous electrical nerve stimulation)	↔	↑	↔	✓	$$	⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Total joint replacement	↓	↔	↑↑	✓ - first- and	$$$	⧗⧗⧗⧗	♿	⚠	⚠⚠⚠	⚕⚕⚕
Weight management	↑↑	↑↑	↑↑	✓	$$	⧗⧗⧗	♿♿	⚠	⚠	⚕
Acupuncture (electroacupuncture)	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕
Acupuncture (laser)	↓↓	↓↓	↓↓	✓✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕
Acupuncture (traditional with manual stimulation)	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕
Anti-nerve growth factor (NGF)	↓↓	↓↓	↓↓	✓✓✓	$$	⧗⧗⧗	♿	⚠⚠	⚠⚠⚠	⚕⚕⚕
Arthroscopic cartilage repair	↓↓	↓↓	↓↓	✓	$$	⧗	♿	⚠	⚠⚠⚠	⚕
Arthroscopic lavage and debridement	↓↓	↓↓	↓↓	✓	$$	⧗	♿	⚠	⚠⚠⚠	⚕
Arthroscopic meniscectomy	↓↓	↓↓	↓↓	✓✓	$$	⧗	♿	⚠	⚠⚠⚠	⚕
Avocado-soybean unsaponifiables	↓	↓	↓	✓	$	⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕
Biphosphonates	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕
Boswellia serrata extract	↓	↓	↓	✓	$	⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕
Calcitonin	↓↓	↓↓	↓↓	✓	$$	⧗	♿♿♿	⚠	⚠	⚕
Chondroitin	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕
Colchicine	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿	⚠	⚠	⚕
Cold therapy	↓↓	↓	↓↓	✓	$	⧗⧗	♿♿	⚠	⚠	⚕
Collagen	↓	↓	↓	✓✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕
Curcuma/curcuminoid	↓	↓	↓	✓✓	$	⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕
Dextrose prolotherapy	↓↓	↓↓	↓↓	✓✓	$	⧗⧗⧗	♿	⚠	⚠	⚕⚕⚕
Diacerein	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿	⚠	⚠⚠⚠	⚕⚕
Doxycycline	↓↓	↓↓	↓↓	✓✓	$	⧗	♿♿♿	⚠	⚠	⚕
Fibroblast growth factor (FGF)	↓↓	↓↓	↓↓	✓	$$	⧗	♿	⚠	⚠	⚕
Footwear (minimalist footwear)	↓↓	↓↓	↓↓	✓	$	⧗	♿	⚠	⚠	⚕
Footwear (rocker soled shoes)	↓↓	↓↓	↓↓	✓✓	$	⧗	♿	⚠	⚠	⚕
Footwear (unloading shoes)	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿	⚠	⚠	⚕
Glucosamine	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕

Glucosamine and chondroitin in compound form	↓↓	↓↓	↓↓	✓	$	⧗	♿♿♿	⚠	⚠	⚕
Interleukin-1 (IL-1) inhibitors	↓↓	↓↓	↓↓	✓✓	$$$	⧗	♿	⚠	⚠	⚕
Knee braces (re-aligning patellofemoral braces)	↓↓	↓↓	↓↓	✓	$	⧗⧗	♿	⚠	⚠	⚕⚕
Knee braces (valgus unloading/re-alignment braces)	↓↓	↓↓	↓↓	✓✓	$	⧗	♿	⚠	⚠	⚕⚕
Knee braces (varus unloading/re-alignment braces)	↓	↓	↓	✓	$	⧗	♿	⚠	⚠	⚕
Knee exercise: Land-based exercise (stationary cyclin	↑	↑	↑	✓	$	⧗⧗	♿♿	⚠	⚠	⚕⚕
Knee exercise: MUSCLE STRENGTHENING ONLY for k	↑↑	↑↑	↑↑	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Knee exercise: MUSCLE STRENGTHENING ONLY for c	↑↑	↑↑	↑↑	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Knee exercise: Stationary cycling only	↑	↑	↑	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕
Knee exercise: Tai Chi only	↑↑	↑↑	↑↑	✓✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Knee exercise: Walking only	↑↑	↑↑	↑↑	✓	$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕
Knee exercise: Yoga only	↑	↑	↑	✓	$	⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Methotrexate	↓↓	↓↓	↓↓	✓✓	$	⧗	♿♿♿	⚠⚠	⚠⚠⚠	⚕
Methylsulfonylmethane	↓	↓	↓	✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕
Omega-3 fatty acids	↓↓	↓↓	↓↓	✓	$	⧗	♿♿♿	⚠	⚠	⚕
Oral opioids	↓↓	↓↓	↓↓	✓✓	$	⧗⧗⧗	♿♿♿	⚠⚠	⚠⚠⚠	⚕⚕
Other electrotherapy (interferential)	↓↓	↓↓	↓↓	✓	$$	⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Other electrotherapy (laser)	↓↓	↓↓	↓↓	✓✓	$$	⧗⧗	♿♿	⚠	⚠	⚕⚕
Other electrotherapy (shockwave)	↓↓	↓↓	↓↓	✓	$$	⧗⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Paracetamol	↓	↔	↓	✓	$	⧗⧗	♿♿♿	⚠	⚠	⚕⚕
Platelet-rich plasma (PRP) injection	↓	↓	↓	✓	$$	⧗⧗⧗	♿	⚠⚠⚠	⚠⚠⚠	⚕⚕⚕
Pulsed electromagnetic/ shortwave therapy	↓	↓	↓	✓✓	$$	⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Pycnogenol	↓	↓	↓	✓✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕⚕
Shoe orthotics (lateral wedge insoles for medial tibiof	↓↓	↓↓	↓↓	✓	$	⧗⧗⧗	♿	⚠	⚠	⚕
Shoe orthotics (medial wedged insoles for lateral tibic	↓	↓	↓	✓	$	⧗⧗	♿	⚠	⚠	⚕⚕⚕
Shoe orthotics (shock absorbing insoles or arch supp	↓	↓	↓	✓	$	⧗	♿	⚠	⚠	⚕
Stem cell therapy	↓↓	↓↓	↓↓	✓	$$	⧗⧗⧗	♿	⚠⚠⚠	⚠	⚕⚕⚕
Strontium ranelate	↓↓	↓↓	↓↓	✓✓✓	$$	⧗	♿	⚠	⚠⚠	⚕
Taping (kinesio taping)	↓↓	↓↓	↓↓	✓	$	⧗	♿	⚠	⚠	⚕⚕
Taping (patellar taping)	↓	↓	↓	✓	$	⧗⧗	♿	⚠	⚠	⚕⚕
Therapeutic ultrasound	↓↓	↓↓	↓↓	✓✓✓	$$	⧗⧗	♿♿	⚠	⚠	⚕⚕⚕
Topical capsaicin	↓↓	↓↓	↓↓	✓✓	$	⧗⧗	♿♿♿	⚠⚠⚠	⚠	⚕
Transdermal buprenorphine	↓↓	↓↓	↓↓	✓✓	$	⧗⧗	♿♿♿	⚠⚠	⚠⚠⚠	⚕
Transdermal Fentanyl	↓↓	↓↓	↓↓	✓✓	$	⧗⧗	♿♿♿	⚠⚠	⚠⚠⚠	⚕⚕
Viscosupplementation injection	↓↓	↓↓	↓↓	✓✓	$$	⧗⧗⧗	♿	⚠⚠	⚠⚠⚠	⚕⚕
Vitamin D	↓↓	↓↓	↓↓	✓✓	$	⧗⧗⧗	♿♿♿	⚠	⚠	⚕⚕

Table S16. Knee OA interventions (N=75) ranked by the full sample mean criteria weights for first-, second- and third-line OA care. Highlighted interventions are RACGP OA knee guideline recommended algorithm interventions: green= first-line (core) interventions; orange=second-line (optional or advanced pharmacological) interventions and blue=third-line interventions (total joint replacement). The "*" denotes one-off payment interventions and "="denotes interventions which have equal total scores, and therefore equal ranks

Intervention	**Recommended interventions Rank, total score (%) (N=15)**			**All interventions Rank, total score (%) (N=75)**		
	First-line OA care	Second-line OA care	Third-line OA care	First-line OA care	Second-line OA care	Third-line OA care
Tai Chi only				1st (76.9)	1st (76.9)	1st (76.9)
All land-based exercise	1st (71.7)	3rd (71.7)	3rd (71.7)	2nd (71.7)	4th (71.7)	4th (71.7)
Muscle strengthening only for lower limb strengthening				3rd= (71.0)	5th= (71.0)	5th= (71.0)
Muscle strengthening only for quadriceps strengthening				3rd= (71.0)	5th= (71.0)	5th= (71.0)
Collagen				5th= (69.9)	7th= (69.9)	7th= (69.9)
Pycnogenol				5th= (69.9)	7th= (69.9)	7th= (69.9)
NSAIDs (topical)	2nd (69.5)	1st (74.2)	2nd (72.1)	7th (69.5)	2nd (74.2)	3rd (72.1)
Aquatic exercise	3rd (67.9)	6th (67.9)	4th (67.9)	8th (67.9)	11th (67.9)	9th (67.9)
Duloxetine	4th (67.9)	2nd (72.1)	5th (67.9)	9th (67.9)	3rd (72.1)	10th (67.9)
Curcuma/curcuminoid				10th (66.4)	12th (66.4)	11th (66.4)
Walking only				11th (65.8)	13th (65.8)	12th (65.8)
Massage therapy	5th (65.3)	4th (69.5)	6th (65.3)	12th (65.3)	9th (69.5)	13th (65.3)
Walking cane*	6th (64.7)	5th (68.9)	7th (64.7)	13th (64.7)	10th (68.9)	14th (64.7)
Yoga only				14th (63.7)	16th (63.7)	15th (63.7)
Stationary cycling only				15th (62.0)	17th (62.0)	16th (62.0)
Therapeutic ultrasound				16th (61.6)	18th (61.6)	17th (61.6)

Intervention	**Recommended interventions Rank, total score (%) (N=15)**			**All interventions Rank, total score (%) (N=75)**		
	First-line OA care	Second-line OA care	Third-line OA care	First-line OA care	Second-line OA care	Third-line OA care
Pulsed electromagnetic/shortwave therapy				17th (60.7)	20th (60.7)	18th (60.7)
Cognitive behavioural therapy	7th (60.6)	10th (60.6)	8th (60.6)	18th (60.6)	21st (60.6)	19th (60.6)
Avocado-soybean unsaponifiables				19th= (60.5)	22nd= (60.5)	20th= (60.5)
Boswellia serrata extract				19th= (60.5)	22nd= (60.5)	20th= (60.5)
Corticosteroid injection	8th (59.9)	7th (64.2)	9th (59.9)	21st (59.9)	14th (64.2)	22nd (59.9)
NSAIDs (oral)	9th (59.7)	8th (64)	10th (59.7)	22nd (59.7)	15th (64)	23rd (59.7)
Methylsulfonylmethane				23rd (58.9)	26th (58.9)	24th (58.9)
Stationary cycling & hatha yoga				24th (58.5)	27th (58.5)	25th (58.5)
Vitamin D				25th (58.3)	28th (58.3)	26th (58.3)
Chondroitin				26th= (57.6)	29th= (57.6)	27th= (57.6)
Glucosamine				26th= (57.6)	29th= (57.6)	27th= (57.6)
Dextrose prolotherapy				28th (57.2)	31st (57.2)	29th (57.2)
Heat therapy*	10th (56.8)	9th (61)	11th (56.8)	29th (56.8)	19th (61.0)	30th (56.8)
Weight management	11th (56.0)	11th (56.0)	12th (56.0)	30th (56.0)	32nd (56.0)	31st (56.0)
TENS	12th (55.4)	12th (59.6)	13th (55.4)	31st (55.4)	25th (59.6)	32nd (55.4)
Paracetamol				32nd (55.3)	24th (59.9)	33rd (55.3)
Medial wedged insoles*				33rd (54.2)	33rd (54.2)	34th (54.2)
Shockwave electrotherapy				34th (51.9)	34th (51.9)	35th (51.9)
Electroacupuncture				35th= (49.3)	35th= (49.3)	36th= (49.3)
Traditional acupuncture with manual stimulation				35th= (49.3)	35th= (49.3)	36th= (49.3)

Intervention	**Recommended interventions Rank, total score (%) (N=15)**			**All interventions Rank, total score (%) (N=75)**		
	First-line OA care	Second-line OA care	Third-line OA care	First-line OA care	Second-line OA care	Third-line OA care
Laser electrotherapy				37th (49.1)	37th (49.1)	38th (49.1)
Patellar taping				38th (49.1)	38th (49.1)	39th (49.1)
Interferential electrotherapy				39th (48.4)	41st (48.4)	40th (48.4)
Self-management education	13th (48.0)	15th (48.0)	14th (48.0)	40th (48.0)	42nd (48.0)	41st (48.0)
Laser acupuncture				41st (45.3)	43rd (45.3)	43rd (45.3)
Anti-nerve growth factor (NGF)				42nd (45.0)	44th (45.0)	44th (45.0)
Valgus unloading/re-alignment braces[*]				43rd (44.5)	45th (44.5)	45th (44.5)
Mobilisation & manipulation	14th (44.4)	14th (48.7)	15th (47.0)	44th (44.4)	40th (48.7)	42nd (47)
Total joint replacement[*]	15th (44.1)	13th (48.7)	1st (74.3)	45th (44.1)	39th (48.7)	2nd (74.3)
Re-aligning patellofemoral braces[*]				46th (42.6)	46th (42.6)	46th (42.6)
Biphosphonates				47th (42.6)	47th (42.6)	47th (42.6)
Oral opioids				48th (41.3)	49th (41.3)	48th (41.3)
Doxycycline				49th (41.0)	50th (41.0)	49th (41.0)
Stem cell therapy				50th (39.4)	51st (39.4)	50th (39.4)
Kinesio taping				51st (38.7)	52nd (38.7)	51st (38.7)
Transdermal Fentanyl				52nd (37.8)	53rd (37.8)	52nd (37.8)
Colchicine				53rd= (36.3)	54th= (36.3)	53rd= (36.3)
Unloading shoes[*]				53rd= (36.3)	54th= (36.3)	53rd= (36.3)
Lateral wedge insoles[*,†]				53rd= (36.3)	54th= (36.3)	53rd= (36.3)
Cold therapy				56th (35.9)	48th (42.4)	56th (35.9)
Topical capsaicin				57th (35.5)	57th (35.5)	57th (35.5)

Intervention	**Recommended interventions Rank, total score (%) (N=15)**			**All interventions Rank, total score (%) (N=75)**		
	First-line OA care	Second-line OA care	Third-line OA care	First-line OA care	Second-line OA care	Third-line OA care
Knee braces varus unloading/re-alignment braces*				58th= (35.3)	58th= (35.3)	58th= (35.3)
Shock absorbing insoles or arch supports*				58th= (35.3)	58th= (35.3)	58th= (35.3)
Glucosamine and chondroitin in compound form				60th= (35.1)	60th= (35.1)	60th= (35.1)
Omega-3 fatty acids				60th= (35.1)	60th= (35.1)	60th= (35.1)
Rocker-soled shoes*				62nd (34.7)	62nd (34.7)	62nd (34.7)
Strontium ranelate				63rd (34.6)	63rd (34.6)	63rd (34.6)
Diacerein				64th (33.3)	64th (33.3)	64th (33.3)
Platelet-rich plasma injection				65th (33.0)	65th (33.0)	65th (33.0)
Calcitonin				66th (32.6)	66th (32.6)	66th (32.6)
Viscosupplementation injection				67th (32.5)	67th (32.5)	67th (32.5)
Minimalist footwear*				68th (28.8)	68th (28.8)	68th (28.8)
Interleukin-1 inhibitors				69th (28.1)	69th (28.1)	69th (28.1)
Transdermal buprenorphine				70th (27.9)	70th (27.9)	70th (27.9)
Fibroblast growth factor				71st (26.3)	71st (26.3)	71st (26.3)
Methotrexate				72nd (24)	72nd (24)	72nd (24)
Arthroscopic meniscectomy				73rd (19.4)	73rd (19.4)	73rd (19.4)
Arthroscopic cartilage repair				74th= (13.5)	74th= (13.5)	74th= (13.5)
Arthroscopic lavage and debridement				74th= (13.5)	74th= (13.5)	74th= (13.5)

† For medial tibiofemoral knee OA

Table S17. Results of the full-sample ordinary least squares (OLS) regression exploring associations between sociodemographic characteristics and full-sample mean preference weights for the eight criteria.

	OLS Regression							
Variables	Accessibility	Risk Mild	Risk Serious	Effectiveness	Recommend-ation	Duration	Cost	Quality of Evidence
Healthcare providers	-0.010	0.013	-0.004	-0.000	0.039**	-0.039**	-0.012	0.014
(Ref: Consumers)	(0.013)	(0.015)	(0.018)	(0.015)	(0.015)	(0.019)	(0.011)	(0.012)
Health policy-makers	-0.017	-0.018	-0.014	0.019	0.021	-0.046***	0.008	0.047**
(Ref: Consumers)	(0.015)	(0.016)	(0.018)	(0.016)	(0.016)	(0.017)	(0.012)	(0.022)
Content area experts	-0.014	-0.005	0.034	-0.034	0.056**	-0.038	-0.006	0.006
(Ref: Consumers)	(0.019)	(0.016)	(0.035)	(0.021)	(0.023)	(0.026)	(0.012)	(0.026)
Female	0.004	0.004	0.009	-0.008	0.018	-0.014	0.002	-0.015
(Ref: Male)	(0.010)	(0.009)	(0.012)	(0.010)	(0.012)	(0.014)	(0.006)	(0.011)
Australian	0.006	-0.026**	0.013	0.012	-0.008	0.018	-0.011	-0.004
(Ref: New Zealander)	(0.016)	(0.013)	(0.026)	(0.017)	(0.022)	(0.024)	(0.015)	(0.022)
Works for a Gov. Agency	0.013	-0.018*	-0.004	0.016	-0.021*	0.010	-0.002	0.005
(Ref: No)	(0.009)	(0.011)	(0.013)	(0.013)	(0.013)	(0.016)	(0.006)	(0.011)
Age	0.000	-0.000	0.000	-0.001	0.001**	-0.000	0.000	-0.000
(years)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Years experience in primary role	-0.000	-0.001	0.000	-0.000	-0.000	-0.000	0.000	0.000
(years)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Constant	0.061**	0.113***	0.103***	0.190***	0.103***	0.178***	0.062***	0.191***
	(0.024)	(0.034)	(0.037)	(0.029)	(0.029)	(0.033)	(0.023)	(0.025)
Observations	178	178	178	178	178	178	178	178
R-squared	0.021	0.073	0.037	0.072	0.052	0.052	0.043	0.097

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regression diagnostics								
Ramsey RESET (Prob>F)	0.689	0.9216	0.4571	0.9981	0.5886	0.023	0.6275	0.281
Breusch-Pagan Test (Prob>Chi2)	0.012	0.0073	0.0179	0.9754	0.1828	0	0.0006	0.0094

Table S18. Average partial effects of the consumer group fractional multinomial logistic (FMNL) regression between consumer mean preference weights and gender, age and years lived with OA. The regression coefficients indicate how the weight of each criterion depends at the margin on each sociodemographic variable

	Average Partial Effects							
Variables	Accessibility	Risk Mild	Risk Serious	Effectiveness	Recommend-ation	Duration	Cost	Quality of Evidence
Male	-0.023	0.029	0.009	0.020	0.007	-0.035	0.013	-0.020
(Ref: Female)	(0.020)	(0.017)	(0.021)	(0.016)	(0.021)	(0.027)	(0.012)	(0.018)
Age	0.000	0.001	0.000	-0.002*	0.001	-0.000	-0.000	-0.001*
(at mean 63 years)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years living with OA	0.000	-0.001	0.000	-0.000	0.000	0.000	-0.000	0.000
(at mean 13 years)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	63	63	63	63	63	63	63	63

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table S19. Demographic characteristics of the healthcare provider group (N=70)

Healthcare provider role	Gender (n)		Age (years)		Experience in primary role (years)		Works for a government agency (n)	
	Male	Female	Mean	SD	Mean	SD	Yes	No
Physiotherapist (n=17)	8	9	40	11	14	10	8	9
Orthopaedic surgeon (n=14)	13	1	53	10	21	12	0	14
Nurse (n=11)	1	10	51	10	17	10	7	4
General Practitioner (n=28)	9	19	50	11	20	10	25	3
Pharmacist (n=3)	0	3	40	11	19	10	1	2
Occupational Therapist (n=2)	0	2	46	25	8	2	0	2
Geriatrician (n=2)	1	1	54	12	25	18	0	2
Rheumatologist (n=2)	2	0	61	1	30	6	1	1

Table S20. Results of the health care provider group fractional multinomial logistic (FMNL) regression exploring associations between health care provider mean preference weights and sociodemographic characteristics. The regression coefficients indicate how the weight of each criterion depends at the margin on each sociodemographic variable.

	Average Partial Effects							
Variables	Accessibility	Risk Mild	Risk Serious	Effectiveness	Recommendation	Duration	Cost	Quality
Physiotherapist	0.003	0.024	0.042	-0.019	-0.064**	0.020	-0.004	-0.002
(Ref: General Practitioner)	(0.013)	(0.020)	(0.023)	(0.017)	(0.022)	(0.017)	(0.011)	(0.019)
Orthopaedic surgeon	0.012	-0.004	-0.001	-0.008	-0.046	0.010	0.004	0.033
(Ref: General Practitioner)	(0.021)	(0.020)	(0.030)	(0.027)	(0.032)	(0.025)	(0.015)	(0.024)
Nurse	0.001	0.010	-0.036*	-0.043*	-0.028	0.002	0.004	0.090***
(Ref: General Practitioner)	(0.020)	(0.018)	(0.017)	(0.021)	(0.040)	(0.024)	(0.012)	(0.024)
Works for a Government Agency	0.020	-0.018	0.019	0.002	-0.042	-0.011	0.004	0.026
(Ref: No)	(0.013)	(0.015)	(0.019)	(0.016)	(0.025)	(0.019)	(0.009)	(0.017)
Female	0.021	-0.026	-0.014	0.003	0.015	0.022	-0.004	-0.017
(Ref: Male)	(0.012)	(0.015)	(0.016)	(0.014)	(0.021)	(0.013)	(0.009)	(0.014)
Age, years	-0.001	-0.000	0.001	-0.000	0.000	0.002	0.000	-0.001
(at mean 48 years)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Work experience in primary role, years	0.001	-0.000	0.000	-0.001	-0.000	-0.002	0.000	0.002*
(at mean 18 years)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Observations	70	70	70	70	70	70	70	70

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table S21. Results of Kendall's coefficient of concordance indicate a very strong level of agreement across groups (W=0.990, p<0.000)

Intervention	**N**	**Mean rank**	**Std. Deviation**	**Minimum rank**	**Maximum rank**
Exercise - Tai Chi only	4	1.0	0.0	1.0	1.0
All land-based exercise	4	2.0	0.0	2.0	2.0
Exercise - muscle strengthening only for lower limb strengthening	4	4.5	2.0	3.5	7.5
Exercise - muscle strengthening only for quadriceps strengthening	4	4.5	2.0	3.5	7.5
NSAIDs (topical)	4	5.5	1.9	3.0	7.0
Collagen	4	5.8	1.3	4.5	7.5
Pycnogenol	4	5.8	1.3	4.5	7.5
Duloxetine	4	7.8	2.1	6.0	10.0
Aquatic exercise	4	8.5	0.6	8.0	9.0
Curcuma/ curcuminoid	4	10.0	0.8	9.0	11.0
Exercise - walking only	4	11.3	1.3	10.0	13.0
Massage	4	11.8	0.5	11.0	12.0
Walking cane*	4	13.0	0.8	12.0	14.0
Exercise - yoga only	4	14.5	0.6	14.0	15.0
Exercise - stationary cycling only	4	15.8	2.5	13.0	19.0
Therapeutic ultrasound	4	17.0	3.6	14.0	22.0
Cognitive behavioural therapy	4	18.3	2.1	16.0	21.0
Pulsed electromagnetic/ shortwave therapy	4	19.0	2.8	17.0	23.0
Avocado-soybean unsaponifiables	4	19.5	2.0	16.5	20.5
Boswellia serrata extract	4	19.5	2.0	16.5	20.5
Corticosteroid injection	4	20.0	2.3	18.0	22.0
Methylsulfonylmethane	4	21.5	2.4	19.0	24.0
NSAIDs (oral)	4	21.5	4.9	16.0	27.0
Exercise - stationary cycling & hatha yoga	4	24.3	2.1	22.0	26.0

Intervention	N	Mean rank	Std. Deviation	Minimum rank	Maximum rank
Vitamin D	4	24.3	1.0	23.0	25.0
Chondroitin	4	27.0	1.9	24.5	28.5
Glucosamine	4	27.0	1.9	24.5	28.5
Dextrose prolotherapy	4	27.8	2.1	25.0	30.0
Heat therapy*	4	28.3	1.5	27.0	30.0
Weight management	4	29.8	3.0	26.0	33.0
Paracetamol	4	31.3	1.0	30.0	32.0
TENS	4	31.3	1.0	30.0	32.0
medial wedged insoles for lateral tibiofemoral OA and valgus deformity	4	32.5	1.0	31.0	33.0
Other electrotherapy - shockwave	4	34.0	0.0	34.0	34.0
Acupuncture - electroacupuncture	4	36.8	1.3	35.5	38.5
Acupuncture - traditional with manual stimulation	4	36.8	1.3	35.5	38.5
Other electrotherapy - laser	4	37.0	2.3	35.0	39.0
Taping - patellar	4	37.5	1.0	36.0	38.0
Self-management & education	4	38.5	3.1	35.0	42.0
Other electrotherapy - interferential	4	39.8	0.5	39.0	40.0
Acupuncture - laser	4	42.3	1.5	41.0	44.0
Anti-nerve growth factor (NGF)	4	42.8	4.6	37.0	47.0
Knee braces - valgus unloading/ re-alignment braces	4	43.3	2.1	41.0	45.0
Mobilisation & manipulation	4	44.0	2.4	42.0	47.0
Total joint replacement*	4	45.0	2.7	43.0	49.0
Knee braces - re-aligning patellofemoral braces	4	45.5	1.0	44.0	46.0
Biphosphonates	4	46.3	4.3	43.0	52.0
Oral opioids	4	48.0	2.2	45.0	50.0
Doxycycline	4	48.8	1.0	48.0	50.0

Intervention	N	Mean rank	Std. Deviation	Minimum rank	Maximum rank
Stem cell therapy	4	49.0	1.8	47.0	51.0
Taping - kinesio	4	51.8	2.2	50.0	55.0
Transdermal Fentanyl	4	54.0	4.4	49.0	59.0
Colchicine	4	55.5	4.4	53.0	62.0
Footwear - unloading shoes	4	55.5	4.4	53.0	62.0
Shoe orthotics - lateral wedge insoles for medial tibiofemoral knee OA	4	55.5	4.4	53.0	62.0
Topical capsaicin	4	57.5	5.5	51.0	63.0
Cold therapy	4	58.0	4.1	55.0	64.0
Knee braces varus unloading/ re-alignment braces	4	58.8	1.5	57.5	60.5
Shoe orthotics - shock absorbing insoles or arch supports	4	58.8	1.5	57.5	60.5
Strontium ranelate	4	59.5	5.3	54.0	65.0
Glucosamine and chondroitin in compound form	4	60.5	3.6	57.5	65.5
Omega-3 fatty acids	4	60.5	3.6	57.5	65.5
Footwear - rocker soled shoes	4	61.3	1.5	59.0	62.0
Platelet-rich plasma injection	4	62.0	6.2	53.0	66.0
Diacerein (not available in NZ)	4	63.8	2.6	60.0	66.0
Viscosupplementation injection	4	63.8	5.3	56.0	67.0
Calcitonin	4	65.5	1.7	64.0	67.0
Footwear - minimalist footwear	4	68.5	0.6	68.0	69.0
Interleukin-1 inhibitors	4	69.3	1.0	68.0	70.0
Transdermal buprenorphine	4	69.3	1.0	68.0	70.0
Fibroblast growth factor (not available in NZ)	4	71.0	0.0	71.0	71.0
Methotrexate	4	72.0	0.0	72.0	72.0
Arthroscopic meniscectomy	4	73.0	0.0	73.0	73.0
Arthroscopic cartilage repair	4	74.5	0.0	74.5	74.5

Intervention	N	Mean rank	Std. Deviation	Minimum rank	Maximum rank
Arthroscopic lavage and debridement	4	74.5	0.0	74.5	74.5

Table S22. Average partial effects of the weighted FMNL regression for equal stakeholder representativeness

	Average partial effects adjusted for equal subgroup sample size							
Stakeholder group	Recommendation to use the intervention now	Quality of evidence about the intervention	Effectiveness of the intervention	Duration of the intervention effect	Risk of serious harm	Risk of mild to moderate side-effects	Cost of the intervention	Accessibility to the intervention
Providers (ref: consumers)	0.047 (0.016)**	0.014 (0.015)	-0.001 (0.014)	-0.051 (0.019)**	-0.001 (0.02)	0.008 (0.014)	-0.012 (0.01)	-0.001 (0.013)
Policy-makers (ref: consumers)	0.030 (0.017)	0.045 (0.021)*	0.02 (0.015)	-0.051 (0.016)	-0.02 (0.02)	-0.022 (0.014)	0.007 (0.011)	-0.008 (0.015)
OA researchers (ref: consumers)	0.062 (0.026)*	0.003 (0.019)	-0.027 (0.018)	-0.041 (0.021)	0.037 (0.026)	-0.023 (0.018)	-0.01 (0.011)	0.001 (0.016)
Policy-makers (ref: Providers)	-0.016 (0.015)	-0.016 (0.015)	-0.016 (0.015)	0 (0.015)	-0.018 (0.014)	-0.03 (0.015)	0.020 (0.008)*	-0.007 (0.011)
OA researchers (ref: Providers)	0.014 (0.019)	0.014 (0.019)	0.014 (0.019)*	0.01 (0.021)	0.039 (0.022)	-0.032 (0.011)**	0.001 (0.008)	0.002 (0.013)
OA researchers (ref: policy-makers)	0.031 (0.023)	-0.042 (0.023)	-0.048 (0.015)**	0.01 (0.023)	0.058 (0.025)*	-0.001 (0.02)	-0.018 (0.009)	0.009 (0.016)
Female (ref: male)	0.022 (0.013)	-0.023 (0.014)	-0.014 (0.009)	-0.014 (0.013)	0.029 (0.014)*	0.002 (0.008)	-0.005 (0.006)	0.004 (0.009)
Australian (ref: New Zealander)	-0.023 (0.020)	-0.010 (0.018)	0.013 (0.015)	0.022 (0.02)	0.035 (0.017)	-0.02 (0.017)	-0.009 (0.015)	-0.007 (0.016)

DHB or MoH employee (ref: other employer)	-0.028 (0.013)*	0.001 (0.013)	0.019 (0.011)	0.023 (0.015)	-0.002 (0.016)	-0.013 (0.012)	-0.003 (0.006)	0.003 (0.01)
Age (at 54 years old)	0.001 (0.001)*	-0.001 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Work experience (at 16 years)	0.000 (0.001)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

*p<0.05, **p<0.01, standard error are in parentheses, "goodness of fit" wald Chi-Square p<0.001

After weighting the regression for equal group sample size, the regression produced similar results to the unweighted regression. For those with a significance of *p*<0.01, the differences between stakeholder groups were again small, and no larger than 5.1 percent points, or equivalently, a change in <2 ranks. On average, health care providers placed 4.7% more importance on Recommendation and 5.1% less importance on Duration (*p*<0.01), relative to the other criteria. On average, OA researchers placed 3.2% less importance on Risk-Mild compared to healthcare providers, relative to the other criteria. Compared to policy-makers, OA researchers on average placed 4.8% less importance on Effectiveness, relative to the other criteria.

Cluster analysis

Cluster analysis was performed to identify groups of participants with similar patterns of weights[403] using the NBClust package (ver.3) in R (R Core Team 2013, Vienna, Austria). Clusters were formed using recommended distance and linkage methods.[405] Chi-square and ANOVA were used to assess differences between clusters. Multinomial logistic regression was used to investigate if the clusters were associated with their members' socio-demographic characteristics.

Results

Two clusters were identified: cluster 1 had higher preference weights for Risk-Serious, Risk-Mild, Cost and Accessibility whereas cluster 2 had higher weights for Recommendation, Quality and Effectiveness (Table S23). No noteworthy socio-demographic characteristics were significantly associated with cluster membership in the logistic regression model (Table S24).

Interpretation

The log likelihood ratio of the regression model was not significant (p=0.348) indicating that none of the independent variables were significantly associated with the dependent variables: it cannot be ruled out that all the coefficients in the model are zero. This could be caused by noise in the data, small sample size, differences so small they're not detectable, or because the variation is not explained by the included sociodemographic variables. Including these other variables could enhance the LR chi-square result/precision of the data. Goodness of fit was assessed using pearson's chi2 (p=0.286) and Hosmer-Lemeshow chi2 (p=0.511). Both of these statistics were not significant, suggesting that the model was not mis-specified.

Table S23. Descriptive comparison between clusters 1 (n=121) and 2 (n=57) preference weights. The total sample (n=178) preference weights are included for reference.

	Total sample weight (SD) n= 178	Cluster 1 weight (SD) n= 121	Criteria rank	Cluster 2 weight (SD) n= 57	Criteria rank
Criteria					
Recommendation for using the intervention now	0.190 (0.072)	0.179 (0.050)	1	0.216 (0.089)†	=1
Quality of the evidence	0.176 (0.064)	0.158 (0.452)	2	0.215 (0.081)†	=1
Effectiveness of the intervention	0.150 (0.062)	0.132 (0.048)	=3	0.188 (0.070)†	3
Duration of the intervention effect	0.132 (0.073)	0.135 (0.074)	=3	0.126 (0.710)	4
Risk of serious harm	0.128 (0.070)	0.141 (0.044)†	4	0.100 (0.041)	5
Risk of mild to moderate side-effects	0.094 (0.060)	0.106 (0.063)†	6	0.068 (0.040)	6
Cost of the intervention	0.066 (0.041)	0.072 (0.433)*	=7	0.054 (0.033)	7
Accessibility to the intervention	0.063 (0.055)	0.076 (0.059)†	=7	0.034 (0.284)	8
Sociodemographic characteristics (%)					
Primary area of work					
Consumers	35	41*		25	
Providers	44	41		51	
Policy-makers	14	11		19	
OA researchers	7	7		5	
Gender (female)	64	65		61	
Australian	8	8		7	
Works for a New Zealand government agency	33	31		39	
Mean years work experience (SD)	16 years (11.2)	15 years (10.9)		17 years (11.8)	
Mean age (SD)	54 years (12.7)	55 years (13.4)		52 years (11.0)	

*p<0.05, †p<0.001 significantly higher criterion weight compared to the other cluster

Pearson's t-test for continuous data comparisons

Chi-squared test for categorical comparisons

Table S24. Results of the regression on cluster assignment (reference: cluster 1)

				95% CI	
Variables	Odds Ratio	Std. Err.	p-value	Lower	Upper
Gender (female)	0.8421911	0.299236	0.629	0.419733	1.68985
Age	0.9875509	0.0177798	0.487	0.9533108	1.023021
Work Experience (years)	1.01845	0.0178237	0.296	0.9841082	1.053989
Works for a New Zealand government agency	1.354947	0.5591423	0.462	0.6034753	3.042183
Providers	1.321861	0.6996266	0.598	0.4684543	3.729964
Policy-makers	2.358383	1.34844	0.133	0.7690017	7.232714
OA researchers	0.7725362	0.645305	0.757	0.150281	3.971308
Reference category: Consumers	0.5365136	0.5927098	0.573	0.0615491	4.676705

Goodness of fit pearson’s chi2 (p=0.286) and Hosmer-Lemeshow chi2 (p=0.511); log likelihood ratio p=0.348

Sensitivity analysis

Perceived value-for-money charts

For the 15 guideline-recommended interventions, 'value-for-money' charts were created for all phases of OA care so that the total score of the interventions could be assessed subject to their cost and quality of evidence separately.[297] This analysis may help to address issues associated with these criteria such as double-counting.[297, 439, 440] Hence, the weights for the Cost and Quality criteria were redistributed among the remaining six criteria and the interventions total scores recalculated. 'TJR' was an outlier, in terms of very high initial cost and long duration of treatment effect relative to the other interventions. It was therefore excluded from the sensitivity analysis. The 'efficiency frontier' was plotted on the charts to highlight potential trade-offs between interventions where higher cost can be compensated by higher value.

Results

The perceived value-for-money chart for first-line care (Figure S1) suggest that a number of interventions, subject to their cost and quality of evidence, offer value-for-money at first-line care relative to the other interventions. For example, if an arbitrary threshold of intervention cost is set at <$600 per month and a total score of >67% is selected, then five interventions - 'all land-based exercise'([total score] 81.5%), 'aquatic exercise' (76.5%), 'corticosteroid injection' (70.4%-76.0%), 'NSAIDs (topical)' (69.0%-75.1%) and 'walking cane' (69.0%-74.5%) - represent value-for-money at first-line care (i.e. the interventions in the top-left quadrant of Figure 4). Changes to the total scores from first- to second-line care for 'duloxetine' (66.9%-72.4%) and 'heat therapy' (66.3%-71.8%) also suggest that these interventions may offer perceived value-for-money later in disease management. The total score for TJR at first-, second- and third-line care was 58.2%, 64.3% and 74.8%, respectively (one-off cost 2017NZD $24,050 [$17,500-$30,600]).

The interventions on the efficiency frontier which offer perceived value-for-money in decreasing order of importance at first-line care are 'all land-based exercise', 'aquatic exercise', 'corticosteroid injections', 'walking cane', 'heat therapy' and 'NSAIDs (oral)'. For second-line care, the relative importance of interventions on the efficiency frontier mirrored first-line care, except 'massage' now appears after 'all land-based exercise'.

Interpretation

These findings suggest that the overall results are robust to confounding caused by the cost and quality of evidence criteria, given the similar pattern of interventions showing good perceived value-for-money compared with the rankings of the interventions scored across all the criteria.

Figure S1. Perceived value for money chart. Interventions along the efficiency frontier highlight potential trade-offs between interventions where higher cost can be compensated by higher value.

Disaggregated intervention scores

Figue description. The top left figure shows the aggregated intervention scores (i.e. total scores), with each colour representing one of the eight criteria. Each figure thereafter shows the weight and corresponding level of achievement by criterion for the same interventions (disaggregated).

The top left figure shows the aggregated intervention scores (total scores). Each list of interventions thereafter shows the weight and corresponding level of achievement for the criteria (i.e. disaggregated).

CHAPTER 7: META-DISCUSSION AND CONCLUSION

7.1 SUMMARY OF THE RESEARCH OBJECTIVES

The overall aim of this body of research was to establish a framework to prioritise interventions for managing knee osteoarthritis (OA) and evaluate the relative importance of these interventions across the healthcare sector in New Zealand (NZ) using multi-criteria decision analysis (MCDA).

The background chapter established that OA is a highly prevalent, long-term disease that is a global public health issue, with a disproportionate response from health systems despite the burden it imparts. Although high quality clinical practice guidelines (CPGs) for OA consistently recommend exercise, self-management education and weight loss (if appropriate), significant evidence-practice gaps persist. This can be attributed to a complex interplay of implementation barriers and enablers at every level of a healthcare system; from clinical encounters and patient's self-care choices through to the system manager. Strong evidence indicates that it is unlikely that any single strategy can maximise the implementation of recommended OA care in primary care settings due to the complexity of this problem.[27] However, addressing context-sensitive factors that influence implementation success, such as cross-sectoral stakeholders' preferences for the characteristics of OA interventions, may help to reduce the overall resistance to a given change effort. In turn, this may support the uptake of high-value OA care. Establishing the 'fit' of interventions in a health system may help achieve this goal.

The ‘fit’ of an intervention in this thesis is conceptualised as *the aggregation of the preferences of cross-sectoral stakeholders with evidence and recommendations about OA care from a CPG using MCDA*. The background chapter established that the literature mostly concerns the preferences of patients with OA and healthcare providers for the characteristics of drug interventions – for example their therapeutic benefits, risk profile and cost – with limited examination of the preferences of cross-sectoral stakeholders (Section 2.13). There is also scant evidence concerning which interventions stakeholders want in a CPG for managing OA, with only one study investigating physiotherapists’ preferences for OA CPG recommendations.[441] Thus, the background chapter confirmed a strong rationale to research what matters to cross-sectoral stakeholders when choosing or recommending OA interventions, and the ‘fit’ between their preferences and evidence and recommendations about OA interventions in a CPG.

7.1.1 Outcomes of the first research objective

To identify what criteria are important to cross-sectoral stakeholders when choosing or recommending OA interventions in NZ and their relative importance.

The first research objective produced a thematic framework of three over-arching categories and sub-categories, derived from the 56 themes generated by the focus group discussions. Nine characteristics of interventions were ranked by importance following the eDelphi survey (Figure 21). The characteristics of interventions informed the second Objective, which was to create a MCDA model for prioritising OA interventions.

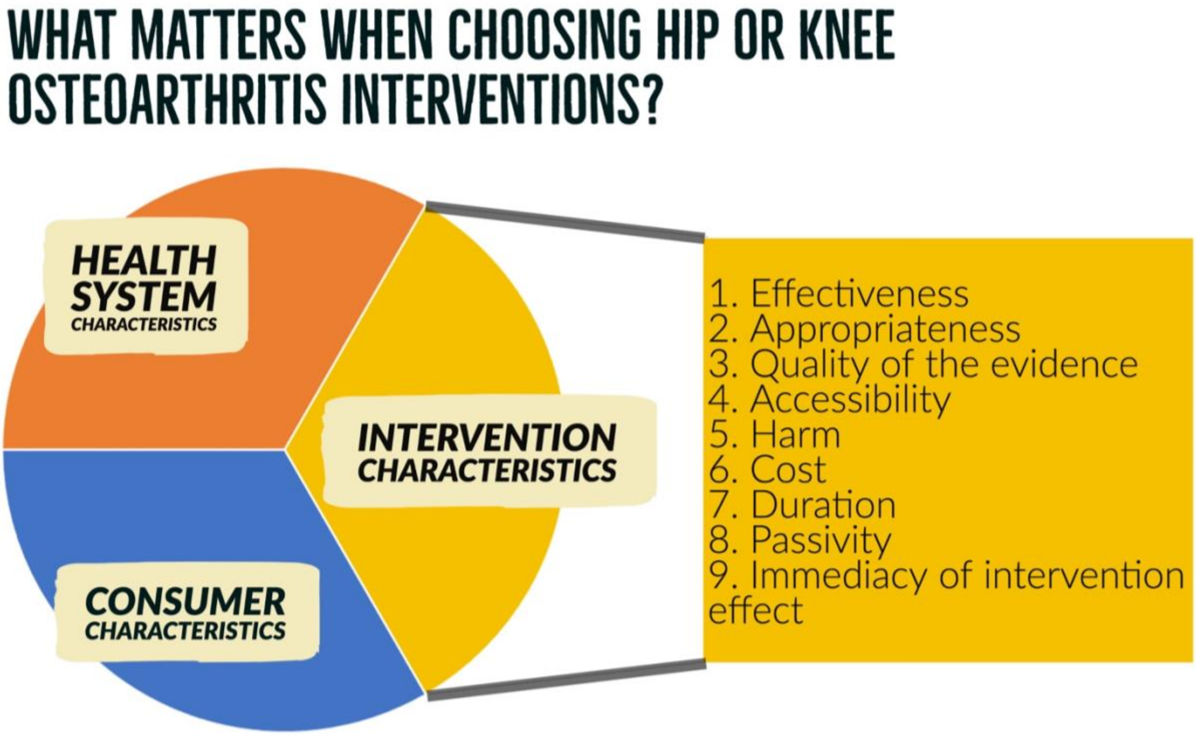


Figure 21. Thematic framework of three overarching characteristics of the: health system, consumer and intervention that cross-sectoral stakeholders consider when choosing or recommending hip or knee OA interventions in NZ. The nine characteristics of interventions presented in the figure, are in decreasing order of importance.

7.1.2 Outcomes of the second research objective

To use MCDA to: (i) evaluate and prioritise knee OA interventions according to stakeholders' preferences and evidence about intervention performance on criteria identified in Objective 1, and (ii) to explore if stakeholders' preferences differed by group.

The results, after specifying the criteria identified in Objective 1 for MCDA, were:

1. The relative importance of eight criteria, as weights (%), representing stakeholders' preferences (in decreasing order of importance):
    1. Recommendation: 19.0%
    2. Quality of evidence: 17.7%
    3. Effectiveness: 15.0%
    4. Duration of effect: 13.2%
    5. Risk of serious harm: 12.8%
    6. Risk of mild side-effects: 9.4%
    7. Cost: 6.6%
    8. Accessibility: 6.3%.
2. The relative importance of the intervention's 'total score', as the sum of the weights associated with the performance ratings on the criteria for a given intervention:
    - For first-, second- and third-line OA interventions respectively (as defined by the RACGP CPG management algorithm[13]), all land-based exercise (total score= 71.7%), NSAIDs (topical) (74.2%) and total joint replacement (74.3%) were ranked first
    - For first-, second-, and third-line OA care, the recommended core interventions of weight management and self-management education ranked between 11th and 15th (total score= 48.0% to 56.0%).
3. Regression analysis identified only small differences in weights ($\leq 5.7\%$; $p<0.01$) between stakeholder groups, suggesting that stakeholders' preferences do not meaningfully differ by group.

The original research contributions from this work, and their potential for impact are summarised in Table 23.

Table 23. Original research contributions in this thesis and their potential impact in policy and practice

Original research contribution	**Impact**
An empirically derived thematic framework representing what influences cross-sectoral stakeholders' choice of intervention for OA intervention in NZ	• To support stakeholder engagement about what to consider (e.g. barriers and incentives for change) for system-strengthening OA care reform, such as Models of Care and other programs, services and policies.[171] For example, elements of the framework could inform content in established domains of 'readiness' and 'success'[292] for an OA Model of Care in NZ (which is currently absent in NZ[8]). • To support better patient-centred care by alerting health care providers to the characteristics of consumers and interventions that may enhance their clinical practice.
Within the thematic framework, the identification of nine generalisable characteristics of interventions, ranked by importance	• To alert health service decision makers, such as health service planners, funders and coordinators, about the characteristics of interventions that influence implementation in the NZ healthcare system. • To inform health care providers about the characteristics of interventions that matter to patients in their pursuit of person-centred, value-based care and to directly inform shared decision making about OA care.
A MCDA framework for prioritising OA interventions in a clinical practice guideline through the integration of context-sensitive preference data, clinical guideline recommendations and evidence about their performance on the criteria	• To guide the identification of OA care with good 'fit' in the development of Models of Service Delivery[171, 230] or tailored care bundles.[442] • To support the evaluation of existing programs/service delivery for 'fit' within the NZ health system. • To test intervention 'fit' at a population level as a 'proof-of-concept' prior to upscaling local Models of Service Delivery to a national level; and potentially other musculoskeletal conditions, e.g. pain,[15] as a policy tool for the health system. • To support the evaluation, development, redesign or regulation of these, or other MoSDs for NZ, by involving and informing decision-makers about what interventions for OA care people want across the health system. • To enhance the utility of clinical practice guidelines by providing a prioritised list of interventions, based on the preferences of OA patients, at a population level. • To support patient-centred care as a tool to guide the selection of interventions in CPGs for OA consumers and health care providers.

Original research contribution	Impact
	• A framework for facilitating transparent, inclusive and rigorous evidence-based and stakeholder-informed decision-making for complex, systemwide implementation problems involving cross-sectoral stakeholders – e.g., in the formulation of recommendations for OA-related public health care, clinical care standards and clinical practice guidelines.
Insight into the preferences of stakeholders in NZ: stakeholders' preferences do not meaningfully differ by stakeholder group, and the performance of two first-line (core) interventions – self-management education and weight management – do not align with their preferences, suggesting that their preferences may represent a barrier to their system-wide implementation.	• To inform healthcare planners, coordinators and providers to population groups which may have specific OA care preferences, thus potentially requiring specifically tailored services • To provide evidence for OA care decision-makers such as policy-makers, about the need to embark on system-strengthening efforts to overcome system-wide barriers to the uptake of high-value care for OA (i.e., weight loss (if appropriate) and education to self-manage).

7.2 META-DISCUSSION OF FINDINGS

Good health care planning and delivery should result in the right care, delivered at the right time, by the right team, in the right place, with the right resources.[171] That is, OA care that is delivered early in the disease course, uses safe and clinically and cost effective interventions (high-value), coordinated and patient-centred; which generates good patient and providers experience, good health care outcomes, and sustainable health care costs – the quadruple aim of value-based care. For people living with OA however, a tendency to deliver acute and episodic care has largely resulted in unmet population OA health care needs, and the tendency to delivery low-value care.

It is clear that evidence alone and discrete evidence dissemination strategies such as CPGs, care paradigms (e.g. patient-centred care) and delivery/service programs in and of themselves are not enough to address the complex nature of implementing high-value OA care at scale. In light of this complexity, thinking about how to create value across a whole healthcare system could help to coordinate, redesign and calibrate discrete strategies to address implementation complexity and support the delivery of value-based care.

One way to support the generation of value-based care is to align the choice of OA interventions more closely with what matters to people – i.e. their preferences and values – with the evidence and recommendations about OA interventions in CPGs. The remainder of this section will describe how the research in this thesis can support and guide implementation of value-based care in the context of an interconnected healthcare system, as seen in the conceptual framework described in Figure 22.

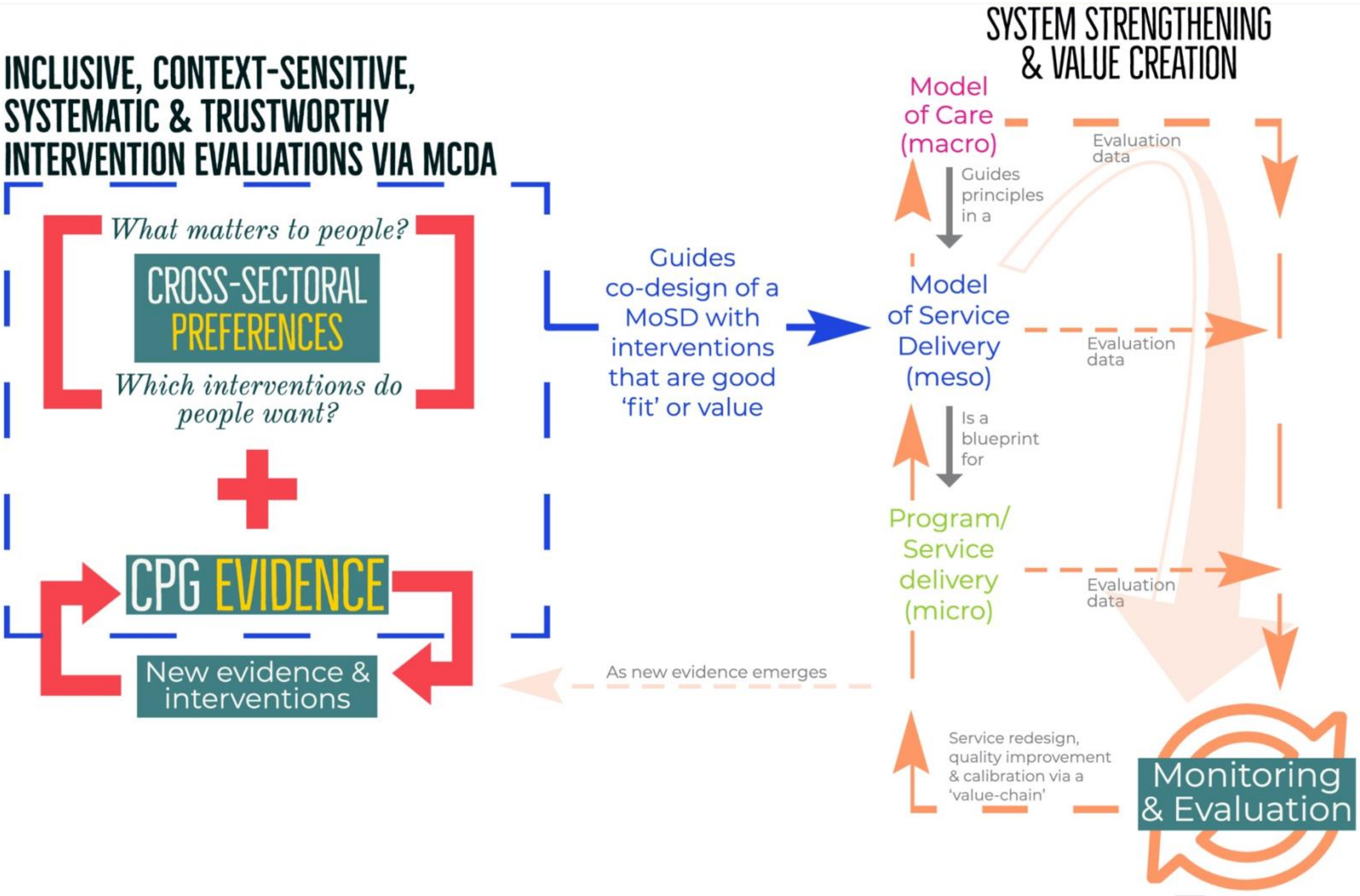


Figure 22. Conceptual framework of the relationship between the MCDA framework in this thesis and a model of system-strengthening healthcare. The MCDA framework in this thesis guides the co-design of Models of Service Delivery (MoSD) by informing which interventions offer good 'fit' in a health system for implementation. In a system-strengthening health system, all level of the health system are aligned to coordinate and support

best-practice care, starting with macro-level Models of Care (MoC), through MoSD, to the micro-level of service delivery and the patients' participation in their own care. Evaluation data relevant to each level of the health system contribute to a continuous cycle of monitoring and evaluation, which help calibrate health system redesign and quality improvement; including the evidence informing clinical practice guidelines (CPG) recommendations.

7.2.1 Theoretical implications for implementation

Conceptual frameworks[170, 443] and empirical evidence[27, 28] point toward integrating contextually sensitive information into strategies aiming to close evidence-practice gaps. A premise of implementation theory is that strategies which are tailored to particular stakeholders and settings, are more likely to create change that is sustainable and scalable. Whilst the research in this thesis does not empirically test this hypothesis, it does address an important evidence-gap in the literature about how to systematically link contextually-sensitive information with strategies in a national system context.[444] That is: an MCDA framework for systematically, inclusively and transparently combining cross-sectoral stakeholders' preferences with evidence and recommendations from a CPG for OA in the NZ healthcare system. This is important for supporting the delivery of value-based care by capitalising on opportunities to incorporate what people value in areas in and of clinical practice, CPG and policy-making decisions, health service design and evaluation.

7.2.2 Clinical practice implications

This research has implications for improving the utility of CPGs for end-users such as OA consumers and providers. For example, the criteria, their weights, intervention performance ratings (performance matrices) and the interventions ranked by importance, could be reported in CPGs to enhance the simplicity and design of CPGs.[445] This could potentially help consumers and providers navigate CPGs, and help providers to deliver care that is more closely aligned with what people want. An example of an interactive tool that could allow users to see their preferences for OA interventions, based on the approach and data in this thesis for example, has been translated to a website that could be used by patients and providers alike (https://cmor.shinyapps.io/cmor-evidence-table/).

While it is difficult to compare choice-based studies directly, due to methodological differences such as question framing, criteria selection and methods, it is likely that patient preferences are not homogenous at the individual level given the typically heterogeneous presentation of OA. For example, while there were no meaningful between or within-group differences revealed in the main research of this thesis, others have reported that patient preferences are not homogenous[38, 285] and may, for example, be influenced by disease severity[269, 280, 284] and other social determinants of health.[446, 447]

With respect to differences between providers' and consumers' preferences, three studies[269-271] have examined these groups together and all three studies found that patients' and healthcare providers' preferences were different. However, the criteria selection were markedly different: two studies[270, 271] focussed on the characteristics of drug interventions and did not include criteria such as accessibility to the treatment or cost of the treatment, and the third study[269] included criteria that were process related or non-health outcomes (e.g. duration of consultation, access to specialist equipment and involved healthcare providers).

In contrast to the current research, which found that preferences did not differ by group (Table 22), there are points of difference worth mentioning. For example, the survey instrument was developed to be highly generalisable for use among cross-sectoral stakeholders, and the survey considered criteria and preferences relevant to the NZ context. The sample size was also modest (N=178), and OA consumers' disease severity was not collected, which may have limited the ability to detect within-group differences. It would therefore be prudent to replicate the study in a larger sample, perhaps with better characterisation of OA consumers' health status, to more thoroughly evaluate their preferences.

Another key finding was that people living with OA did not prefer self-management education programs or weight loss. Given that the MCDA framework reveals the relative importance of the characteristics of interventions, this knowledge could be used by healthcare providers to better understand patients' preferences, and better deliver these recommended interventions to reduce the burden of OA. For example, knowledge of patients' preferences for risk of serious harm may help to juxtapose self-management education and weight-loss relative to other interventions that may incur greater risk of harm, which they may not be alert to at a level that is commensurate with their preferences.

7.2.3 MCDA for evidence translation

Developers of CPGs and policy-makers are faced with making recommendations about the appropriateness of interventions or treatments on behalf of the stakeholders they represent. According to the 'accountability for reasonableness' framework[448] priority setting should be based on reasons and criteria that are relevant and valued by the relevant stakeholders. Therefore, CPG 'expert' development group opinions or preferences should be representative of stakeholders' preferences, however, this is not always guaranteed.[449, 450] For CPGs and policy-makers, broad stakeholder engagement is typically constrained with limited patient engagement,[30-34] perhaps in the absence of health networks.[293] Furthermore, the trade-offs driving decisions may not always be explicit. These factors may compromise the trustworthiness of CPGs and highlight an important limitation with respect to current methods for developing guideline and policy recommendations.

For example, a comparison of recommendations between two CPGs for OA found that methodological differences used to reach consensus could potentially be harmonised by better standardising of procedures and broader stakeholder engagement.[451] This thesis outlines a decision-making framework that potentially addresses both of these problems.

The MCDA framework in this thesis involved participants making explicit pairwise trade-offs between 8 criteria (Table 21): recommendation to use the intervention now, quality of evidence, effectiveness of treatment, duration of the intervention effect, risk serious harm, risk mild to moderate side effects, intervention cost and accessibility. This conveys several benefits:

1. Transparency, because decision-makers' preferences for the criteria are revealed in a structured, systematic process.
2. Inclusivity, which simultaneously allows for a broader range and number of stakeholders to access complex decisions while making better decisions, including OA consumers/patients.
3. Efficiency, because the PAPRIKA method ranks all current and hypothetical interventions
4. Currency, both preference data and intervention performance data can be updated separately and on a rolling basis to prioritise interventions as needed.

The MCDA framework also shares similarity to the GRADE Evidence to Decision (EtD) framework[217] for health system and public health decisions.

The GRADE EtD framework is a decision-making framework – much like MCDA – which steps from problem identification, through making trade-offs about alternatives (e.g. interventions), to producing recommendations or decisions. Compared to the MCDA framework presented in this thesis, the GRADE EtD framework includes similar components to inform decisions in a transparent and systematic way: criteria, research evidence, contextual factors and the judgement of decision-makers.

The criteria considered in the GRADE EtD conceptual framework include:[217, 417] priority of the problem, test accuracy, benefits and harms of the options, certainty of the evidence, outcome importance, values and balance of effects, resources required, cost-effectiveness, equity, acceptability and feasibility. These criteria overlap with those in this thesis' MCDA framework. However, the GRADE EtD does not specify how to perform the trade-offs between criteria and decision-makers' preferences. Nor were the criteria defined by users in a local setting, which is a critical component of deriving intervention 'fit' in a local context. Thus, the MCDA framework in this thesis showcases how MCDA can be used as a robust framework to achieve the same goal as the GRADE EtD framework, with enhanced specification of context-specific criteria and stakeholders' preferences: transparent and systematic decision-making based on the best available evidence to support decisions for guidelines and/or policy.

7.2.4 System strengthening implications

This research in this thesis highlighted that two out of three first-line interventions – self-management education and weight loss – did not align with stakeholders' preferences for intervention criteria (Figure 19). This finding suggests that stakeholders' preferences are likely to pose a barrier to the uptake of these interventions across the healthcare system. This raises important questions about how recommended OA care can be delivered and more generally how healthcare systems can be strengthened by harnessing the preferences and values of patients and other stakeholders.

Engagement and communication are critical tools in creating change. For patient-centred care, engaging with patients at every level of the health system can help to close the evidence practice gap, increase transparency and patients' adherence to treatment, and improve health system efficiency.[186] To support and guide a value-based healthcare system, opportunities exist to 'inject' value by involving patients – and other stakeholders – into health system reform, such as policy and regulation, as defined recently in a systematic appraisal of policy for NCDs.[159] In the context of the current research, the application of the MCDA framework lies in engagement, decision-making and a tool for establishing interventions with good 'fit'. These characteristics are particularly suited in the design or evaluation of Models of Service Delivery (MoSD), as described in the conceptual framework, seen in Figure 22.

Several evidence-based MoSD for OA care have been developed internationally.[165] For example, the Good Life with Arthritis in Denmark (GLA;D) programme, Better Management of patients with Osteoarthritis Sweden (BOA), the OA Chronic Care Program[201] and more recently the primary care management on knee pain and function in patients with knee osteoarthritis (PARTNER) model.[233, 452] In NZ, a systemwide MoSD could help drive evidence into practice by coordinating a national programme of OA care at the level of the health system which concerns service design and workforce capacity building initiatives (the meso level; e.g. health services, clinical workforce capacity and competencies; service delivery systems; funding models, and clinical infrastructure).

Currently, NZ is trialling and evaluating several musculoskeletal MoSDs aiming to improve earlier access to care, self-management education and intervention through the Mobility Action Programme.[453] The NZ Ministry of Health intends to transition MoSDs which are perceived as good value-for-money into systemwide, publicly funded services once the period of trial and evaluation has ceased, presumably sometime after the 2017/18 financial period. This research could support the evaluation, development, redesign or regulation of these, or other MoSDs for NZ, by involving and informing decision-makers about what interventions for OA care people want across the health system. As a result, local health system managers and coordinators could potentially focus healthcare resources on configuring health services that are more closely aligned with evidence and what people value to support sustainable and scalable implementation; value-based care.

Further system strengthening could also be achieved by coupling a national MoSD, informed by this research, with a national OA Model of Care (MoC). Importantly, MoCs are principle-based, macro-level policies which influence the building blocks of health systems, such health policy and strategy priorities, resourcing, workforce configurations, service delivery options and health governance decisions. In contrast, MoSD, operationalise aspects of MoC at the service (meso) level of the healthcare system, which concern health service delivery and workforce capacity building. Thus a MoC supports the operational activities outlined in MoSD by influencing macro-level health system levers to align with the features delivered in a MoSD.

In summary, a deliberate and coordinated approach to implementation across the health system, can help patient and health care providers achieve better health outcomes, better patient and provider experiences, and better use of healthcare resources – the quadruple aim of value-based health care.[438]

7.2.4.1 New Zealand perspective

Countries such as Australia and Canada have developed evidence-based and stakeholder-informed national and/or jurisdictional policies for musculoskeletal conditions (e.g. MoSD[454] and MoC[455]) as strategies to drive evidence into practice through closing the evidence-policy gap. While NZ does have strategies for long-term disease management (also referred to as non-communicable diseases in the NZ government; e.g. a healthy ageing strategy[456]), the guidance for OA management could be strengthened (e.g. there are no OA specific surveillance indicators), and there is also currently no national framework or strategy addressing the evidence-policy gap specifically for the management of OA.[8]

In NZ, decisions about which health interventions should be delivered are made by a narrow group of publicly-funded entities consisting of the Ministry of Health, Pharmaceutical Management Agency (PHARMAC), District Health Boards (DHB) and the Accident Compensation Corporation (ACC).[457] Each group has its own focus with respect to providing funding recommendations or advice about specific types of health interventions. For example, The Ministry of Health is responsible for establishing and funding programmes, planning, strategic prioritisation and national services. DHBs (currently 20) are charged with making decisions regarding how health services are configured in the areas they service, including which services are to be funded, for whom and where they should be located. PHARMAC is responsible for providing advice and recommendations with regard to pharmaceuticals, vaccines and medical devices, while the ACC scope is limited to interventions for injury care.

What is common to DHBs, PHARMAC and ACC is that the assessment, prioritisation and final funding decision about interventions are made within the same entity with Ministerial devolved authority.[458] Herein lies an opportunity for the Ministry of Health, DHBs and/or PHARMAC to deliver higher-value OA care by incorporating stakeholders' preferences for OA care across the health system (or within a DHB's jurisdiction) along the continuum of OA care (i.e. first-, second- or third-line OA care). These publicly-funded entities could consider adopting the framework described in this thesis to guide decisions about which interventions to promote as new interventions, recommendations or evidence about interventions for OA care emerge.

7.2.5 Strengths and limitations of the research in this thesis

The strengths and limitations of the discrete research activity conducted to achieve Objective 1 and Objective 2 are described in Chapters 3 and 5, respectively. This section will focus on the overarching strengths and limitations of the research in this thesis.

7.2.5.1 Strengths

7.2.5.1.1 Mixed-methods research

The overarching research approach in this thesis followed the 3 stage sequential equal-status mixed method design (QUAL → QUAN) → QUAN, described in Chapter 3, Section 3.2.2.2. A strength of this approach is that it enhances the strengths and minimises the weaknesses of either a purely qualitative or quantitative approach. The first two stages (underlined for emphasis) (QUAL → QUAN) → QUAN involved focus groups and eDelphi surveys conducted in two phases (Figure 13).

Phase I involved diverse stakeholders (people with lived experience of OA, Māori health advocates, healthcare providers, health policy-makers and OA experts) who were convened into six heterogenous focus group discussions using the Nominal Group Technique (NGT). Data from the individual focus groups were collated and then thematically analysed into a framework of categories and sub-categories to describe what influences stakeholders choice of OA interventions (Table 10 & Figure 15). Importantly, this approach generated data about the considerations of these cross-sectoral stakeholders in NZ and assessed valuable descriptions of their considerations, that would otherwise be difficult to attain using purely quantitative approaches.

Phase II involved a dual-panel, two-round eDelphi survey. The eDelphi strengthened the validity of the qualitatively-derived framework developed in phase I. It achieved this by quantitively evaluating the focus group participants' acceptability of the framework (which also included an independent and external content validation panel) and the relative importance of the characteristics of interventions identified in the framework (that would later inform the criteria for the MCDA framework; Figure 15 & Table 11).

The final and third stage of this mixed-methods approach (underlined for emphasis) (QUAL → QUAN) <u>→ QUAN</u> was informed by the output of the eDelphi survey: the relative importance of the characteristics of interventions. This provided a clear justification for which criteria to include in the MCDA framework. Incorrectly specifying the criteria in an MCDA framework may expose the trade-offs in an MCDA process to risk of omitted variable bias, causing spurious results. However, the overall three-stage sequential mixed-method approach adopted in this thesis helped mitigate this risk of bias.

7.2.5.1.2 Verification processes used throughout the thesis to monitor reliability and validity

This thesis included verification checks throughout each stage of research activity. In the first stage, the NGT process involved seeking confirmation from focus group participants at the conclusion of the meeting (NGT stage 5, Figure 12). Once the data were collated, a preliminary primary thematic framework was developed by the primary author, whilst two other researchers also independently thematically analysed the data. The three initial frameworks were then combined in a meeting arbitrated by a third and independent researcher, to confirm the reliability of the primary analysis (Section 4.5.2). These checks mitigated the potential risk of bias caused by only one investigator performing the thematic analysis.

In the second stage of research, a dual-panel eDelphi survey was used to evaluate the results of the thematic analysis (Figure 13). The acceptability of the framework was verified by the focus group participants (panel A) and also concurrently, by another panel to confirm the external content validity of the framework (panel B). These steps ensured that the thematic framework and the characteristics of interventions were acceptable to the eDelphi panellists before they were ranked by importance in the second eDelphi round.

The third stage of research involved three verification processes to assess internal validity and reliability. First, the choice-based survey was pilot-tested using cognitive interviewing to identify problems (e.g. reading, instructions) with the survey (SUPPLEMENT 9). Second, through the 1000minds survey platform, three survey questions were repeated to the survey participants to assess the reliability of their responses. This did not produce concerning results (Table S12). Third, the choice-based survey also included a response verification step (Section 6.3.4.3). Immediately after completing the choice-based survey, each participant reviewed their own results and evaluated the extent to which their criteria weights aligned with their overall intuitive and holistic evaluations of the relative importance of the criteria. The results of this process also did not produce concerning results (Section 6.4.1.2.1).

The multiple verification checks used to monitor the reliability and validity of the outputs at each stage of the thesis provide strong evidence for the reader to assess the overall internal validity of the research.

7.2.5.1.3 Representativeness of the study sample

Another strength of this research was the representativeness of the cross-sectoral participants included in this research. The background chapter established that studies investigating stakeholder preferences tend to focus mostly on people with a lived experience of OA or healthcare providers (see section 2.13). Throughout each stage of this mixed-methods research, a deliberate effort was made to recruit a representative sample of participants from each level of the health system; i.e. people living with OA, health care providers, Māori health advocates, health policy-makers and OA experts. The sample characteristics of the participants involved in the focus group and eDelphi survey (Objective 1) are shown in Table 9. While this may be considered a limitation because the sampling was non-probabilistic, the benefits of purposeful sampling include breadth and depth (especially so when paired with maximum heterogeneity sampling – which was the approach taken in stage 1 of this thesis) of a studied phenomenon for informing qualitative research. Similarly, for Objective 2, the sociodemographic characteristics of the respondents who completed the choice-based survey (Table 20) included a wide range of representatives across the NZ healthcare system, from the Ministry of Health (national), through District Health Boards (jurisdictional), to Public Health Organisations (primary care). Moreover, healthcare providers across the continuum of care (i.e. from primary care through to tertiary care such as general practitioners and orthopaedic surgeons, both in public and private settings), including representatives from the peak consumer advocacy group for arthritis in NZ and a broadly representative[2] cross-section of consumers with OA were included in the sample. Arguably, other sampling methods, such as simple randomised sampling may produce a sample that is of lower risk of sampling bias, however this is was not feasible, due to unknown population-level data required to conduct such a sampling method. Also other groups, such as carers, people from socioeconomically deprived areas and younger people experiencing OA may be underrepresented in the current sample. Notwithstanding these subgroups, the research in this thesis covered the primary groups experiencing[2] or delivering OA care in NZ and, as a blueprint or framework for future research, is generalisable of the NZ population in this respect.

7.2.5.1.4 MCDA framework

A strength of this study is the PAPRIKA method used in the MCDA framework.[369] 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 (e.g. Figure 18). This means that the relative importance of the interventions could not have been influenced by any preconceived bias for, or against certain interventions, irrespective of the evidence about the intervention's performance on the criteria. For example, a person may choose to discount any form of exercise therapy, regardless of its alignment with that individual's preferences and evidence about the intervention on the criteria.

The development of the MCDA framework was also multi-stage, involved a mixed-methods approach with multiple verification checks (see Sections 7.2.5.1.1 & 7.2.5.1.2 above). As a result of the rigorous development, involving broadly representative stakeholders in the NZ healthcare system, this approach produced an acceptable, valid and generalisable MCDA framework for the NZ setting. For example, the participants agreed their weights reflected their overall assessment of the importance of the criteria (median=agree; Table S11); there was virtually no difference between participants who answered all or none of the consistency checks correctly (<2.4% weight difference; Table S12), and the majority of participants felt the survey process was not difficult to complete (Table S13 & Table S14).

Due to the separation of preference data and intervention data in MCDA, this method of prioritising the interventions also 'future-proofs' these results against new interventions or evidence about the interventions currently unavailable. For example, interventions' performance were rated on the criteria using data extracted from the 2018 RACGP CPG[13] evidence summary tables; the ratings were summarised into performance matrices (Table S15). The performance matrices could be updated with new evidence, such as from the 2019 OARSI[17] or 2020 ACR[16] OA CPGs, and then combined with the preference data to generate the most up to date total scores of existing and new interventions.

7.2.5.2Limitations

This research is not without limitations. Limitations are discussed in Chapters 3 and 5, however two key limitations are discussed here.

7.2.5.2.1 Criteria in the MCDA framework

An important methodological limitation of the aggregation function used to combine preference data with intervention performance data (i.e. the weighted sum model) is that criteria should be structurally independent. That is, the ratings on one criterion should not be constrained or influenced by the ratings on another criterion, and vice-versa (i.e. non-overlapping).[297] In this research the selection of criteria included the Recommendation criterion, which is typically evaluated through an intervention's effectiveness, harms, and quality of evidence.

Secondly, the Quality of Evidence and Cost criteria are not strictly measuring value. The former criterion is a measure of confidence (validity) which can compensated by interventions' performance across the other criteria. The latter may also be described as "a measure of something that is given up to achieve the value criteria" (p.22).[297] Therefore, the conclusions from this research may be imprecise with respect to the criteria selection. However, the selection of criteria was informed by the mixed method study involving local stakeholders (Chapter 4), and the choice-based survey (used to generate the preference weights) was pilot-tested prior to dissemination. The choice-based survey also included a post-survey response verification and process evaluation component (Section 6.3.4.3). These two approaches strengthened the internal validity of the choice-based survey.

Another methodological limitation is that the relative importance of the interventions was never verified by the stakeholders. Confidence in the primary result of this research – prioritised interventions for OA care – could have been strengthened during pilot testing by asking a representative group of stakeholders to directly rank the relative importance of interventions for OA care based on their appraisal of hypothetical patient vignettes and evidence profiles for a subset of interventions. These 'labelled' interventions, ranked in order of importance for each patient vignette, could have then been compared against the relative importance of the same interventions generated indirectly through the framework described in this thesis. This approach has been used previously to develop a system for prioritising patients for elective health services in NZ.[459]

7.2.5.2.2 Sample size

A limitation of the primary research arising from Objective 2 in this thesis, is the modest sample size (N=178). There are several potential consequences. First, the small sample size may have introduced selection bias into the mean preference weights of the sample. The small sample size also limited the ability to conduct meaningful subgroup analysis. The number of Māori participants in the research activity was also small (n=5), severely limiting the generalisability of the study findings to this group. The combination of these factors suggest that the choice-based activity should be replicated in a larger sample to confirm the study results: stakeholders' preference weights for the eight criteria (Table 21), intervention total scores and their rankings (Figure 19), and the result which suggests that preferences between groups are not meaningfully different (Table 22 & Table S21).

7.2.5.2.3 Social deprivation

The findings arising from Objective 2 in this thesis may be limited by the representativeness of the study sample with respect to the social deprivation of the included participants. For example, a finding from this research was that stakeholders did not value intervention cost or accessibility, relative to the other criteria (Table 21). Assuming that the study participants were from areas of less deprivation, participants from areas of greater deprivation may have different preferences, such as preferring the cost and accessibility criteria more, for example due to these being significant barriers to healthcare in their context. This is important for equity and access to healthcare, given that the prevalence of OA is increased in areas of greater deprivation[460-462] and more broadly, that social determinants of health (e.g. housing, transport, education and access to nutritious food options) are known to influence health outcomes. Future research would benefit from including socioeconomic indicators such as decile, income and/or education level to better characterise the representativeness of the study sample and hence the generalisability of the research findings.

7.3 IMPLICATIONS FOR FUTURE RESEARCH

7.3.1 Direct implications

It is clear that the next research step is to translate the decision-making framework and evidence presented in this thesis into practice and policy through implementation of a plan for a national Model of Service Delivery in NZ. This could potentially occur in 5 phases described below, guided by WHO guidelines.[463]

7.3.1.1 Phase I: Empirically test the model for generalisability across the population of New Zealand

This research established a blueprint or framework for prioritising OA interventions in NZ. The limitations section (7.2.5.2) highlighted that the generalisability of the study findings may be limited by the sample size and representativeness of the study sample. It would therefore be prudent to replicate the study findings in a larger and better characterised sample of participants. The first step might include conducting an eDelphi study with an expanded group of participants to establish the acceptability of the Framework of considerations when choosing OA interventions and intervention characteristics established in Objective 1 of the thesis. Included in this activity would be the collection of additional participant sociodemographic characteristics (e.g. disease severity, education level, decile) to assess the acceptability and generalisability of the Framework. Assuming the results are acceptable, the activity carried in Objective 2 to prioritise OA interventions could be conducted again in a larger sample, again with better characterisation of the participants, to explore if preferences for OA interventions meaningfully differ by group (and subgroup). These activities would provide a strong foundation for the subsequent activities described in Phases II-IV.

7.3.1.2 Phase II: Translate the research evidence into recommendations for consumption among policy-makers

This step entails translating the research methods, findings and practical system strengthening implications (i.e. Chapters 4 and 5) into a plain language report for national or jurisdictional policy-makers to consume, based on their informational needs[261, 464-466] (e.g. report on Painful Transitions[467]). For example a report, 'white paper' or evidence brief to support communication, engagement, knowledge translation and evidence-informed decision-making at the levels of the health system which have responsibilities of funding, planning and/or configuring OA health services at a national level; policy-makers (e.g. within the Ministry of Health, PHARMAC, ACC and DHBs). Leadership and engagement is a critical component in implementation for long-term buy-in, commitment and cultural change;[170] ideally, the report should be endorsed by key entities in the health system before proceeding with a national MoSD to utilise macro level system influencers,[171] promote ownership and a nationally consistent response to subsequent change efforts.

7.3.1.3 Phase III (intervention design): Co-design of a national Model of Service Delivery (MoSD)

Collaborate with key stakeholders to co-design an evidence-based and stakeholder-informed national Model of Service Delivery – e.g. OA consumers, healthcare providers, policy-makers and OA experts. The MCDA framework in this thesis could be used to support-value-based service redesign by driving inclusive, transparent, trustworthy and evidence-based decisions to evaluate intervention 'fit' (Figure 22). Discussion about all the possible effects of the proposed MoSD should also take place. This discussion could be guided by the thematic framework of key categories and subcategories describing what influences the choice of interventions in the NZ healthcare system (Figure 15) to help identify issues and redesign any components of the MoSD that may not work in practice (for example, the proposed MoSD may require data system infrastructure that do not exist in the NZ context). Recommendations from these discussions could also be linked to the conceptional frameworks for implementation to help guide interpretation, such as the Consolidated Framework For Implementation Research[170] or Lau et al.'s conceptual framework of factors influencing change (Figure 7).[27]

7.3.1.4Phase IV: Develop an evaluation plan for the national MoSD

Evaluating the systemwide effects of interventions is the objective of the evaluation. The WHO[463] and a framework for evaluating musculoskeletal Models of Care[292] recommends that large-scale evaluation include four components:

- Process evaluation: what processes of change lead to observed effects and why?
- Context evaluation: Are the effects caused by the intervention or due to other factors? Are the effects of the intervention transferrable in different contexts?
- Effects evaluation: What quantity of health gain or quality of service are gained?
- Economic evaluation: How efficient or sustainable is the service relative to status quo?

Decision-makers in this step should determine the key evaluation outcomes of the national MoSD that are meaningful and relevant to the intervention at every level of the healthcare system.[292] For example, at the micro level, quality indicators may be useful to detect changes in practice behaviours; at the meso level, patient reported outcome measures and experience measures may be useful to detect population-level gains; and at the macro-level, costs and effects of intervention may be useful to determine the health system and/or societal costs of the MoSD. These ‘evaluation data’ feed into ‘monitoring and evaluation’ (see Figure 22), therefore it is important that they are carefully selected.

Once the appropriate evaluation outcomes have been determined the next step is to design the evaluation which may follow probability, plausibility and adequacy methodologies. For large-scale, system-level interventions plausibility designs are recommended;[463] they evaluate if a specific intervention, when adequately delivered, is effective in its context.[468]

7.3.1.5 Phase IV (trial, monitor, evaluate and redesign): Implement the MoSD nationally

The last step in the implementation journey is to implement the proposed MoSD and determine its impact according to the proposed evaluation plan, and the pace at which measurable changes are expected. Ideally, this information informs subsequent redesign at every level of the health system to 'calibrate' the MoSD and establish its overall impact on delivering value-based OA care.

7.3.2 Implications beyond transforming care for knee OA

7.3.2.1 Expand the generalisability of the framework

The generalisability of the MCDA framework in this thesis could expanded to explore other non-communicable diseases with similar management principles. For example, CPGs recommendations for low back pain, neck pain and shoulder pain share similar core management principles to hip/knee OA CPGs.[15, 469] Future work could explore the preferences of stakeholders with these conditions by (i) evaluating if the trade-offs in the MCDA framework (i.e. the criteria) are acceptable and complete for each condition (e.g. in an eDelphi survey) and (ii) depending on the outcome, either modify the criteria and/or explore if stakeholders' preferences differ by condition using MCDA. The generalisability of the MCDA framework could also be expanded to low to middle income countries, and countries with different health systems. This could help inform whether stakeholders' preferences for care are influenced by a health system's configuration, and if so, what the consequences are and strategies to manage change (e.g. qualitative focus group discussions, based on the results of the MCDA framework). These expansions could help to inform the design of new models of care that are holistic and patient-centred, and have greater implementation 'fit', recognising that OA does not exist in a vacuum.

7.3.2.1.1 Expand the process for all complex interventions

The background chapter of this thesis established that implementation of complex interventions, particularly in primary care, is influenced by many interdependent factors, involving a range of stakeholders at every level of the health system: context. Therefore, to expand the generalisability of this process to all complex interventions, it would be necessary to repeat the same MCDA process (Table 6) to ensure that the emerging MCDA model reflects contextually important considerations. For example, to structure the overarching decision objective, the type of decision, who is involved, and what criteria stakeholders consider important would be required to build the MCDA model. While potentially resource intensive, the benefits of undertaking this process are arguably high-value and far-reaching. For example, the MCDA process is able to capture and incorporate contextually sensitive preference information from local stakeholders to inform clinical practice (Section 7.2.2), evidence translation (Section 7.2.3), and system strengthening strategies (Section 7.2.4) – the latter which is typically under-served in research and implementation efforts.

7.3.2.2Focus on process-specific implementation factors

This research rests on a hypothesis which is that tailoring or contextualising implementation strategies can help enhance implementation efforts (because a strategy in one context may not necessarily be effective in another). However, there is limited evidence to confirm which tailored implementation approaches are effective.[470] Therefore, future research could focus on investigating other methods for incorporating stakeholder preference and intervention performance data to create tailored models of care for OA, and investigate if tailored interventions are more effective than current practices in a given setting.

For example, the research in this thesis identified that pooled intervention evidence and CPG recommendations about exercise was highly valued by cross-sectoral stakeholders (Figure 19); however, it was unable to discern preferences for program-level characteristics of exercise interventions. A recent study of low back pain consumers' preferences for exercise found that people prefer exercise programmes that differ from those known to be effective, with evidence of preference heterogeneity.[471] Therefore, work arising from this thesis could focus specifically on tailoring program-level characteristics, such as those pertaining to education to self-manage programmes and weight loss, based on what matters to stakeholders. In turn, this might support better clinical practice through 'tipping points' (e.g. promote exercise by educating patients about the risks of drug intervention[287]) and guide better design of models of care implementing behavioural change interventions. This could also address an important evidence-gap in the literature about the preferences of stakeholders for non-drug and non-surgical interventions.

7.4 CONCLUSION

The novel MCDA framework in this thesis was grounded by research evidence from a clinical practice guideline and the values of people with a lived experience of OA, health care providers, Māori health advocates, health policy-makers and OA experts. The mixed-methods research, multiple verification checks, representativeness and MCDA approach give strength to the findings of this research.

Focus groups and an eDelphi were used in the first objective of this thesis to generate rich insight into what matters to cross-sectoral stakeholders' when choosing or recommending OA interventions in the NZ healthcare system. This research adds to knowledge by providing context-sensitive information about the nature of intervention choice in the NZ healthcare system; essential knowledge to initiate transformative change. Stakeholders made choices within a framework of characteristics of interventions, of consumers and of the health system. Nine characteristics of interventions were found to influence intervention choice.

The second objective of this thesis developed an MCDA framework to prioritise OA interventions based on combining the preferences of stakeholders with evidence and recommendations about OA interventions in a clinical practice guideline. The novel research found that cross-sectoral stakeholders preferences for the first-line interventions, self-management education and weight loss represent a barrier to their uptake. No meaningful differences in the preferences of stakeholders were found between groups.

Incorporating what matters to people across the health system in decisions for OA care is a crucial part of enhancing implementation of complex interventions. The MCDA framework in this thesis represents an important step towards making inclusive, systematic and transparent decisions, which can help support the design of patient-centred, high-value healthcare systems for people experiencing OA.

REFERENCES

1. Safiri S, Kolahi A-A, Smith E, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. *Annals of the Rheumatic Diseases*. 2020;79(6):819-828.

2. Ministry of Health. Annual Data Explorer 2018/19: New Zealand Health Survey [Data File]. Accessed 24 Jun, 2020. https://minhealthnz.shinyapps.io/nz-health-survey-2018-19-annual-data-explorer/

3. Access Economics. *The economic cost of arthritis in New Zealand in 2010*. 2010.

4. Wilson R, Abbott JH. The projected burden of knee osteoarthritis in New Zealand: healthcare expenditure and total joint replacement provision. *New Zealand Medical Journal*. Oct 4 2019;132(1503):53-65.

5. MSD's 2019 Household Incomes Report: Background and Selected Key Findings (2019).

6. Ackerman IN, Pratt C, Gorelik A, Liew D. Projected Burden of Osteoarthritis and Rheumatoid Arthritis in Australia: A Population-Level Analysis. *Arthritis Care & Research*. Jun 2018;70(6):877-883. doi:10.1002/acr.23414

7. Abbott JH, Usiskin IM, Wilson R, Hansen P, Losina E. The quality-of-life burden of knee osteoarthritis in New Zealand adults: A model-based evaluation. *PloS one*. 2017;12(10):e0185676.

8. Baldwin J, Briggs AM, Bagg W, Larmer P. An osteoarthritis model of care should be a national priority for New Zealand. *New Zealand Medical Journal*. 2017;30(1467):78-86.

9. 50 Prevalence of disabilities and associated health conditions among adults—United States, 1999 120-5 (2001).

10. Guccione AA, Felson DT, Anderson JJ, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *Am J Public Health*. 1994;84(3):351-358.

11. Schofield DJ, Shrestha RN, Cunich M, et al. Lost productive life years caused by chronic conditions in Australians aged 45-64 years, 2010-2030. *Medical Journal of Australia*. Sep 21 2015;203(6):260.e1-6.

12. Schofield DJ, Callander EJ, Shrestha RN, Percival R, Kelly SJ, Passey ME. Labour force participation and the influence of having arthritis on financial status. *Rheumatology International*. Jul 2015;35(7):1175-81. doi:10.1007/s00296-015-3224-2

13. Royal Australian College of General Practitioners. *Guideline for the management of knee and hip osteoarthritis*. 2018.

14. Larmer PJ, Reay ND, Aubert ER, Kersten P. Systematic Review of Guidelines for the Physical Management of Osteoarthritis. *Archives of Physical Medicine and Rehabilitation*. Feb 2014;95(2):375-389. doi:10.1016/j.apmr.2013.10.011

15. Lin I, Wiles L, Waller R, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. *British Journal of Sports Medicine*. 2020;54(2):79. doi:10.1136/bjsports-2018-099878

16. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. *Arthritis Rheumatololgy*. Feb 2020;72(2):220-233. doi:10.1002/art.41142

17. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthritis and Cartilage*. 2019/11/01/ 2019;27(11):1578-1589. doi:https://doi.org/10.1016/j.joca.2019.06.011

18. Hagen KB, Smedslund G, Østerås N, Jamtvedt G. Quality of community-based osteoarthritis care: a systematic review and meta-analysis. *Arthritis Care & Research*. 2016;68(10):1443-1452.

19. Basedow M, Esterman A. Assessing appropriateness of osteoarthritis care using quality indicators: a systematic review. *Journal of Evaluation in Clinical Practice*. 2015;21(5):782-789. doi:10.1111/jep.12402

20. Leech RD, Eyles J, Batt ME, Hunter DJ. Lower extremity osteoarthritis: optimising musculoskeletal health is a growing global concern: a narrative review. *British Journal of Sports Medicine*. 2019;53(13):806. doi:10.1136/bjsports-2017-098051

21. Pabinger C, Lothaller H, Geissler A. Utilization rates of knee-arthroplasty in OECD countries. *Osteoarthritis and Cartilage*. Oct 2015;23(10):1664-73. doi:10.1016/j.joca.2015.05.008

22. Glasziou P, Straus S, Brownlee S, et al. Evidence for underuse of effective medical services around the world. *The Lancet*. Jul 8 2017;390(10090):169-177. doi:10.1016/s0140-6736(16)30946-1

23. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. *The Lancet*. 2017/07/08/ 2017;390(10090):156-168. doi:https://doi.org/10.1016/S0140-6736(16)32585-5

24. Papandony MC, Chou L, Seneviwickrama M, et al. Patients' perceived health service needs for osteoarthritis (OA) care: a scoping systematic review. *Osteoarthritis and Cartilage*. 2017/07/01/ 2017;25(7):1010-1025. doi:https://doi.org/10.1016/j.joca.2017.02.799

25. Egerton T, Diamond LE, Buchbinder R, Bennell KL, Slade SC. A systematic review and evidence synthesis of qualitative studies to identify primary care clinicians' barriers and enablers to the management of osteoarthritis. *Osteoarthritis and Cartilage*. 2017;25(5):625-638. doi:10.1016/j.joca.2016.12.002

26. Briggs AM, Houlding E, Hinman RS, et al. Health professionals and students encounter multi-level barriers to implementing high-value osteoarthritis care: a multi-national study. *Osteoarthritis and Cartilage*. Jan 19 2019;27(5):788-804. doi:10.1016/j.joca.2018.12.024

27. Lau R, Stevenson F, Ong BN, et al. Achieving change in primary care--causes of the evidence to practice gap: systematic reviews of reviews. *Implementation Science*. Mar 22 2016;11:40. doi:10.1186/s13012-016-0396-4

28. Bach-Mortensen AM, Lange BCL, Montgomery P. Barriers and facilitators to implementing evidence-based interventions among third sector organisations: a systematic review. *Implementation Science*. Jul 30 2018;13(1):103. doi:10.1186/s13012-018-0789-7

29. Speerin R, Needs C, Chua J, et al. Implementing Models of Care for musculoskeletal conditions in health systems to support value-based care. *Best Practice & Research Clinical Rheumatology*. 2020;https://doi.org/10.1016/j.berh.2020.101548doi:https://doi.org/10.1016/j.berh.2020.101548

30. Wahlster P, Goetghebeur M, Kriza C, Niederländer C, Kolominsky-Rabas P, on behalf of the National Leading-Edge Cluster Medical Technologies ‘Medical Valley E. Balancing costs and

benefits at different stages of medical innovation: a systematic review of Multi-criteria decision analysis (MCDA). *BMC health services research*. 2015/07/09 2015;15(1):262. doi:10.1186/s12913-015-0930-0

31. Young CE, Boyle FM, Brooker KS, Mutch AJ. Incorporating patient preferences in the management of multiple long-term conditions: is this a role for clinical practice guidelines? *Journal of comorbidity*. 2015;5:122-131. doi:10.15256/joc.2015.5.53

32. Walsh L, Hill S, Wluka AE, et al. Harnessing and supporting consumer involvement in the development and implementation of Models of Care for musculoskeletal health. *Best Practice & Research Clinical Rheumatology*. 2016/06/01/ 2016;30(3):420-444. doi:https://doi.org/10.1016/j.berh.2016.09.004

33. Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Research Policy and Systems*. 2006;4doi:10.1186/1478-4505-4-12

34. Chong CA, Chen IJ, Naglie G, Krahn MD. How well do guidelines incorporate evidence on patient preferences? *Journal of General Internal Medicine*. Aug 2009;24(8):977-82. doi:10.1007/s11606-009-0987-8

35. Muhlbacher AC, Kaczynski A. Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. *Appl Health Econ Hea*. Feb 2016;14(1):29-40. doi:10.1007/s40258-015-0203-4

36. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. Feb 2012;21(2):145-172. doi:10.1002/hec.1697

37. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Pharmacoeconomics*. Sep 2014;32(9):883-902. doi:10.1007/s40273-014-0170-x

38. Pinto D, Bockenholt U, Lee J, et al. Preferences for physical activity: a conjoint analysis involving people with chronic knee pain. *Osteoarthritis and Cartilage*. 2019;27(2):240-247. doi:https://dx.doi.org/10.1016/j.joca.2018.10.002

39. Moorman CT, 3rd, Kirwan T, Share J, Vannabouathong C. Patient Preferences Regarding Surgical Interventions for Knee Osteoarthritis. *Clinical Medicine Insights: Arthritis and Musculoskeletal Disorders*. 2017;10:1179544117732039. doi:https://dx.doi.org/10.1177/1179544117732039

40. Fraenkel L, Suter L, Cunningham CE, Hawker G. Understanding Preferences for Disease-Modifying Drugs in Osteoarthritis. *Arthritis Care & Research*. 2014;66(8):1186-1192.

41. Chua J, Briggs AM, Hansen P, Chapple C, Abbott JH. Choosing interventions for hip or knee osteoarthritis - What matters to stakeholders? A mixed-methods study. *Osteoarthritis and Cartilage Open*. 09-MAY-2020 2020;2(3):100062. doi:https://doi.org/10.1016/j.ocarto.2020.100062

42. Chua J, Hansen P, Briggs AM, Wilson R, Gwynne-Jones D, Abbott JH. Stakeholders' preferences for osteoarthritis interventions in health services: a cross-sectional study using multi-criteria decision analysis. *Osteoarthritis and Cartilage Open*. 2020/10/10/ 2020:100110. doi:https://doi.org/10.1016/j.ocarto.2020.100110

43. Wilson R, Chua J, Briggs AM, Abbott JH. The cost-effectiveness of recommended adjunctive osteoarthritis management options in New Zealand: results from a computer simulation model. *Osteoarthritis and Cartilage*. 2019;27:S313. doi:10.1016/j.joca.2019.02.711

44. Wilson R, Chua J, Briggs AM, Abbott JH. The cost-effectiveness of recommended adjunctive treatments for knee osteoarthritis: Results from a computer simulation model *Osteoarthritis and Cartilage Open*. 2020;In press:doi: 10.1016/j.ocarto.2020.100123.

45. Chua J, Hansen P, Briggs AM, Abbott H. Establishing the core factors considered by stakeholders choosing or recommending treatment options for hip or knee osteoarthritis in New Zealand (poster presentation). Internal Medicine Journal; 2017:5-41.

46. Chua J, Hansen P, Briggs AM, Abbott JH. What attributes of interventions for osteoarthritis drive preferences? A discrete choice experiment involving cross-sectoral and multi-disciplinary stakeholder groups. *Osteoarthritis and Cartilage*. 2019/04/01/ 2019;27:S302. doi:https://doi.org/10.1016/j.joca.2019.02.698

47. Chua J, Hansen P, Briggs AM, Wilson R, Gwynne-Jones DP, Abbott JH. Integrating values and preferences with the best available evidence: a multi-criteria decision analysis approach. *Osteoarthritis and Cartilage*. 2019/04/01/ 2019;27:S308-S309. doi:https://doi.org/10.1016/j.joca.2019.02.705

48. Ackerman IN, Bucknill A, Page RS, et al. The substantial personal burden experienced by younger people with hip or knee osteoarthritis. *Osteoarthritis and Cartilage*. Aug 2015;23(8):1276-84. doi:10.1016/j.joca.2015.04.008

49. Liu-Bryan R. Inflammation and intracellular metabolism: new targets in OA. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. Nov 2015;23(11):1835-42. doi:10.1016/j.joca.2014.12.016

50. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *Journal of Rheumatology*. Jun 2000;27(6):1513-7.

51. Javaid M, Kiran A, Guermazi A, et al. Individual magnetic resonance imaging and radiographic features of knee osteoarthritis in subjects with unilateral knee pain: the health, aging, and body composition study. *Arthritis & Rheumatism*. 2012;64(10):3246-3255.

52. Guermazi A, Niu J, Hayashi D, et al. Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). *Bmj*. 2012;345:e5339.

53. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *Journal of Rheumatololgy*. Dec 1988;15(12):1833-40.

54. Wilson R, Abbott JH. Age, period and cohort effects on body mass index in New Zealand, 1997–2038. *Australian and New Zealand Journal of Public Health*. 2018;42(4):396-402.

55. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Medical Care*. Sep 2004;42(9):851-9. doi:10.1097/01.mlr.0000135827.18610.0d

56. National Clinical Guideline Centre. National Institute for Health and Clinical Excellence: Guidance. *Osteoarthritis: Care and Management in Adults*. National Institute for Health and Care Excellence (UK); 2014.

57. Skou ST, Koes BW, Grønne DT, Young J, Roos EM. Comparison of three sets of clinical classification criteria for knee osteoarthritis: a cross-sectional study of 13,459 patients treated in primary care. *Osteoarthritis and Cartilage*. Feb 2020;28(2):167-172. doi:10.1016/j.joca.2019.09.003

58. Litwic A, Edwards MH, Dennison EM, Cooper C. Epidemiology and burden of osteoarthritis. *British Medical Bulletin*. March 1, 2013 2013;105(1):185-199. doi:10.1093/bmb/lds038

59. Wang X, Perry TA, Arden N, et al. Occupational Risk in Knee Osteoarthritis: A Systematic Review and Meta-Analysis of Observational Studies. *Arthritis Care & Research*. Sep 2020;72(9):1213-1223. doi:10.1002/acr.24333

60. Canetti EFD, Schram B, Orr RM, Knapik J, Pope R. Risk factors for development of lower limb osteoarthritis in physically demanding occupations: A systematic review and meta-analysis. *Applied Ergonomics*. Jul 2020;86:103097. doi:10.1016/j.apergo.2020.103097

61. Berenbaum F, Wallace IJ, Lieberman DE, Felson DT. Modern-day environmental factors in the pathogenesis of osteoarthritis. *Nature Reviews Rheumatology*. Nov 2018;14(11):674-681. doi:10.1038/s41584-018-0073-x

62. The Organisation for Economic Co-operation and Development. *OECD Health Statistics 2016 - frequently requested data*. 2016.

63. Palazzo C, Nguyen C, Lefevre-Colau M-M, Rannou F, Poiraudeau S. Risk factors and burden of osteoarthritis. *Annals of physical and rehabilitation medicine*. 2016;59(3):134-138.

64. Hunter DJ, Bierma-zeinstra S. Osteoarthritis. *The Lancet*. 2019;393:1749-59.

65. YELIN E. Arthritis. The cumulative impact of a common chronic condition. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*. 1992;35(5):489-497.

66. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis and cartilage*. 2005;13(9):769-781.

67. Glyn-Jones S, Palmer A, Agricola R, et al. Osteoarthritis. *The Lancet*. 2015;386(9991):376-387.

68. Tanamas SK, Wijethilake P, Wluka AE, et al. Sex hormones and structural changes in osteoarthritis: a systematic review. *Maturitas*. 2011;69(2):141-156.

69. Valdes AM, Spector TD. Genetic epidemiology of hip and knee osteoarthritis. *Nature Reviews Rheumatology*. 2011;7(1):23.

70. Spector TD, MacGregor AJ. Risk factors for osteoarthritis: genetics. *Osteoarthritis and cartilage*. 2004;12:39-44.

71. O'Neill TW, McCabe PS, McBeth J. Update on the epidemiology, risk factors and disease outcomes of osteoarthritis. *Best Practice & Research Clinical Rheumatology*. 2018;

72. Jiang L, Tian W, Wang Y, et al. Body mass index and susceptibility to knee osteoarthritis: a systematic review and meta-analysis. *Joint Bone Spine*. 2012;79(3):291-297.

73. Felson DT, Zhang Y, Anthony JM, Naimark A, Anderson JJ. Weight loss reduces the risk for symptomatic knee osteoarthritis in women: the Framingham Study. *Annals of internal medicine*. 1992;116(7):535-539.

74. Cicuttini FM, Wluka AE. Not just loading and age: the dynamics of osteoarthritis, obesity and inflammation. *Medical Journal of Australia*. 2016;204(2):47.

75. Yusuf E, Nelissen RG, Ioan-Facsinay A, et al. Association between weight or body mass index and hand osteoarthritis: a systematic review. *Annals of the rheumatic diseases*. 2010;69(4):761-765.

76. Ohfuji S, Jingushi S, Kondo K, et al. Factors associated with diagnostic stage of hip osteoarthritis due to acetabular dysplasia among Japanese female patients: a cross-sectional study. *BMC musculoskeletal disorders*. 2016;17(1):320.

77. Reyes C, Leyland KM, Peat G, Cooper C, Arden NK, Prieto-Alhambra D. Association Between Overweight and Obesity and Risk of Clinically Diagnosed Knee, Hip, and Hand Osteoarthritis: A Population-Based Cohort Study. *Arthritis Rheumatology*. Aug 2016;68(8):1869-75. doi:10.1002/art.39707

78. Reijman M, Pols HA, Bergink AP, et al. Body mass index associated with onset and progression of osteoarthritis of the knee but not of the hip: the Rotterdam Study. *Annals of the Rheumatic Diseases*. Feb 2007;66(2):158-62. doi:10.1136/ard.2006.053538

79. Gelber AC, Hochberg MC, Mead LA, Wang NY, Wigley FM, Klag MJ. Body mass index in young men and the risk of subsequent knee and hip osteoarthritis. *Am J Med*. Dec 1999;107(6):542-8. doi:10.1016/s0002-9343(99)00292-2

80. Bastick AN, Belo JN, Runhaar J, Bierma-Zeinstra SMA. What Are the Prognostic Factors for Radiographic Progression of Knee Osteoarthritis? A Meta-analysis. *Clinical Orthopaedics and Related Research*. 2015/09/01 2015;473(9):2969-2989. doi:10.1007/s11999-015-4349-z

81. Chapple CM, Nicholson H, Baxter GD, Abbott JH. Patient characteristics that predict progression of knee osteoarthritis: A systematic review of prognostic studies. *Arthritis Care & Research*. 2011/08/01 2011;63(8):1115-1125. doi:10.1002/acr.20492

82. Wright AA, Cook C, Abbott JH. Variables associated with the progression of hip osteoarthritis: A systematic review. *Arthritis Care & Research*. 2009/07/15 2009;61(7):925-936. doi:10.1002/art.24641

83. Sharma L, Song J, Dunlop D, et al. Varus and valgus alignment and incident and progressive knee osteoarthritis. *Annals of the Rheumatic Diseases*. Nov 2010;69(11):1940-5. doi:10.1136/ard.2010.129742

84. Muraki S, Akune T, Oka H, et al. Association of occupational activity with radiographic knee osteoarthritis and lumbar spondylosis in elderly patients of population-based cohorts: a large-scale population-based study. *Arthritis and rheumatism*. Jun 15 2009;61(6):779-86. doi:10.1002/art.24514

85. Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. *Best Practice & Research Clinical Rheumatology*. 2// 2014;28(1):5-15. doi:http://dx.doi.org/10.1016/j.berh.2014.01.004

86. Neogi T, Zhang Y. Epidemiology of Osteoarthritis. *Rheumatic Disease Clinics of North America*. 2// 2013;39(1):1-19. doi:http://dx.doi.org/10.1016/j.rdc.2012.10.004

87. Harris EC, Coggon D. HIP osteoarthritis and work. *Best Practice & Research Clinical Rheumatology*. 2015;29(3):462-482.

88. Ezzat AM, Li LC. Occupational physical loading tasks and knee osteoarthritis: a review of the evidence. *Physiotherapy Canada*. 2014;66(1):91-107.

89. Vigdorchik JM, Nepple JJ, Eftekhary N, Leunig M, Clohisy JC. What is the association of elite sporting activities with the development of hip osteoarthritis? *The American journal of sports medicine*. 2017;45(4):961-964.

90. Driban JB, Hootman JM, Sitler MR, Harris KP, Cattano NM. Is participation in certain sports associated with knee osteoarthritis? A systematic review. *Journal of Athletic Training.* 2017;52(6):497-506.

91. Muthuri S, McWilliams D, Doherty M, Zhang W. History of knee injuries and knee osteoarthritis: a meta-analysis of observational studies. *Osteoarthritis and Cartilage.* 2011;19(11):1286-1293.

92. Slauterbeck JR, Kousa P, Clifton BC, et al. Geographic mapping of meniscus and cartilage lesions associated with anterior cruciate ligament injuries. *The Journal of Bone and Joint Surgery.* 2009;91(9):2094-2103.

93. Øiestad BE, Engebretsen L, Storheim K, Risberg MA. Winner of the 2008 systematic review competition: knee osteoarthritis after anterior cruciate ligament injury. *The American journal of sports medicine.* 2009;37(7):1434-1443.

94. Suter LG, Smith SR, Katz JN, et al. Projecting lifetime risk of symptomatic knee osteoarthritis and total knee replacement in individuals sustaining a complete anterior cruciate ligament tear in early adulthood. *Arthritis care & Research.* 2017;69(2):201-208.

95. Zbrojkiewicz D, Vertullo C, Grayson JE. Increasing rates of anterior cruciate ligament reconstruction in young Australians, 2000–2015. *Medical Journal of Australia.* 2018;208(8):354-358.

96. Thomas AC, Hubbard-Turner T, Wikstrom EA, Palmieri-Smith RM. Epidemiology of posttraumatic osteoarthritis. *Journal of athletic training.* 2017;52(6):491-496.

97. Nevitt MC, Zhang Y, Javaid MK, et al. High systemic bone mineral density increases the risk of incident knee OA and joint space narrowing, but not radiographic progression of existing knee OA: the MOST study. *Annals of the Rheumatic Diseases.* 2010;69(01):163-168. doi:10.1136/ard.2008.099531

98. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. *Arthritis & Rheumatism.* 1986;29(8):1039-1049. doi:10.1002/art.1780290816

99. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis & Rheumatism.* 1986;29(8):1039-1049.

100. Zhang W, Doherty M, Peat G, et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. *Annals of the rheumatic diseases*. 2010;69(3):483-489.

101. McAlindon TE, Bannuru RR, Sullivan MC, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis and Cartilage*. Mar 2014;22(3):363-88. doi:10.1016/j.joca.2014.01.003

102. Sakellariou G, Conaghan PG, Zhang W, et al. EULAR recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis. *Annals of the Rheumatic Diseases*. Sep 2017;76(9):1484-1494. doi:10.1136/annrheumdis-2016-210815

103. Adelani MA, Mall NA, Brophy RH, Halstead ME, Smith MV, Wright RW. The Use of MRI in Evaluating Knee Pain in Patients Aged 40 Years and Older. *The Journal of the American Academy of Orthopaedic Surgeons*. Sep 2016;24(9):653-9. doi:10.5435/jaaos-d-15-00681

104. Croft P, Cooper C, Wickham C, Coggon D. DEFINING OSTEOARTHRITIS OF THE HIP FOR EPIDEMIOLOGIC STUDIES. *American Journal of Epidemiology*. 1990;132(3):514-522. doi:10.1093/oxfordjournals.aje.a115687

105. Tönnis D. *Congenital dysplasia and dislocation of the hip in children and adults*. Springer Science & Business Media; 2012.

106. Hunter DJ, Arden N, Conaghan PG, et al. Definition of osteoarthritis on MRI: results of a Delphi exercise. *Osteoarthritis and Cartilage*. 2011/08/01/ 2011;19(8):963-969. doi:https://doi.org/10.1016/j.joca.2011.04.017

107. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. IHME, University of Washington. Accessed 1 Nov, 2020. http://vizhub.healthdata.org/gbd-compare

108. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*. 2016/10/08/ 2016;388(10053):1545-1602. doi:https://doi.org/10.1016/S0140-6736(16)31678-6

109. Pereira D, Peleteiro B, Araujo J, Branco J, Santos RA, Ramos E. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. *Osteoarthritis and*

cartilage / OARS, Osteoarthritis Research Society. Nov 2011;19(11):1270-85. doi:10.1016/j.joca.2011.08.009

110. Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. *Annals of the Rheumatic Diseases*. July 1, 2014 2014;73(7):1323-1330. doi:10.1136/annrheumdis-2013-204763

111. Hootman JM, Helmick CG, Barbour KE, Theis KA, Boring MA. Updated projected prevalence of self-reported doctor-diagnosed arthritis and arthritis-attributable activity limitation among US adults, 2015–2040. *Arthritis & Rheumatology*. 2016;68(7):1582-1587.

112. Brennan-Olsen SL, Cook S, Leech MT, et al. Prevalence of arthritis according to age, sex and socioeconomic status in six low and middle income countries: analysis of data from the World Health Organization study on global AGEing and adult health (SAGE) Wave 1. *BMC Musculoskelet Disorders*. Jun 21 2017;18(1):271. doi:10.1186/s12891-017-1624-z

113. Wallis JA, Taylor NF, Bunzli S, Shields N. Experience of living with knee osteoarthritis: a systematic review of qualitative studies. *BMJ open*. 2019;9(9):e030060. doi:10.1136/bmjopen-2019-030060

114. Sale JE, Gignac M, Hawker G. The relationship between disease symptoms, life events, coping and treatment, and depression among older adults with osteoarthritis. *The Journal of rheumatology*. 2008;35(2):335-342.

115. Ackerman IN, Bennell KL, Osborne RH. Decline in Health-Related Quality of Life reported by more than half of those waiting for joint replacement surgery: a prospective cohort study. journal article. *BMC Musculoskeletal Disorders*. 2011;12(1):108. doi:10.1186/1471-2474-12-108

116. Berkovic D, Briggs AM, Ayton D, Parker C, Ackerman I. Arthritis-related work outcomes experienced by younger to middle-aged adults: a systematic review. *Occupational and Environmental Medicine*. Aug 26 2020;doi:10.1136/oemed-2020-106640

117. Berkovic D, Ayton D, Briggs AM, Ackerman IN. "The financial impact is depressing and anxiety inducing": A qualitative exploration of the personal financial toll of arthritis. *Arthritis Care & Research*. 2020/02/26 2020; https://doi.org/10.1002/acr.24172doi:10.1002/acr.24172

118. Gelber AC, Hochberg MC, Mead LA, Wang N-Y, Wigley FM, Klag MJ. Joint injury in young adults and risk for subsequent knee and hip osteoarthritis. *Annals of internal medicine.* 2000;133(5):321-328.

119. Deshpande BR, Katz JN, Solomon DH, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. *Arthritis care & research.* 2016;68(12):1743-1750.

120. Sharma A, Kudesia P, Shi Q, Gandhi R. Anxiety and depression in patients with osteoarthritis: impact and management challenges. *Open Aaccess Rheumatology: Research and Reviews.* 2016;8:103.

121. Maurer J, Rebbapragada V, Borson S, et al. Anxiety and depression in COPD: current understanding, unanswered questions, and research needs. *Chest.* 2008;134(4):43S-56S.

122. Januzzi JL, Stern TA, Pasternak RC, DeSanctis RW. The influence of anxiety and depression on outcomes of patients with coronary artery disease. *Archives of Internal Medicine.* 2000;160(13):1913-1921.

123. Calders P, Van Ginckel A. Presence of comorbidities and prognosis of clinical symptoms in knee and/or hip osteoarthritis: a systematic review and meta-analysis. *Seminars in Arthritis and Rheumatism.* 2018;47(6):805-813.

124. Williams A, Kamper SJ, Wiggers JH, et al. Musculoskeletal conditions may increase the risk of chronic disease: a systematic review and meta-analysis of cohort studies. *BMC Medicine.* 2018/09/25 2018;16(1):167. doi:10.1186/s12916-018-1151-2

125. Hall AJ, Stubbs B, Mamas MA, Myint PK, Smith TO. Association between osteoarthritis and cardiovascular disease: Systematic review and meta-analysis. *European Journal of Preventive Cardiology.* Jun 2016;23(9):938-46. doi:10.1177/2047487315610663

126. Wshah A, Guilcher SJ, Goldstein R, Brooks D. Prevalence of osteoarthritis in individuals with COPD: a systematic review. *International Journal of Chronic Obstructive Pulmonary Disease.* 2018;13:1207-1216. doi:10.2147/copd.S158614

127. Eymard F, Parsons C, Edwards MH, et al. Diabetes is a risk factor for knee osteoarthritis progression. *Osteoarthritis and Cartilage.* Jun 2015;23(6):851-9. doi:10.1016/j.joca.2015.01.013

128. Louati K, Vidal C, Berenbaum F, Sellam J. Association between diabetes mellitus and osteoarthritis: systematic literature review and meta-analysis. *Rheumatic and Musculoskeletal Disorders Open*. 2015;1(1):e000077. doi:10.1136/rmdopen-2015-000077

129. Magnusson K, Bech Holte K, Juel NG, et al. Long term type 1 diabetes is associated with hand pain, disability and stiffness but not with structural hand osteoarthritis features - The Dialong hand study. *PLoS One*. 2017;12(5):e0177118. doi:10.1371/journal.pone.0177118

130. Frey N, Hugle T, Jick SS, Meier CR, Spoendlin J. Type II diabetes mellitus and incident osteoarthritis of the hand: a population-based case-control analysis. *Osteoarthritis and Cartilage*. Sep 2016;24(9):1535-40. doi:10.1016/j.joca.2016.04.005

131. Garessus ED, de Mutsert R, Visser AW, Rosendaal FR, Kloppenburg M. No association between impaired glucose metabolism and osteoarthritis. *Osteoarthritis and Cartilage*. Sep 2016;24(9):1541-7. doi:10.1016/j.joca.2016.04.007

132. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis and Cartilage*. 2015/04/01/ 2015;23(4):507-515. doi:https://doi.org/10.1016/j.joca.2014.11.019

133. no. 25 Musculoskeletal conditions and comorbidity in Australia (AIHW) (2019).

134. Marengoni A, Angleman S, Melis R, et al. Aging with multimorbidity: a systematic review of the literature. *Ageing Research Reviews*. 2011;10(4):430-439.

135. Wilmot EG, Edwardson CL, Achana FA, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia*. 2012;55:2895-2905.

136. Kadam UT, Croft PR. Clinical comorbidity in osteoarthritis: associations with physical function in older patients in family practice. *The Journal of rheumatology*. 2007;34(9):1899-1904.

137. Salmon JH, Rat AC, Sellam J, et al. Economic impact of lower-limb osteoarthritis worldwide: a systematic review of cost-of-illness studies. *Osteoarthritis and Cartilage*. 2016/09/01/ 2016;24(9):1500-1508. doi:https://doi.org/10.1016/j.joca.2016.03.012

138. March L, Bachmeier C. Economics of osteoarthritis: a global perspective. *Occupational Health and Industrial Medicine*. 1998;3(38):154.

139. Xie F, Kovic B, Jin X, He X, Wang M, Silvestre C. Economic and humanistic burden of osteoarthritis: a systematic review of large sample studies. *Pharmacoeconomics*. 2016;34(11):1087-1100.

140. van der Zee-Neuen A, Putrik P, Ramiro S, et al. Impact of chronic diseases and multimorbidity on health and health care costs: the additional role of musculoskeletal disorders. *Arthritis care & Research*. 2016;68(12):1823-1831.

141. Agaliotis M, Mackey MG, Jan S, Fransen M. Burden of reduced work productivity among people with chronic knee pain: a systematic review. *Occupational and Environmental Medicine*. 2014;71(9):651-659.

142. Busija L, Buchbinder R, Osborne RH. A grounded patient-centered approach generated the personal and societal burden of osteoarthritis model. *Journal of clinical epidemiology*. 2013;66(9):994-1005.

143. Arthritis and Osteoporosis Victoria. *A problem worth solving*. 2013.

144. Ministry of Health. *Annual update of key results 2014/15: New Zealand Health Survey*. 2015.

145. National Population Projects: 2014(base)-2068 (2014).

146. Briefing to the incoming minister: Health (2014).

147. Ministry of Health. Annual Data Explorer: New Zealand Health Survey [Data File]. Accessed 29 January, 2019. https://minhealthnz.shinyapps.io/nz-health-survey-2019-20-annual-data-explorer/_w_d50ac752/#!/download-data-sets

148. Stanley J, Semper K, Millar E, Sarfati D. Epidemiology of multimorbidity in New Zealand: a cross-sectional study using national-level hospital and pharmaceutical data. *BMJ open*. 2018;8(5):e021689-e021689. doi:10.1136/bmjopen-2018-021689

149. New Zealand Period Life Tables: 2012-14 (2015).

150. Loyola-Sanchez A, Hurd K, Barnabe C. Healthcare utilization for arthritis by indigenous populations of Australia, Canada, New Zealand, and the United States: A systematic review. *Seminars in Arthritis and Rheumatism*. Apr 2017;46(5):665-674. doi:10.1016/j.semarthrit.2016.10.011

151. Hudson ML, Russell K. The Treaty of Waitangi and Research Ethics in Aotearoa. journal article. *Journal of Bioethical Inquiry*. 2009;6(1):61-68. doi:10.1007/s11673-008-9127-0

152. Hooper G, Lee AJ, Rothwell A, Frampton C. Current trends and projections in the utilisation rates of hip and knee replacement in New Zealand from 2001 to 2026. *New Zealand Medical Journal*. Aug 29 2014;127(1401):82-93.

153. Larmer PJ, Bennett K, Baldwin JN, Bassett S, O'Brien DW. Quality indicators for hip and knee osteoarthritis management in New Zealand: A patient survey. *New Zealand Journal of Physiotherapy*. 2019;47(3)

154. Jolly JB, Bassett SFDPBAMP, O'Brien DBM, Parkinson CBSSDP, Larmer PJDMPH. An exploration of the sequence and nature of treatment options available to people living with osteoarthritis of the hip and/or knee within a New Zealand context. *New Zealand Journal of Physiotherapy*. Jul 2017

2018-04-03 2017;45(2):90-95. doi:http://dx.doi.org/10.15619/NZJP/45.2.05

155. Darlow B, Brown M, Thompson B, et al. Living with osteoarthritis is a balancing act: an exploration of patients' beliefs about knee pain. *BMC Rheumatology*. 2018/06/12 2018;2(1):15. doi:10.1186/s41927-018-0023-x

156. Darlow B, Brown M, Grainger R, et al. Stakeholder views about a novel consumer health resource for knee osteoarthritis. *Osteoarthritis and Cartilage Open*. 2020/06/01/ 2020;2(2):100058. doi:https://doi.org/10.1016/j.ocarto.2020.100058

157. Ministry of Health. The mobility action programme. Ministry of Health. Updated 30 May 2019. Accessed 20 November, 2019. https://www.health.govt.nz/our-work/preventative-health-wellness/mobility-action-programme

158. Jones P. MAPping New Zealand Bridging the evidence-practice gap for people with painful musculoskeletal conditions. Health Quality & Safety Commission New Zealand; 2017:

159. Briggs AM, Persaud JG, Deverell ML, et al. Integrated prevention and management of non-communicable diseases, including musculoskeletal health: a systematic policy analysis among OECD countries. *BMJ Global Health*. 2019;4(5):e001806. doi:10.1136/bmjgh-2019-001806

160. National Osteoarthritis Strategy Project Group. *National Osteoarthritis Strategy*. 2018.

161. Osteoarthritis of the Knee Clinical Care Standard (ACSQHC) (2017).

162. Victorial Musculoskeletal Clinical Leadership Group. *Victorian model of care for osteoarthritis of the hip and knee*. 2018.

163. Osteoarthritis Chronic Care Program Model of Care (2012).

164. Elective Joint Replacement Service Model of Care (2010).

165. Allen KD, Choong PF, Davis AM, et al. Osteoarthritis: Models for appropriate care across the disease continuum. *Best Practice & Research Clinical Rheumatology*. 2016;30(3):503-535. doi:10.1016/j.berh.2016.09.003

166. New Zealand Manipulative Physiotherapists Association. GLA:D NZ Certified Practitioners Courses 2020. Accessed 11 Oct, http://www.nzmpa.org.nz/glad-nz-certified-practitioners-courses-2020/

167. Allen-Duck A, Robinson JC, Stewart MW. Healthcare Quality: A Concept Analysis. *Nurs Forum*. 2017;52(4):377-386. doi:10.1111/nuf.12207

168. Lin I, Wiles L, Waller R, et al. Patient-centred care: the cornerstone for high-value musculoskeletal pain management. BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine; 2020.

169. Hoffmann TC, Montori VM, Del Mar C. THe connection between evidence-based medicine and shared decision making. *JAMA*. 2014;312(13):1295-1296. doi:10.1001/jama.2014.10186

170. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*. Aug 7 2009;4doi:Artn 50

10.1186/1748-5908-4-50

171. Briggs AM, Chan M, Slater H. Models of Care for musculoskeletal health: Moving towards meaningful implementation and evaluation across conditions and care settings. *Best Practice & Research Clinical Rheumatology*. 2016;30(3):359-374. doi:10.1016/j.berh.2016.09.009

172. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72.

173. Pope C. Resisting evidence: the study of evidence-based medicine as a contemporary social movement. *Health:*. 2003;7(3):267-282.

174. Satterfield JM, Spring B, Brownson RC, et al. Toward a Transdisciplinary Model of Evidence-Based Practice. *The Milbank Quarterly*. 2009;87(2):368-390.

175. Cochrane AL. *Effectiveness and efficiency: random reflections on health services*. Nuffield Provincial Hospitals Trust; 1972.

176. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews*. 2015;(1)doi:10.1002/14651858.CD004376.pub3

177. Montori VM, Brito JP, Murad MH. The optimal practice of evidence-based medicine: incorporating patient preferences in practice guidelines. *JAMA*. 2013;310(23):2503-2504.

178. van der Weijden T, Legare F, Boivin A, et al. How to integrate individual patient values and preferences in clinical practice guidelines? A research protocol. *Implementation Science*. Feb 2 2010;5doi:Artn 10

10.1186/1748-5908-5-10

179. Preference Collaborative Review Group. Patients' preferences within randomised trials: systematic review and patient level meta-analysis. *BMJ*. 2008;337

180. Haynes RB, Devereaux PJ, Guyatt GH. Physicians' and patients' choices in evidence based practice: Evidence does not make decisions, people do. *BMJ*. 2002;324(7350):1350-1351.

181. Straus SE, Jones G. What has evidence based medicine done for us? *BMJ*. 2004;329(7473):987-988.

182. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ*. Jun 13 2014;348doi:ARTN g3725

10.1136/bmj.g3725

183. Steinberg E, Greenfield S, Mancher M, Wolman DM, Graham R. *Clinical practice guidelines we can trust*. National Academies Press; 2011.

184. Kitson A, Marshall A, Bassett K, Zeitz K. What are the core elements of patient-centred care? A narrative review and synthesis of the literature from health policy, medicine and nursing. *Journal of Advanced Nursing*. 2013;69(1):4-15.

185. Little P, Everitt H, Williamson I, et al. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. *Bmj*. Oct 20 2001;323(7318):908-11.

186. de Wit M, Cooper C, Tugwell P, et al. Practical guidance for engaging patients in health research, treatment guidelines and regulatory processes: results of an expert group meeting organized by the World Health Organization (WHO) and the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO). *Aging clinical and experimental research*. 2019:1-11.

187. Spitaels D, Vankrunkelsven P, Grypdonck L, et al. Quality of Care for Knee Osteoarthritis in Primary Care: A Patient's Perspective. https://doi.org/10.1002/acr.24034. *Arthritis Care & Research*. 2020/10/01 2020;72(10):1358-1366. doi:https://doi.org/10.1002/acr.24034

188. Paskins Z, Sanders T, Hassell AB. Comparison of patient experiences of the osteoarthritis consultation with GP attitudes and beliefs to OA: a narrative review. *BMC family practice*. 2014;15(1):46.

189. Robertsson O, Dunbar M, Pehrsson T, Knutson K, Lidgren L. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. *Acta Orthopaedica Scandinavica*. Jun 2000;71(3):262-7. doi:10.1080/000164700317411852

190. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: A cohort study from the National Joint Registry for England and Wales. *Bone and Joint Journal*. Oct 2013;95-b(10):1359-65. doi:10.1302/0301-620x.95b10.32281

191. Nakano N, Shoman H, Olavarria F, Matsumoto T, Kuroda R, Khanduja V. Why are patients dissatisfied following a total knee replacement? A systematic review. Article. *International orthopaedics*. 2020;44(10):1971-2007. doi:10.1007/s00264-020-04607-9

192. Porter ME. What is value in health care. *New England Journal of Medicine*. 2010;363(26):2477-2481.

193. Shah A. Value-based healthcare: A global assessment. *The Economist Intelligence Unit*. 2016;

194. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. *Health Affairs*. May/Jun

May/Jun 2008

2018-10-06 2008;27(3):759-69. doi:http://dx.doi.org/10.1377/hlthaff.27.3.759

195. Bodenheimer T, Sinsky C. From Triple to Quadruple Aim: Care of the Patient Requires Care of the Provider. *The Annals of Family Medicine*. 2014;12(6):573. doi:10.1370/afm.1713

196. Ishaque S, Karnon J, Chen G, Nair R, Salter AB. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). *Quality of Life Research*. 2019/03/01 2019;28(3):567-592. doi:10.1007/s11136-018-2016-z

197. Rutherford C, Campbell R, Tinsley M, et al. Implementing Patient-Reported Outcome Measures into Clinical Practice Across NSW: Mixed Methods Evaluation of the First Year. *Applied Research in Quality of Life*. 2020:1-20.

198. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology*. 2010;63(11):1179-1194.

199. Rolfson O, Wissig S, van Maasakkers L, et al. Defining an International Standard Set of Outcome Measures for Patients With Hip or Knee Osteoarthritis: Consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. https://doi.org/10.1002/acr.22868. *Arthritis Care & Research*. 2016/11/01 2016;68(11):1631-1639. doi:https://doi.org/10.1002/acr.22868

200. NEJM Catalyst. What is value-based healthcare? Updated January 2017. Accessed Sep 01, 2020. https://catalyst.nejm.org/doi/full/10.1056/CAT.17.0558

201. Deloitte Access Economics. *Osteoarthritis Chronic Care Program evaluation*. 2014.

202. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative. *Seminars in Arthritis and Rheumatism*. 6// 2014;43(6):701-712. doi:http://dx.doi.org/10.1016/j.semarthrit.2013.11.012

203. Zhang W, Moskowitz RW, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis and Cartilage*. 2008/02/01/ 2008;16(2):137-162. doi:https://doi.org/10.1016/j.joca.2007.12.013

204. Charlesworth J, Fitzpatrick J, Perera NKP, Orchard J. Osteoarthritis-a systematic review of long-term safety implications for osteoarthritis of the knee. *BMC musculoskeletal disorders*. 2019;20(1):151.

205. Smith TO, Purdy R, Lister S, Salter C, Fleetcroft R, Conaghan P. Living with osteoarthritis: a systematic review and meta-ethnography. *Scandinavian Journal of Rheumatology*. 2014/11/01 2014;43(6):441-452. doi:10.3109/03009742.2014.894569

206. Feuerstein JD, Pelsis JR, Lloyd S, Cheifetz AS, Stone KR. Systematic analysis of the quality of the scientific evidence and conflicts of interest in osteoarthritis of the hip and knee practice guidelines. *Seminars in Arthritis and Rheumatism*. Feb 2016;45(4):379-385. doi:10.1016/j.semarthrit.2015.09.002

207. Choojaturo S, Sindhu S, Utriyaprasit K, Viwatwongkasem C. Factors associated with access to health services and quality of life in knee osteoarthritis patients: a multilevel cross-sectional study. *BMC health services research*. 2019/10/11 2019;19(1):688. doi:10.1186/s12913-019-4441-2

208. Smith T, Purdy R, Lister S, Salter C, Fleetcroft R, Conaghan PG. Attitudes of people with osteoarthritis towards their conservative management: a systematic review and meta-ethnography. *Rheumatology International*. 2014;34(3):299-313. doi:10.1007/s00296-013-2905-y

209. Dougados M, Betteridge N, Burmester GR, et al. EULAR standardised operating procedures for the elaboration, evaluation, dissemination, and implementation of recommendations endorsed by the EULAR standing committees. *Annals of the rheumatic diseases*. 2004;63(9):1172-1176.

210. American Academy of Orthopaedic Surgeons. *Treatment of the knee*. 2013. *Evidence-based guideline*.

211. Fernandes L, Hagen KB, Bijlsma JWJ, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Annals of the Rheumatic Diseases*. April 17, 2013 2013;doi:10.1136/annrheumdis-2012-202745

212. Bruyère O, Honvo G, Veronese N, et al. An updated algorithm recommendation for the management of knee osteoarthritis from the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO). *Seminars in Arthritis and Rheumatism*. 2019/12/01/ 2019;49(3):337-350. doi:https://doi.org/10.1016/j.semarthrit.2019.04.008

213. Royal Australian College of General Practitioners. *Guideline for the management of knee and hip osteoarthritis*. 2017.

214. Shekelle P, Woolf S, Grimshaw JM, Schünemann HJ, Eccles MP. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. journal article. *Implementation Science*. 2012;7(1):1-7. doi:10.1186/1748-5908-7-62

215. Samanta A, Samanta J, Gunn M. Legal considerations of clinical guidelines: will NICE make a difference? *Journal of the Royal Society of Medicine*. March 1, 2003 2003;96(3):133-138.

216. Grol R. Improving the quality of medical care - Building bridges among professional pride, payer profit, and patient satisfaction. *JAMA*. Nov 28 2001;286(20):2578-2585. doi:DOI 10.1001/jama.286.20.2578

217. Moberg J, Oxman AD, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Research Policy and Systems*. 2018/05/29 2018;16(1):45. doi:10.1186/s12961-018-0320-2

218. National Institute for Health and Care Excellence. *Developing NICE guidelines: the manual*. 2014. *Process and methods*.

219. Social value judgements: principles for the development of NICE guidance (2nd edition) (NICE) (2008).

220. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, et al. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. *Health Research Policy and Systems*. 2020/06/29 2020;18(1):74. doi:10.1186/s12961-020-00588-8

221. Krahn M, Naglie G. The next step in guideline development - Incorporating patient preferences. *JAMA*. Jul 23 2008;300(4):436-438. doi:DOI 10.1001/jama.300.4.436

222. Boivin A, Green J, van der Meulen J, Legare F, Nolte E. Why Consider Patients' Preferences? A Discourse Analysis of Clinical Practice Guideline Developers. *Medical Care*. Aug 2009;47(8):908-915.

223. Fraenkel L, Miller AS, Clayton K, et al. When Patients Write the Guidelines: Patient Panel Recommendations for the Treatment of Rheumatoid Arthritis. *Arthritis Care & Research*. Jan 2016;68(1):26-35. doi:10.1002/acr.22758

224. Crawford MJ, Rutter D, Manley C, et al. Systematic review of involving patients in the planning and development of health care. *BMJ*. 2002;325(7375):1263.

225. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines?: The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*. 1999;281(20):1900-1905.

226. Kung J, Miller RR, Mackowiak PA. Failure of clinical practice guidelines to meet institute of medicine standards: two more decades of little, if any, progress. *Archives of internal medicine*. 2012;172(21):1628-1633.

227. Lenzer J. Why we can't trust clinical guidelines. Article. *BMJ (Online)*. 2013;346(7913)f3830. doi:10.1136/bmj.f3830

228. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *Journal of Clinical Epidemiology*. 2010;63doi:10.1016/j.jclinepi.2010.07.001

229. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines International Network: Toward International Standards for Clinical Practice Guidelines. *Annals of Internal Medicine*. 2012;156(7):525-531. doi:10.7326/0003-4819-156-7-201204030-00009

230. Dziedzic KS, French S, Davis AM, Geelhoed E, Porcheret M. Implementation of musculoskeletal Models of Care in primary care settings: Theory, practice, evaluation and outcomes for musculoskeletal health in high-income economies. *Best Practice & Research Clinical Rheumatology*. 2016/06/01/ 2016;30(3):375-397. doi:https://doi.org/10.1016/j.berh.2016.08.004

231. Briggs AM, Towler SC, Speerin R, March LM. Models of care for musculoskeletal health in Australia: now more than ever to drive evidence into health policy and practice. *Australian Health Review*. Sep 2014;38(4):401-5. doi:10.1071/ah14032

232. MacKay C, Veinot P, Badley EM. Characteristics of evolving models of care for arthritis: A key informant study. *BMC health services research*. 2008/07/14 2008;8(1):147. doi:10.1186/1472-6963-8-147

233. Egerton T, Hinman RS, Hunter DJ, et al. PARTNER: a service delivery model to implement optimal primary care management of people with knee osteoarthritis: description of development. *BMJ open*. Oct 7 2020;10(10):e040423. doi:10.1136/bmjopen-2020-040423

234. Waltz TJ, Powell BJ, Fernández ME, Abadie B, Damschroder LJ. Choosing implementation strategies to address contextual barriers: diversity in recommendations and future directions. *Implementation Science*. Apr 29 2019;14(1):42. doi:10.1186/s13012-019-0892-4

235. World Health Organisation. *Quality of Care: A process for making strategic choices in health systems*. 2006. https://www.who.int/health-topics/quality-of-care#tab=tab_1

236. Hagen KB, Smedslund G, Østerås N, Jamtvedt G. Quality of Community-Based Osteoarthritis Care: A Systematic Review and Meta-Analysis. *Arthritis Care & Research*. 2016;68(10):1443-1452. doi:10.1002/acr.22891

237. Waugh E, King L, Gakhal N, Hawker G, Webster F, White D. Physical activity intervention in primary care and rheumatology for the management of knee osteoarthritis: a review. *Arthritis care & Research*. 2019;71(2):189-197.

238. DeHaan MN, Guzman J, Bayley MT, Bell MJ. Knee osteoarthritis clinical practice guidelines--how are we doing? *The Journal of rheumatology*. 2007;34(10):2099-2105.

239. Mazzuca SA, Brandt KD, Katz BP, et al. Comparison of general internists, family physicians, and rheumatologists managing patients with symptoms of osteoarthritis of the knee. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*. 1997;10(5):289-299.

240. Hinman RS, Nicolson PJA, Dobson FL, Bennell KL. Use of Nondrug, Nonoperative Interventions by Community-Dwelling People With Hip and Knee Osteoarthritis. *Arthritis Care & Research*. 2015;67(2):305-309. doi:10.1002/acr.22395

241. Hunter DJ. Quality of osteoarthritis care for community-dwelling older adults. *Clinics in geriatric medicine*. 2010;26(3):401-417.

242. Brand CA, Harrison C, Tropea J, Hinman RS, Britt H, Bennell K. Management of osteoarthritis in general practice in Australia. *Arthritis Care & Research*. Apr 2014;66(4):551-8. doi:10.1002/acr.22197

243. Basedow M, Williams H, Shanahan EM, Runciman WB, Esterman A. Australian GP management of osteoarthritis following the release of the RACGP guideline for the non-surgical management of hip and knee osteoarthritis. *BMC research notes*. 2015;8:536. doi:10.1186/s13104-015-1531-z

244. Allen KD, Golightly YM, White DK. Gaps in appropriate use of treatment strategies in osteoarthritis. *Best Practice & Research Clinical Rheumatology*. 2017/10/01/ 2017;31(5):746-759. doi:https://doi.org/10.1016/j.berh.2018.05.003

245. Health Funds Association of New Zealand. *Indicative costs of surgery*. 2017.

246. Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. *BMC Musculoskeletal Disorders*. 2019/02/23 2019;20(1):90. doi:10.1186/s12891-019-2411-9

247. Arthritis Australia. *Counting the cost*. 2016. *Part 1: healthcare costs.*

248. Ackerman IN, Skou ST, Roos EM, et al. Implementing a national first-line management program for moderate-severe knee osteoarthritis in Australia: A budget impact analysis focusing on knee replacement avoidance. *Osteoarthritis and Cartilage Open*. 2020/09/01/ 2020;2(3):100070. doi:https://doi.org/10.1016/j.ocarto.2020.100070

249. Hunter DJ, Neogi T, Hochberg MC. Quality of osteoarthritis management and the need for reform in the US. *Arthritis Care & Research*. 2011;63(1):31-38. doi:10.1002/acr.20278

250. Hunter DJ. Lower extremity osteoarthritis management needs a paradigm shift. *British Journal of Sports Medicine*. Apr 2011;45(4):283-8. doi:10.1136/bjsm.2010.081117

251. Kadu MK, Stolee P. Facilitators and barriers of implementing the chronic care model in primary care: a systematic review. *BMC family practice*. 2015/02/06 2015;16(1):12. doi:10.1186/s12875-014-0219-0

252. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. *Frontiers in Pharmacology*. 07/25

01/19/received

06/28/accepted 2013;4:91. doi:10.3389/fphar.2013.00091

253. Erwin J, Chance-Larsen K, Backhouse M, Woolf AD. Exploring what patients with musculoskeletal conditions want from first point-of-contact health practitioners. *Rheumatology Advances in Practice*. 2020;4(1)doi:10.1093/rap/rkz048

254. Basedow M, Hibbert P, Hooper T, Runciman W, Esterman A. Australians with osteoarthritis: satisfaction with health care providers and the perceived helpfulness of treatments and information sources. *Journal of Multidisciplinary Healthcare*. 2016;9:387-94. doi:https://dx.doi.org/10.2147/JMDH.S110751

255. Dobson F, Bennell KL, French SD, et al. Barriers and Facilitators to Exercise Participation in People with Hip and/or Knee Osteoarthritis: Synthesis of the Literature Using Behavior Change Theory. *American journal of physical medicine & rehabilitation*. May 2016;95(5):372-89. doi:10.1097/phm.0000000000000448

256. Wluka AE, Chou L, Briggs AM, Cicuttini FM. *Understanding the needs of consumers with musculoskeletal conditions: Consumers' perceived needs of health information, health services and other non-medical services: A systematic scoping review*. 2016.

257. Wallis JA, Webster KE, Levinger P, Taylor NF. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. *Osteoarthritis and Cartilage*. 2013/11/01/ 2013;21(11):1648-1659. doi:https://doi.org/10.1016/j.joca.2013.08.003

258. Kanavaki AM, Rushton A, Efstathiou N, et al. Barriers and facilitators of physical activity in knee and hip osteoarthritis: a systematic review of qualitative evidence. *BMJ open*. Dec 26 2017;7(12):e017042. doi:10.1136/bmjopen-2017-017042

259. Cottrell E, Roddy E, Rathod T, Porcheret M, Foster NE. What influences general practitioners' use of exercise for patients with chronic knee pain? Results from a national survey. *BMC family practice*. 2016;17(1):172.

260. Briggs AM, Hinman RS, Darlow B, et al. Confidence and attitudes toward osteoarthritis care among the current and emerging health workforce: a multinational interprofessional study. *ACR Cpen Rheumatology*. 2019;1(4):219-235.

261. Tricco AC, Zarin W, Rios P, et al. Engaging policy-makers, heath system managers, and policy analysts in the knowledge synthesis process: a scoping review. *Implementation Science*. 2018/02/12 2018;13(1):31. doi:10.1186/s13012-018-0717-x

262. Tricco AC, Cardoso R, Thomas SM, et al. Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: a scoping review. *Implementation Science*. 2016/01/12 2016;11(1):4. doi:10.1186/s13012-016-0370-1

263. Nilsen P. Making sense of implementation theories, models and frameworks. *Implementation Science*. 2015/04/21 2015;10(1):53. doi:10.1186/s13012-015-0242-0

264. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*. Jun 2008;41(3-4):327-50. doi:10.1007/s10464-008-9165-0

265. Belton V. *Multiple Criteria Decision Analysis: an Integrated Approach*. Springer US; 2002.

266. Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making-An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. Jan 2016;19(1):1-13. doi:10.1016/j.jval.2015.12.003

267. Adunlin G, Diaby V, Xiao H. Application of multicriteria decision analysis in health care: a systematic review and bibliometric analysis. *Health Expect*. Dec 2015;18(6):1894-905. doi:10.1111/hex.12287

268. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the Value of Healthcare Interventions Using Multi-Criteria Decision Analysis: A Review of the Literature. *Pharmacoeconomics*. Apr 2014

2015-01-10 2014;32(4):345-65.

269. Arslan IG, Huls SPI, de Bekker-Grob EW, et al. Patients', healthcare providers', and insurance company employees' preferences for knee and hip osteoarthritis care: a discrete choice experiment. *Osteoarthritis and Cartilage*. 2020/10/01/ 2020;28(10):1316-1324. doi:https://doi.org/10.1016/j.joca.2020.07.002

270. Byun JH, Kwon SH, Lee JE, Cheon JE, Jang EJ, Lee EK. Comparison of benefit-risk preferences of patients and physicians regarding cyclooxygenase-2 inhibitors using discrete choice experiments. *Patient Preference and Adherence*. 2016;10:641-50. doi:10.2147/ppa.S98228

271. Chancellor J, Martin M, Liedgens H, Baker MG, Müller-Schwefe GHH. Stated Preferences of Physicians and Chronic Pain Sufferers in the Use of Classic Strong Opioids. *Value Health*. 2012;15(1):106-117. doi:10.1016/j.jval.2011.07.002

272. Fraenkel L, Wittink DR, Concato J, Fried T. Are preferences for cyclooxygenase-2 inhibitors influenced by the certainty effect? *The Journal of rheumatology*. 2004;31(3):591-3.

273. Fraenkel L, Bogardus ST, Jr., Concato J, Wittink DR. Treatment options in knee osteoarthritis: the patient's perspective. *Archives of internal medicine*. 2004;164(12):1299-304. [Comment in: Arch Intern Med. 2005 Mar 28;165(6):702; PMID: 15795352 [https://www.ncbi.nlm.nih.gov/pubmed/15795352]].

274. Laba T-L, Brien J-a, Fransen M, Jan S. Patient preferences for adherence to treatment for osteoarthritis: the MEdication Decisions in Osteoarthritis Study (MEDOS). *BMC musculoskeletal disorders*. 2013;14:160. doi:https://dx.doi.org/10.1186/1471-2474-14-160

275. Posnett J, Dixit S, Oppenheimer B, Kili S, Mehin N. Patient preference and willingness to pay for knee osteoarthritis treatments. *Patient Preference and Adherence*. 2015;9:733-44. doi:https://dx.doi.org/10.2147/PPA.S84251

276. Hauber AB, Arden NK, Mohamed AF, et al. A discrete-choice experiment of United Kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. *Osteoarthritis and Cartilage*. 2013;21(2):289-97. doi:https://dx.doi.org/10.1016/j.joca.2012.11.007

277. Fraenkel L, Suter L, Cunningham CE, Hawker G. Understanding preferences for disease-modifying drugs in osteoarthritis. *Arthritis Care & Research*. 2014;66(8):1186-92. doi:https://dx.doi.org/10.1002/acr.22280

278. Fraenkel L, Wittink DR, Concato J, Fried T. Informed choice and the widespread use of antiinflammatory drugs. *Arthritis & Rheumatism*. 2004;51(2):210-4.

279. Dennis C, Boeri M, Abraham L, et al. Patient Preferences for Osteoarthritis Pain and Chronic Low Back Pain Treatments in the United States: A Discrete-Choice Experiment. *Osteoarthritis and Cartilage*. 2020;in pressdoi:10.1016/j.joca.2020.06.006

280. Copsey B, Buchanan J, Fitzpatrick R, Lamb SE, Dutton SJ, Cook JA. Duration of Treatment Effect Should Be Considered in the Design and Interpretation of Clinical Trials: Results of a Discrete Choice Experiment. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2019;39(4):461-473. doi:https://dx.doi.org/10.1177/0272989X19841877

281. Berchi C, Degieux P, Halhol H, Danel B, Bennani M, Philippe C. Impact of falling reimbursement rates on physician preferences regarding drug therapy for osteoarthritis using a discrete choice experiment. *The International Journal of Pharmacy Practice*. 2016;24(2):114-22. doi:https://dx.doi.org/10.1111/ijpp.12220

282. Arden NK, Hauber AB, Mohamed AF, et al. How do physicians weigh benefits and risks associated with treatments in patients with osteoarthritis in the United Kingdom? *The Journal of rheumatology*. 2012;39(5):1056-63. doi:https://dx.doi.org/10.3899/jrheum.111066

283. Al-Omari B, Sim J, Croft P, Frisher M. Generating Individual Patient Preferences for the Treatment of Osteoarthritis Using Adaptive Choice-Based Conjoint (ACBC) Analysis. *Rheumatology and therapy*. 2017;4(1):167-182. doi:https://dx.doi.org/10.1007/s40744-017-0056-4

284. Ratcliffe J, Buxton M, McGarry T, Sheldon R, Chancellor J. Patients' preferences for characteristics associated with treatments for osteoarthritis. *Rheumatology*. 2004;43(3):337-45.

285. Hiligsmann M, Dennison E, Beaudart C, et al. A discrete-choice experiment to assess patients' preferences for osteoarthritis treatment: An ESCEO working group. *Seminars in Arthritis and Rheumatism*. 2020/10/01/ 2020;50(5):859-866. doi:https://doi.org/10.1016/j.semarthrit.2020.08.005

286. Franco MR, Howard K, Sherrington C, et al. Eliciting older people's preferences for exercise programs: a best-worst scaling choice experiment. *Journal of Physiotherapy*. 2015/01/01/ 2015;61(1):34-41. doi:https://doi.org/10.1016/j.jphys.2014.11.001

287. Fraenkel L, Fried T. If You Want Patients with Knee Osteoarthritis to Exercise, Tell Them about NSAIDs. *The Patient: Patient-Centered Outcomes Research*. 2008/01/01 2008;1(1):21-26. doi:10.2165/01312067-200801010-00005

288. Byrne MM, Souchek J, Richardson M, Suarez-Almazor M. Racial/ethnic differences in preferences for total knee replacement surgery. *Journal of clinical epidemiology*. 2006;59(10):1078-86. [Comment in: J Clin Epidemiol. 2007 Jun;60(6):593; author reply 594; PMID: 17493515 [https://www.ncbi.nlm.nih.gov/pubmed/17493515]].

289. Harris CA, Shauver MJ, Yuan F, Nasser J, Chung KC. Understanding patient preferences in proximal interphalangeal joint surgery for osteoarthritis: a conjoint analysis. *The Journal of hand surgery*. 2018;43(7):615-624. e4.

290. Marsh K, M IJ, Thokala P, et al. Multiple Criteria Decision Analysis for Health Care Decision Making-Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. Mar 2016;19(2):125-37. doi:10.1016/j.jval.2015.12.016

291. Pinto D, Danilovich MK, Hansen P, et al. Qualitative Development of a Discrete Choice Experiment for Physical Activity Interventions to Improve Knee Osteoarthritis. *Archives of physical medicine and rehabilitation*. 2017;98(6):1210-1216.e1. doi:https://dx.doi.org/10.1016/j.apmr.2016.11.024

292. Briggs AM, Jordan JE, Jennings M, et al. *A framework to evaluate musculoskeletal models of care*. 2016.

293. Briggs AM, Bragge P, Slater H, Chan M, Towler SC. Applying a Health Network approach to translate evidence-informed policy into practice: a review and case study on musculoskeletal health. *BMC health services research*. Nov 14 2012;12:394. doi:10.1186/1472-6963-12-394

294. Belton V, Stewart T. *Multiple criteria decision analysis: an integrated approach*. Springer; 2002.

295. Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care: A bibliometric analysis. *Operations Research for Health Care*. 3// 2013;2(1–2):20-24. doi:http://dx.doi.org/10.1016/j.orhc.2013.03.001

296. Oliveira MD, Mataloto I, Kanavos P. Multi-criteria decision analysis for health technology assessment: addressing methodological challenges to improve the state of the art. *The European Journal of Health Economics*. 2019;20(6):891-918. doi:10.1007/s10198-019-01052-3

297. Hansen P, Devlin N. Multi-Criteria Decision Analysis (MCDA) in Healthcare Decision-Making. Oxford research encyclopedia of economics and finance: Oxford University Press; 2019.

298. Ruwhiu D, Cone M. Advancing a pragmatist epistemology in organisational research. *Qualitative Research in Organizations and Management: An International Journal*. 2010;5(2):108-126.

299. Integrating Qualitative and Quantitative Methods: A Pragmatic Approach. SAGE Publications, Inc.; 2014. https://methods.sagepub.com/book/integrating-qualitative-and-quantitative-methods-a-pragmatic-approach

300. Morgan DL. Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research*. 2007;1(1):48-76. doi:10.1177/2345678906292462

301. Johnson RB, Onwuegbuzie AJ. Mixed methods research: A research paradigm whose time has come. *Educational researcher*. 2004;33(7):14-26.

302. Creswell JW. *Research design : qualitative, quantitative, and mixed methods approaches*. 2nd ed. ed. Sage Publications; 2003.

303. *The SAGE Handbook Qualitative Methods Health Research*. 2010.

304. Sofaer S. Qualitative research methods. *International Journal for Quality in Health Care*. 2002;14(4):329-336. doi:10.1093/intqhc/14.4.329

305. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*. 2007/04/01 2007;1(2):112-133. doi:10.1177/1558689806298224

306. Johnson B, Christensen L. *Educational research: quantitative, qualitative and mixed approaches 5th Edition*. Sage Publication Inc. UK; 2014.

307. Creswell JW. *Designing and conducting mixed methods research*. 2nd ed. SAGE Publications; 2011.

308. Morse JM. *Mixed method design: principles and procedures*. Left Coast Press; 2009.

309. Schoonenboom J, Johnson RB. How to Construct a Mixed Methods Research Design. *Kolner Z Soz Sozpsychol*. 2017;69(Suppl 2):107-131. doi:10.1007/s11577-017-0454-1

310. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008

2019-05-22 2008;26(8):661-677.

311. Coast J, Al-Janabi H, Sutton EJ, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ*. 2012;21(6):730-741. doi:10.1002/hec.1739

312. Louviere JJ. *Stated choice methods: analysis and applications*. Cambridge University Press; 2000.

313. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*. 2015;42(5):533-544.

314. Krueger RA, Casey MA. *Focus groups: A practical guide for applied research*. Sage publications; 2014.

315. Rabiee F. Focus-group interview and data analysis. 2004:655-60.

316. Palinkas L, Horwitz S, Green C, Wisdom J, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*. 2015;42(5):533-544. doi:10.1007/s10488-013-0528-y

317. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *Bmj*. Jul 1 1995;311(6996):42-5. doi:10.1136/bmj.311.6996.42

318. Cleary M, Horsfall J, Hayter M. Data collection and sampling in qualitative research: does size matter? *Journal of Advanced Nursing*. Mar 2014;70(3):473-5. doi:10.1111/jan.12163

319. Burrows D, Kendall S. Focus groups: what are they and how can they be used in nursing and health care research. *Social Sciences in Health*. 1997;3(4):244-253.

320. Harvey N, Holmes CA. Nominal group technique: an effective method for obtaining group consensus. *Int J Nurs Pract*. Apr 2012;18(2):188-94. doi:10.1111/j.1440-172X.2012.02017.x

321. Redman S, Carrick S, Cockburn J, Hirst S. Consulting about priorities for the NHMRC National Breast Cancer Centre: how good is the nominal group technique. *Australian and New Zealand Journal of Public Health*. 1997;21(3):250-256.

322. Hanekom S, Van Aswegen H, Plani N, Patman S. Developing minimum clinical standards for physiotherapy in South African intensive care units: the nominal group technique in action. *Journal of Evaluation in Clinical Practice*. 2015;21(1):118-127. doi:10.1111/jep.12257

323. Hiligsmann M, van Durme C, Geusens P, et al. Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis. *Patient Preference and Adherence*. 2013;7:133. doi:10.2147/PPA.S38408

324. Jaeschke R, Guyatt GH, Dellinger P, et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:a744. doi:10.1136/bmj.a744

325. Potter M, Gordon S, Hamer P. The nominal group technique: a useful consensus methodology in physiotherapy research. *New Zealand Journal of Physiotherapy*. 2004;32(3):126.

326. McMillan SS, Kelly F, Sav A, et al. Using the Nominal Group Technique: how to analyse across multiple groups. journal article. *Health Services and Outcomes Research Methodology*. 2014;14(3):92-108. doi:10.1007/s10742-014-0121-1

327. Van Breda AD. Steps to analysing multiple-group NGT data. *The Social Work Practitioner-Researcher*. 2005;17(1):1-14.

328. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa

329. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*. 2006;27(2):237-246. doi:10.1177/1098214005283748

330. Danermark B. *Explaining Society An Introduction to Critical Realism in the Social Sciences*. 1st ed. ed. Taylor and Francis; 2005.

331. Fletcher AJ. Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology*. 2017;20(2):181-194. doi:10.1080/13645579.2016.1144401

332. Hsieh H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*. 2005/11/01 2005;15(9):1277-1288. doi:10.1177/1049732305276687

333. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*. 12// 2004;42(1):15-29. doi:http://dx.doi.org/10.1016/j.im.2003.11.002

334. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*. 4// 2014;67(4):401-409. doi:http://dx.doi.org/10.1016/j.jclinepi.2013.12.002

335. McMillan S, King M, Tully M. How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*. 2016;38(3):655-662. doi:10.1007/s11096-016-0257-x

336. A. Linstone H, Turoff M. *The Delphi Method: Techniques and Applications*. vol 18. 1975.

337. von der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technological Forecasting and Social Change*. 2012;79(8):1525-1536.

338. Keeney S. *The Delphi technique in nursing and health research*. Wiley-Blackwell; 2011.

339. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *Journal of advanced nursing*. 2000;32(4):1008-1015.

340. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR. *The RAND/UCLA appropriateness method user's manual*. 2001.

341. Powell C. The Delphi technique: myths and realities. *Journal of advanced nursing*. 2003;41(4):376-382.

342. Schmidt RC. Managing Delphi Surveys Using Nonparametric Statistical Techniques*. *Decision Sciences*. 1997;28(3):763-774. doi:10.1111/j.1540-5915.1997.tb01330.x

343. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review. *Palliative Medicine*. 2017;31(8):684-706.

344. Tobin GA, Begley CM. Methodological rigour within a qualitative framework. *Journal of Advanced Nursing*. 2004;48(4):388-396. doi:10.1111/j.1365-2648.2004.03207.x

345. Onwuegbuzie A, Johnson R. The Validity Issue in Mixed Research. *Research in the Schools*. 2006;13(1):48-63.

346. Mixed Methods in Health Sciences Research: A Practical Primer. SAGE Publications, Inc.; 2015. https://methods.sagepub.com/book/mixed-methods-in-health-sciences-research-a-practical-primer

347. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine*. 2010;8(1):18.

348. Lincoln YS. *Naturalistic inquiry*. Sage Publications; 1985.

349. Maria Fenech A, Kiger A. The use of triangulation for completeness purposes. *Nurse Researcher*. 2005

2016-04-09 2005;12(4):19-29.

350. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007-12-01 00:00:00 2007;19(6):349-357. doi:10.1093/intqhc/mzm042

351. Briggs AM, Cross MJ, Hoy DG, et al. Musculoskeletal Health Conditions Represent a Global Threat to Healthy Aging: A Report for the 2015 World Health Organization World Report on Ageing and Health. *The Gerontologist*. 2016;56(Suppl_2):S243-S255. doi:10.1093/geront/gnw002

352. Briggs AM, Jordan JE, Jennings M, et al. Supporting the Evaluation and Implementation of Musculoskeletal Models of Care: A Globally Informed Framework for Judging Readiness and Success. *Arthritis Care & Research*. Apr 2017;69(4):567-577. doi:10.1002/acr.22948

353. Tanios N, Wagner M, Tony M, et al. Which criteria are considered in healthcare decisions? Insights from an international survey of policy and clinical decision makers. *Int J Technol Assess*. Oct 2013;29(4):456-65. doi:10.1017/S0266462313000573

354. Selten EM, Vriezekolk JE, Geenen R, et al. Reasons for treatment choices in knee and hip osteoarthritis: A qualitative study. *Arthritis Care & Research*. Jan 27 2016;68(9):1260-7. doi:10.1002/acr.22841

355. Briggs AM, Jordan JE, Speerin R, et al. Models of care for musculoskeletal health: a cross-sectional qualitative study of Australian stakeholders' perspectives on relevance and standardised

evaluation. journal article. *BMC health services research*. 2015;15(1):1-13. doi:10.1186/s12913-015-1173-9

356. Kitzinger J. Introducing Focus Groups. *BMJ*. 1995;311(7000):299-302.

357. Gray JAM. Evidence based policy making. 10.1136/bmj.329.7473.988. *BMJ*. 2004;329(7473):988.

358. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint Analysis Applications in Health—a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 6// 2011;14(4):403-413. doi:http://dx.doi.org/10.1016/j.jval.2010.11.013

359. GRADE Handbook. Accessed 26 July, 2018. http://gdt.guidelinedevelopment.org/app/handbook/handbook.html

360. Ministry of Health. *The guide to He Korowai Oranga: Māori Health Strategy 2014*. 2014.

361. Ralph LK. Common Mistakes in Making Value Trade-Offs. *Operations Research*. 2002;50(6):935-945. doi:10.1287/opre.50.6.935.357

362. Angelis A, Kanavos P. Value-Based Assessment of New Medical Technologies: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment. *Pharmacoeconomics*. 2016;34(5):435-446. doi:10.1007/s40273-015-0370-z

363. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*. Apr 2011;64(4):401-6. doi:10.1016/j.jclinepi.2010.07.015

364. Cohen JT. A Power Primer. *Psychological Bulletin*. 1992;112:155-159.

365. The New Zealand Joint Registry. *Eighteen Year Report January 1999 to December 2016*. 2017.

366. Ratcliffe J, Buxton M, McGarry T, Sheldon R, Chancellor J. Patients' preferences for characteristics associated with treatments for osteoarthritis. *Rheumatology*. March 1, 2004 2004;43(3):337-345. doi:10.1093/rheumatology/keh038

367. Pinto D, Robertson MC, Abbott JH, Hansen P, Campbell AJ. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee. 2: economic

evaluation alongside a randomized controlled trial. *Osteoarthritis and Cartilage*. 2013/10/01/ 2013;21(10):1504-1513. doi:https://doi.org/10.1016/j.joca.2013.06.014

368. Minstry of Health. Questions and answers – Elective services. Ministry of Health, New Zealand Government. Updated 04 July 2016. Accessed 17 August, 2018. https://www.health.govt.nz/our-work/hospitals-and-specialist-care/elective-services/questions-and-answers-elective-services

369. Hansen P, Ombler F. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis*. 2008;15(3-4):87-107.

370. Kahneman D. *Thinking, fast and slow*. 1st ed. ed. Farrar, Straus and Giroux; 2011.

371. Sullivan T. *Using MCDA (Multi-Criteria Decisions Analysis) to prioritise publicly-funded health care*. University of Otago; 2012.

372. Moshkovich HM, Mechitov AI, Olson DL. Ordinal judgments in multiattribute decision analysis. *European Journal of Operational Research*. 2002/03/16/ 2002;137(3):625-641. doi:https://doi.org/10.1016/S0377-2217(01)00106-0

373. Babashahi S. PMU25 USING MULTIPLE CRITERIA DECISION ANALYSIS (MCDA) TO CREATE A PRIORITY LIST OF CHRONIC NON-COMMUNICABLE DISEASES (NCDS) TO GUIDE HEALTH RESEARCH SPENDING. *Value Health*. 2019;22:S712. doi:10.1016/j.jval.2019.09.1644

374. Sullivan T, Hansen P. Determining Criteria and Weights for Prioritizing Health Technologies Based on the Preferences of the General Population: A New Zealand Pilot Study. *Value Health*. 2017/04/01/ 2017;20(4):679-686. doi:https://doi.org/10.1016/j.jval.2016.12.008

375. Au J, Coleman A, Sullivan T. A Practical Approach to Well-being Based Policy Development: What Do New Zealanders Want from Their Retirement Income Policies? : New Zealand Treasury Wellington; 2015.

376. Weistroffer HR, Li Y. Multiple criteria decision analysis software. *Multiple Criteria Decision Analysis*. Springer; 2016:1301-1341.

377. Moreno-Calderón A, Tong TS, Thokala P. Multi-criteria Decision Analysis Software in Healthcare Priority Setting: A Systematic Review. *Pharmacoeconomics*. 2020:1-15.

378. Johnson TP. *Handbook of health survey methods*. John Wiley & Sons, Inc.; 2015.

379. Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*. 2003;42(1):57-63. doi:10.1046/j.1365-2648.2003.02579.x

380. Choi BCK, Pak AWP. A catalog of biases in questionnaires. *Prev Chronic Dis*. 2005;2(1):A13-A13.

381. Willis G. Cognitive Interviewing : A Tool for Improving Questionnaire Design. Los Angeles, CA, USA: SAGE Publications Inc; 2004.

382. Schwarz N, Oyserman D. Asking questions about behavior: cognition, communication, and questionnaire construction. *The American Journal of Evaluation*. 2001/06/01/ 2001;22(2):127-160. doi:https://doi.org/10.1016/S1098-2140(01)00133-3

383. Beatty PC, Willis GB. Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*. 2007;71(2):287-311. doi:10.1093/poq/nfm006

384. Willis GB. *Cognitive interviewing and questionnaire design: A training manual*. US Department of Health and Human Services, CDCP; 1994.

385. García AA. Cognitive Interviews to Test and Refine Questionnaires. *Public Health Nursing*. 2011;28(5):444-450. doi:10.1111/j.1525-1446.2010.00938.x

386. Wilson BF, Whitehead N, Whitaker K. Cognitive testing proposed questions for PRAMS in the NCHS Questionnaire Design Research Laboratory. 2000:989-993.

387. Faulkner L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*. 2003/08/01 2003;35(3):379-383. doi:10.3758/BF03195514

388. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & Quantity*. 2018/07/01 2018;52(4):1893-1907. doi:10.1007/s11135-017-0574-8

389. Ryan M, Gerard K, Amaya-Amaya M. Discrete Choice Experiments in a Nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, eds. *Using Discrete Choice Experiments to Value Health and Health Care*. Springer Netherlands; 2008:13-46.

390. Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: a systematic review. *BMC health services research*. 2014/09/01 2014;14(1):367. doi:10.1186/1472-6963-14-367

391. Miguel FS, Ryan M, Amaya-Amaya M. 'Irrational' stated preferences: a quantitative and qualitative investigation. *Health Econ*. 2005;14(3):307-322. doi:10.1002/hec.912

392. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ*. 2006;15(8):797-811. doi:10.1002/hec.1104

393. Field AP. *Discovering statistics using SPSS : (and sex and drugs and rock 'n' roll)*. 3rd ed., [Students ed.]. ed. SAGE; 2009.

394. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. *Korean zjournal of Anesthesiology*. 2017;70(2):144-156. doi:10.4097/kjae.2017.70.2.144

395. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*. 2010/12/01 2010;15(5):625-632. doi:10.1007/s10459-010-9222-y

396. Šidák Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*. 1967/06/01 1967;62(318):626-633. doi:10.1080/01621459.1967.10482935

397. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*. 1952;47(260):583-621.

398. Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal*. 2015;15(1):292-300.

399. Wooldridge JM. *Introductory Econometrics*. 5th ed. ed. Cengage Learning US; 2012.

400. McFadden D. The revealed preferences of a government bureaucracy: Empirical evidence. *The Bell Journal of Economics*. 1976:55-72.

401. McFadden D. The revealed preferences of a public bureaucracy. *Department of Economics, University of California, Berkeley*. 1968;

402. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*. 1996;11(6):619-632.

403. Mandara J. The typological approach in child and family psychology: A review of theory, methods, and research. *Clinical Child and Family Psychology Review*. 2003;6(2):129-146.

404. Kaufman L. *Finding groups in data an introduction to cluster analysis*. Wiley-Interscience; 2005.

405. Henry DB, Tolan PH, Gorman-Smith D. Cluster analysis in family psychology research. *Journal of Family Psychology*. 2005;19(1):121.

406. Odilia Y, Kylee TR. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *Tutorials in Quantitative Methods for Psychology*. 2015;11(1):8-21.

407. Pinto D, Robertson MC, Hansen P, Abbott JH. Cost-Effectiveness of Nonpharmacologic, Nonsurgical Interventions for Hip and/or Knee Osteoarthritis: Systematic Review. *Value Health*. 2012/01/01/ 2012;15(1):1-12. doi:https://doi.org/10.1016/j.jval.2011.09.003

408. Welte R, Feenstra T, Jager H, Leidl R. A Decision Chart for Assessing and Improving the Transferability of Economic Evaluation Results Between Countries. *Pharmacoeconomics*. 2004/09/01 2004;22(13):857-876. doi:10.2165/00019053-200422130-00004

409. Deeks JJ, Higgins JP. Statistical algorithms in review manager 5. *Statistical Methods Group of The Cochrane Collaboration*. 2010:1-11.

410. Aronson J, Ferner R. Clarification of Terminology in Drug Safety. *Drug Safety*. 2005;28(10):851-870. doi:10.2165/00002018-200528100-00003

411. Noordzij M, van Diepen M, Caskey FC, Jager KJ. Relative risk versus absolute risk: one cannot be interpreted without the other. *Nephrology Dialysis Transplantation*. 2017;32(suppl_2):ii13-ii18. doi:10.1093/ndt/gfw465

412. Gwynne-Jones D. Quantifying the demand for hip and knee replacement in Otago, New Zealand. *New Zealand Medical Journal*. Jun 28 2013;126(1377):7-17.

413. Shan L, Shan B, Suzuki A, Nouh F, Saxena A. Intermediate and long-term quality of life after total knee replacement: a systematic review and meta-analysis. *The Journal of Bone and Joint Surgery*. Jan 21 2015;97(2):156-68. doi:10.2106/jbjs.m.00372

414. Shan L, Shan B, Graham D, Saxena A. Total hip replacement: a systematic review and meta-analysis on mid-term quality of life. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. Mar 2014;22(3):389-406. doi:10.1016/j.joca.2013.12.006

415. Shan L, Shan B, Graham D, Saxena A. Total hip replacement: a systematic review and meta-analysis on mid-term quality of life. *Osteoarthritis and Cartilage*. 2014/03/01/ 2014;22(3):389-406. doi:https://doi.org/10.1016/j.joca.2013.12.006

416. Stewart TJ. Simplified approaches for multicriteria decision making under uncertainty. *Journal of Multi-Criteria Decision Analysis*. 1995;4(4):246-258. doi:10.1002/mcda.4020040404

417. Vogel JP, Oxman AD, Glenton C, et al. Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: an international survey. *Health Research Policy and Systems*. May 24 2013;11:19. doi:10.1186/1478-4505-11-19

418. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty Analysis: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. *Med Decis Making*. 2012/09/01 2012;32(5):722-732. doi:10.1177/0272989X12458348

419. Østerås N, Jordan KP, Clausen B, et al. Self-reported quality care for knee osteoarthritis: comparisons across Denmark, Norway, Portugal and the UK. 10.1136/rmdopen-2015-000136. *RMD Open*. 2015;1(1)

420. Li LC, Sayre EC, Kopec JA, Esdaile JM, Bar S, Cibere J. Quality of nonpharmacological care in the community for people with knee and hip osteoarthritis. *Journal of Rheumatology*. Oct 2011;38(10):2230-7. doi:10.3899/jrheum.110264

421. Dhawan A, Mather Iii RC, Karas V, et al. An Epidemiologic Analysis of Clinical Practice Guidelines for Non-Arthroplasty Treatment of Osteoarthritis of the Knee. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*. 1// 2014;30(1):65-71. doi:http://dx.doi.org/10.1016/j.arthro.2013.09.002

422. Ferreira de Meneses S, Rannou F, Hunter DJ. Osteoarthritis guidelines: Barriers to implementation and solutions. *Annals of physical and rehabilitation medicine*. Jun 2016;59(3):170-3. doi:10.1016/j.rehab.2016.01.007

423. Boivin A, Currie K, Fervers B, et al. Patient and public involvement in clinical guidelines: international experiences and future perspectives. *Quality and Safety in Health Care*. 2010;19(5):1-4.

424. Lim AY, Doherty M. What of guidelines for osteoarthritis? *International Journal of Rheumatic Diseases*. May 2011;14(2):136-44. doi:10.1111/j.1756-185X.2011.01609.x

425. Woolf S, Schünemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. journal article. *Implementation Science*. 2012;7(1):61. doi:10.1186/1748-5908-7-61

426. Salloum R, Shenkman E, Louviere J, Chambers D. Application of discrete choice experiments to enhance stakeholder engagement as a strategy for advancing implementation: a systematic review. *Implementation Science*. 2017;12doi:10.1186/s13012-017-0675-8

427. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. Oct 20 2007;370(9596):1453-7. doi:10.1016/s0140-6736(07)61602-x

428. Chua J, Briggs AM, Hansen P, Chapple C, Abbott JH. Choosing interventions for hip or knee osteoarthritis - what matters to stakeholders? A mixed-methods study. *Osteoarthritis and Cartilage Open*. 2020;2:100062.

429. Blakely T, Tobias M, Robson B, Ajwani S, Bonné M, Woodward A. Widening ethnic mortality disparities in New Zealand 1981–99. *Soc Sci Med*. 2005/11/01/ 2005;61(10):2233-2251. doi:https://doi.org/10.1016/j.socscimed.2005.02.011

430. Roos EM, Juhl CB. Osteoarthritis 2012 year in review: rehabilitation and outcomes. *Osteoarthritis and Cartilage*. 2012/12/01/ 2012;20(12):1477-1483. doi:https://doi.org/10.1016/j.joca.2012.08.028

431. *FMLOGIT: Stata module fitting a fractional multinomial logit model by quasi maximum likelihood*. Version Revised 17 Feb 2017. 2008.

432. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*. Apr 2011;64(4):383-94. doi:10.1016/j.jclinepi.2010.04.026

433. Dziedzic KS, Allen KD. Challenges and controversies of complex interventions in osteoarthritis management: recognizing inappropriate and discordant care. *Rheumatology*. 2018;57(suppl_4):iv88-iv98.

434. Pellegrini CA, Ledford G, Chang RW, Cameron KA. Understanding barriers and facilitators to healthy eating and physical activity from patients either before and after knee arthroplasty. *Disability and Rehabilitation*. Aug 2018;40(17):2004-2010. doi:10.1080/09638288.2017.1323026

435. Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. journal article. *Health Research Policy and Systems*. November 29 2006;4(1):15. doi:10.1186/1478-4505-4-15

436. Shiffman J, Smith S. Generation of political priority for global health initiatives: a framework and case study of maternal mortality. *The Lancet*. Oct 13 2007;370(9595):1370-9. doi:10.1016/s0140-6736(07)61579-7

437. Gagliardi AR, Brouwers MC, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. journal article. *Implementation Science*. March 22 2011;6(1):26. doi:10.1186/1748-5908-6-26

438. Sikka R, Morath JM, Leape L. The Quadruple Aim: care, health, cost and meaning in work. *BMJ Quality & Safety*. 2015;24(10):608. doi:10.1136/bmjqs-2015-004160

439. Angelis A, Kanavos P. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. *Soc Sci Med*. 2017/09/01/ 2017;188:137-156. doi:https://doi.org/10.1016/j.socscimed.2017.06.024

440. Golan O, Hansen P, Kaplan G, Tal O. Health technology prioritization: which criteria for prioritizing new technologies and what are their relative weights? *Health Policy*. 2011;102(2):126-135.

441. Teo PL, Hinman RS, Egerton T, Dziedzic KS, Bennell KL. Identifying and Prioritizing Clinical Guideline Recommendations Most Relevant to Physical Therapy Practice for Hip and/or Knee Osteoarthritis. *Journal of Orthopaedic & Sports Physical Therapy*. 2019/07/01 2019;49(7):501-512. doi:10.2519/jospt.2019.8676

442. Lavallée JF, Gray TA, Dumville J, Russell W, Cullum N. The effects of care bundles on patient outcomes: a systematic review and meta-analysis. *Implementation Science*. Nov 29 2017;12(1):142. doi:10.1186/s13012-017-0670-0

443. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health*. Jan 2011;38(1):4-23. doi:10.1007/s10488-010-0327-7

444. Powell BJ, Beidas RS, Lewis CC, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. *The Journal of Behavioral Health Services & Research*. Apr 2017;44(2):177-194. doi:10.1007/s11414-015-9475-6

445. Fischer F, Lange K, Klose K, Greiner W, Kraemer A. Barriers and Strategies in Guideline Implementation-A Scoping Review. *Healthcare*. 2016;4(3):36. doi:10.3390/healthcare4030036

446. Luong ML, Cleveland RJ, Nyrop KA, Callahan LF. Social determinants and osteoarthritis outcomes. *Aging Health*. Aug 1 2012;8(4):413-437. doi:10.2217/ahe.12.43

447. Karran EL, Grant AR, Moseley GL. Low back pain and the social determinants of health: a systematic review and narrative synthesis. *Pain*. Nov 2020;161(11):2476-2493. doi:10.1097/j.pain.0000000000001944

448. Daniels N. Accountability for reasonableness : Establishing a fair process for priority setting is easier than agreeing on principles. *BMJ : British Medical Journal*. 2000;321(7272):1300-1301.

449. Gales CL, Moatti J-P. Searching for Consensus Through Multicriteria Decision Analysis: Assessment of Screening Strategies for Hemoglobinopathies in Southeastern France. *Int J Technol Assess*. 1990;6(3):430-449. doi:10.1017/S0266462300001033

450. Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decision analysis. *Journal of Public Health*. 2012;35(3):460-466. doi:10.1093/pubmed/fds099

451. Arden NK, Perry TA, Bannuru RR, et al. Non-surgical management of knee osteoarthritis: comparison of ESCEO and OARSI 2019 guidelines. *Nature Reviews Rheumatology*. 2020:1-8.

452. Bowden JL, Egerton T, Hinman RS, et al. Protocol for the process and feasibility evaluations of a new model of primary care service delivery for managing pain and function in

patients with knee osteoarthritis (PARTNER) using a mixed methods approach. *BMJ open*. Feb 4 2020;10(2):e034526. doi:10.1136/bmjopen-2019-034526

453. Ministry of Health. The Mobility Action Programme. Updated 30 May 2019. Accessed 26 November, 2019. https://www.health.govt.nz/our-work/preventative-health-wellness/mobility-action-programme

454. Brand C, Hunter D, Hinman R, March L, Osborne R, Bennell K. Improving care for people with osteoarthritis of the hip and knee: how has national policy for osteoarthritis been translated into service models in Australia? *International Journal of Rheumatic Diseases*. May 2011;14(2):181-90. doi:10.1111/j.1756-185X.2011.01613.x

455. Speerin R, Slater H, Li L, et al. Moving from evidence to practice: Models of care for the prevention and management of musculoskeletal conditions. *Best Practice & Research Clinical Rheumatology*. Jun 2014;28(3):479-515. doi:10.1016/j.berh.2014.07.001

456. Healthy Ageing Strategy (Ministry of Health) (2016).

457. Barnett RbP. Democratic Governance and Health: Hospitals, Politics and Health Policy in New Zealand. *Australian Health Review*. 2013;37(5):695-695.

458. Holmes E. *Approaches to prioritisation and health technology assessment in New Zealand*. Version 1. ed. 2014.

459. Hansen P, Hendry A, Naden R, Ombler F, Stewart R. A new process for creating points systems for prioritising patients for elective health services. *Clinical Governance: An International Journal*. 2012;17(3):200-209.

460. Wilkie R, Kaur K, Hayward RA. The impact of socioeconomic status on the link between osteoarthritis and the onset of common comorbidities. *Clin Exp Rheumatol*. Jan-Feb 2019;37(1):44-48.

461. Reyes C, Garcia-Gil M, Elorza JM, et al. Socio-economic status and the risk of developing hand, hip or knee osteoarthritis: a region-wide ecological study. *Osteoarthritis and Cartilage*. 2015/08/01/ 2015;23(8):1323-1329. doi:https://doi.org/10.1016/j.joca.2015.03.020

462. Gustafsson K, Kvist J, Eriksson M, Dahlberg LE, Rolfson O. Socioeconomic status of patients in a Swedish national self-management program for osteoarthritis compared with the

general population-a descriptive observational study. *BMC Musculoskelet Disord*. Jan 6 2020;21(1):10. doi:10.1186/s12891-019-3016-z

463. Organization WH. Systems thinking for health systems strengthening. *Systems thinking for health systems strengthening*. 2009:107-107.

464. Oliver K, Lorenc T, Innvær S. New directions in evidence-based policy research: a critical analysis of the literature. journal article. *Health Research Policy and Systems*. 2014;12(1):1-11. doi:10.1186/1478-4505-12-34

465. Oliver KA, de Vocht F. Defining 'evidence' in public health: a survey of policymakers' uses and preferences. *European Journal of Public Health*. 2015;27(suppl_2):112-117. doi:10.1093/eurpub/ckv082

466. Oliver KA, de Vocht F, Money A, Everett M. Identifying public health policymakers' sources of information: comparing survey and network analyses. *European Journal of Public Health*. 2015;27(suppl_2):118-123. doi:10.1093/eurpub/ckv083

467. Slater H, Jordan J, Chua J, Schütze R, Briggs A. Painful transitions: Young people's experiences of living with persistent pain, their interactions with health services and their needs and preferences for pain management including digital technologies. 2016;

468. Victora CG, Schellenberg JA, Huicho L, et al. Context matters: interpreting impact findings in child survival evaluations. *Health policy and planning*. 2005;20(suppl_1):i18-i31. doi:10.1093/heapol/czi050

469. Briggs AM, Slater H, Hsieh E, et al. System strengthening to support value-based care and healthy ageing for people with chronic pain. *Pain*. Jun 2019;160(6):1240-1244. doi:10.1097/j.pain.0000000000001526

470. Baker R, Camosso-Stefinovic J, Gillies C, et al. Tailored interventions to address determinants of practice. *Cochrane Database of Systematic Reviews*. Apr 29 2015;(4):Cd005470. doi:10.1002/14651858.CD005470.pub3

471. Ferreira GE, Howard K, Zadro JR, O'Keeffe M, Lin C-WC, Maher CG. People considering exercise to prevent low back pain recurrence prefer exercise programs that differ from programs known to be effective: a discrete choice experiment. *Journal of Physiotherapy*. 2020/10/01/ 2020;66(4):249-255. doi:https://doi.org/10.1016/j.jphys.2020.09.011

APPENDIX