ECONOMIC EVALUATION OF INTERVENTIONS TO SUPPORT SHARED DECISION-MAKING: AN EXTENSION OF THE VALUATION FRAMEWORK

by

Michael Logan Trenaman

B.Sc., The University of Victoria, 2009

M.Sc., The University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Population and Public Health)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2018

© Michael Logan Trenaman, 2018

The following individuals certify that they have read and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled: Economic evaluation of interventions to support shared decision-making: an extension of the valuation framework submitted by Michael Logan Trenaman in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Population and Public Health.

Examining Committee:

Nick Bansback

Co-supervisor

Stirling Bryan

Co-supervisor

Supervisory Committee Member

Donald Griesdale

University Examiner

Anita Ho

University Examiner

Additional Supervisory Committee Members:

Dawn Stacey

Supervisory Committee Member

Katherine Payne

Supervisory Committee Member

Abstract

Background: Supporting shared decision-making (SDM) between patients and providers is a key health care objective. SDM-interventions can help encourage SDM but may require investment. This thesis used a case study of treatment decision-making for advanced osteoarthritis to quantify the economic value of SDM-interventions in health care.

Methods: A trial-based cost-effectiveness analysis and a longer-term cost analysis using administrative data was undertaken to estimate the value of a SDM-intervention in adults considering total joint arthroplasty. Limitations of conventional cost-effectiveness analysis in assessing the consequences of SDM-interventions were outlined, and methods for valuing the process of SDM presented. A systematic review of discrete choice experiments (DCEs) that have valued the process of SDM was undertaken. A two-step chained valuation technique which included a DCE was completed to estimate the health state utility value of the process of SDM.

Results: The trial-based cost-effectiveness and administrative data analyses suggested that SDM-interventions for total joint arthroplasty provided value, resulting in lower costs at two and seven-years follow-up and similar quality-adjusted life-years (QALYs) over the two-year trial period. QALYs may fail to capture the consequences of SDM-interventions, such as the value of being informed and involved in decision-making. To reflect the opportunity cost of allocating scarce resources toward these non-health benefits, Canadian guidelines suggest that their value be ascertained through the trade-off with health outcomes using societal preferences. The systematic review found 25 studies that have valued SDM using a DCE. No studies valued SDM in advanced osteoarthritis, and most did not include a health outcome attribute or elicit societal preferences. Analysis of the data from the DCE completed by nearly 1,500 Canadians aged 60 and older revealed that respondents were willing to sacrifice health outcomes for greater SDM and estimated the value of SDM.

Conclusions: Evidence suggests that SDM-interventions for adults with advanced osteoarthritis are a cost-effective use of resources. Results from the trial-based cost-effectiveness analysis, systematic review, and DCE suggest that policy-makers may be justified in allocating scarce resources toward SDM-interventions at the expense of other interventions that provide health benefits. Future research is required to quantify the value of SDM-interventions in other contexts.

Lay Summary

Background: Health care systems want to support patients to be active participants in their care. Tools are available to help patients engage in shared decision-making with their doctor. However, these may require health systems to pay money up front. This research aimed to see if these tools provide good value for money in patients with osteoarthritis who were considering total joint replacement.

Methods: This research used data from a trial, a literature review, and survey of Canadians to determine whether tools to support shared decision-making for joint replacement provide value.

Results: Trial results suggest that tools to support shared decision-making may reduce health care costs and improve patient outcomes. Survey results indicate that many (but not all) Canadians value shared decision-making with their doctor regardless of whether it results in improved outcomes.

Conclusions: Investing in tools to support shared decision-making for patients considering joint replacement appears cost-effective in certain circumstances.

Preface

Michael Logan Trenaman was the principal person responsible for identifying and designing this program of research, completing data analysis, interpreting the results, and writing the chapters.

The analysis presented in Chapter Two was based on a randomised controlled trial led by Professor Dawn Stacey at the University of Ottawa. Michael Logan Trenaman created the analysis protocol, performed the analysis, and wrote the chapter. This work was presented at the International Shared Decision Making Conference in Sydney, Australia in 2015, the 37th Annual North American Meeting of the Society for Medical Decision Making in St. Louis in 2015, and the Vancouver Health Economics Meeting in 2016. This chapter was published in the journal *Osteoarthritis and Cartilage* in 2017.

Chapter Three was an extension of the trial-based analysis in Chapter Two. Michael Logan Trenaman created the analysis protocol, performed the analysis, and wrote the chapter. This study was supported by the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. Parts of this material are based on data and/or information compiled and provided by CIHI. However, the analyses, conclusions, opinions, and statements expressed in the material are those of the author(s), and not necessarily those of CIHI. The Institute for Clinical and Evaluative Sciences (ICES) is a prescribed entity under section 45 of Ontario's Personal Health Information Protection Act. Section 45 authorizes ICES to collect personal health information, without consent, for the purpose of analysis or compiling statistical information with respect to the management of, evaluation or monitoring of, the allocation of resources to or planning for all or part of the health system. Projects conducted under section 45, by definition, do not require review by a Research Ethics Board. This project was conducted under section 45, and approved by ICES' Privacy and Compliance Office. Research reported in Chapters Two and Three received ethics approval from the UBC Behavioural Research Ethics Board (H15-01865) and the University of Ottawa (20140661-01H).

The analysis in Chapter Four was completed at the Centre for Health Economics at the University of Manchester, and in Vancouver at the Centre for Health Evaluation and Outcome Sciences (CHEOS) and the Centre for Clinical Epidemiology and Evaluation (C2E2). Michael Logan Trenaman led all aspects of this Chapter. This work was presented at the Health Economist's Study Group Meeting in Gran Canaria, Spain and the Alberta Health Economic Study Group in Calgary, Alberta in 2016.

The analysis in Chapter 5 was completed at the Centre for Health Economics at the University of Manchester and in Vancouver at C2E2 and CHEOS. Michael Logan Trenaman updated an existing database of discrete choice experiments at the University of Manchester and led all aspects of this chapter. This work was presented at the 38th Annual North American Meeting of the Society for Medical Decision Making in Vancouver in 2016.

The analysis in Chapter 6 was completed in Vancouver at CHEOS and C2E2. Michael Logan Trenaman developed the research objectives, created the discrete choice experiment survey, and performed data analysis, interpretation, and wrote the chapter. In 2017, this work was presented at the Vancouver Health Economics Meeting, the Health Economist Study Group Meeting in Aberdeen, Scotland, the International Shared Decision Making Conference in Lyon, France, and the 39th Annual North American Meeting of the Society for Medical Decision Making in Pittsburgh in 2017. Research reported in Chapter Six received ethics approval from the UBC Behavioural Research Ethics Board (H16-03355).

Table of contents

List of Tables

List of Figures

Glossary

Attributes	The 'properties' or 'characteristics' of the goods described in a discrete choice experiment.
Choice set	The alternatives presented to the respondent from which they choose in a discrete choice experiment.
CollaboRATE	A patient-reported measure of the shared decision-making process which includes three questions related to: 1) explanation of the health issue; 2) elicitation of patient preferences; and 3) integration of patient preferences.
Conditional logit model	A regression model with a categorical dependent variable where the values of the variables (usually choice characteristics) vary across the choices but parameters are common across the choices. Also know as the 'multinomial logit' model.
Conjoint analysis	A stated preference technique where respondents are asked to order or score alternatives according to their preferences.
Contingent valuation	A stated preference technique used to elicit willingness-to pay (or accept) values through a direct question.
Cost-benefit analysis	An assessment of the costs and the benefits of an intervention where the benefits are measured in monetary terms.
Cost-effectiveness analysis	An assessment of the costs and the benefits of an intervention where the benefits are measured in a clinical or health-related metric.
Cost-minimization analysis	An assessment of the cost of a health intervention where the benefits are assumed to be identical.
Cost-utility analysis	An assessment of the costs and the benefits of an intervention where the benefits are measured in a quality-adjusted outcome, such as quality-adjusted life years.
Donabedian model	A conceptual framework for assessing the quality of health care which includes three concepts: structures, processes, and outcomes.
Experimental design	A sample from all possible combinations of attribute levels used to construct choice alternatives in a discrete choice experiment.
Extra-welfarist	An evaluative framework where something other than, or in addition to, utility is maximised (for example, health).
Fractional factorial design	A sample from a full factorial design which can estimate effects of interest through interactions in a discrete choice experiment.

Full factorial design	A design using the complete set of all attribute and level combinations in the discrete choice experiment.
Health state utility value	A quality-of-life weight that is measured on a scale anchored at one (equivalent to full health) and zero (equivalent to dead), with some negative values (reflecting states worse than dead). It is also called a "health utility." Notably, true 'utilities' must obey the axioms of von Neumann-Morgenstern utility theory for decisions under uncertainty. The only valuation method that conforms to these axioms is the standard gamble.
Heteroscedastic conditional logit model	A conditional logit regression model accounting for differences in the variance of the error term through estimation of the scale parameter.
Lancaster's theory	A hypothesis which suggests that individuals do not value a good or service *per se*, but instead value its characteristics or attributes.
Latent-class analysis	A regression modelling technique which identifies subsets of respondents with similar preferences.
Lexicographic preferences	When the good providing the most of X is always preferred, no matter what the amount of Y.
Marginal rate of substitution	The willingness to exchange a unit of one good for another to maintain the same level of utility.
Mixed logit model	A regression model which allows for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors utilising any distribution for the random coefficients.
Ngene	A software package used to efficiently design and test the statistical properties of discrete choice experiments.
Opportunity cost	The cost of an alternative that is forgone when making a choice within a fixed budget.
Ordering effect	A phenomenon which occurs when changing the arrangement of questions in a survey affects responses.
Osteoarthritis	A heterogeneous group of conditions that are associated with defective integrity of articular cartilage and changes in the underlying bone at the joint margins which results in joint pain, aching, and stiffness.
Outcomes	Defined by Donabedian as the changes (whether desirable or undesirable) in individuals and populations attributable to health care.
Patient-centred care	An approach to health care that supports people to develop the knowledge, skills, and confidence they need to effectively manage and make informed decisions about their own health and healthcare.

Patient decision aid	Evidence-based tool designed to help patients make specific and deliberated choices among healthcare options.
Processes	Defined by Donabedian as the activities that constitute health care.
Random utility theory	A choice theory where decisions are deterministic and utility has a random component.
Revealed preference	Data collected through observations of behaviour in real markets.
Shared decision-making	The conversation that happens between a patient and their health care professional to reach a health care choice together.
Shared decision-making intervention	An intervention designed to support shared decision-making between patients and providers. Examples of provider-focused interventions include distribution of printed materials, educational meetings, audit and feedback, educational outreach, and skills training. Patient focused interventions include patient decision aids.
Standard gamble	A stated preference technique used to estimate health state utility values through the trade-off with the risk of an undesirable outcome (e.g., death).
Stated preference	Data collected through surveying individuals to attain information about how they would behave in a hypothetical scenario.
Structures	Defined by Donabedian as the conditions under which care is provided.
Time trade-off	A stated preference technique used to estimate health state utility values through the trade-off with life-years.
Total joint arthroplasty	A procedure where parts of the damaged joint are removed and replaced with prosthesis, which can be metal, plastic, or ceramic. Includes both total hip arthroplasty and total knee arthroplasty.
Utility	Term used in economics to describe the satisfaction gained from the consumption of goods or services.
Welfarist	An evaluative framework that maximises total utility for society.

List of abbreviations

BWS	Best-worst scaling
CA	Conjoint analysis
CADTH	Canadian Agency for Drugs and Technologies in Health
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CI	Confidence interval
CMA	Cost-minimization analysis
CMS	Centers for Medicare and Medicaid Services
CSHS	Cost of a standard hospital stay
CV	Contingent valuation
DAD	Discharge abstract database
DCE	Discrete choice experiment
EQ-5D	EuroQoL 5-dimension
GP	General practitioner
HRQoL	Health-related quality-of-life
HTA	Health technology assessment
HTERP	Health technology expert review panel
ICER	Incremental cost-effectiveness ratio
MAR	Missing at random
MRS	Marginal rate of substitution
MCDA	Multi-criteria decision analysis
NACRS	National Ambulatory Care Reporting System
NICE	National Institute for Health and Care Excellence
NRS	National Rehabilitation Reporting System
OA	Osteoarthritis
ODB	Ontario Drug Benefit
OHIP	Ontario Health Insurance Plan
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RIW	Resource intensity weight
RPDB	Registered Persons Database

RUT	Random Utility Theory
SD	Standard deviation
SDM	Shared decision-making
SG	Standard gamble
THA	Total hip arthroplasty
TKA	Total knee arthroplasty
TJA	Total joint arthroplasty
TTO	Time trade-off
UK	United Kingdom
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
WTP	Willingness-to-pay
WTW	Willingness-to-wait

Acknowledgements

I would like to begin by thanking my supervisory committee, including Nick Bansback, Stirling Bryan, Dawn Stacey, and Katherine Payne. I have been incredibly fortunate to have your support and guidance throughout the process of obtaining my doctorate.

I would like to thank the Canadian Institutes for Health Research for supporting this dissertation through a Banting-Best Doctoral Award, and the University of British Columbia for supporting this research through a four-year fellowship.

I would also like to thank the University of British Columbia Faculty of Medicine Friedman Scholar Program for providing financial support that enabled me to study at the Manchester Centre for Health Economics (MCHE) during my PhD. This was an invaluable experience. Thanks to Caroline and Alex for opening their house to a stranger, and to the MCHE for providing an intellectually stimulating environment at work, and a culturally stimulating one away from it.

I would like to thank my family: Mom, Dad, and Lundy, for their unwavering support.

I would like to thank Jack and Finn for their unconditional love, and for not continuing the family trend of deleting one chapter per thesis.

Lastly, I would like to thank my partner, Sarah, for inspiring, encouraging, and supporting me throughout this journey.

1 Introduction

1.1 Background

Patient-centred care is an approach that "supports people to develop the knowledge, skills, and confidence they need to effectively manage and make informed decisions about their own health and healthcare."[1] The Institute of Medicine in the United States defines patient-centred care as health care that is "respectful of and responsive to individual patient preferences, needs, and values, and ensur[es] that patient values guide all clinical decisions."[2] These definitions focus on the patient-provider relationship, and represent a departure from the traditional, paternalistic model of health care where decisions were made by the health care provider with little input from the patient.[3]

Shared decision-making (SDM) is a component of patient-centred care that supports informed decision-making,[4] and has been defined as "the conversation that happens between a patient and their healthcare professional to reach a healthcare choice together."[5] While the specific definition and behaviours of SDM vary, Makoul et al. outline nine essential elements, including: defining/explaining the problem, presenting options, discussing pros/cons, considering patient values/preferences, discussing patient ability/self-efficacy, incorporating the doctor's knowledge and recommendations, checking/clarifying understanding, making or explicitly deferring the decision, and arranging follow-up.[6]

Despite a supportive policy environment, SDM is not widely implemented in clinical practice.[7,8] Research has estimated that SDM only occurs in 10% of consultations.[9] To accelerate implementation, there has been widespread development of interventions to support SDM (SDM-interventions). Broadly speaking, SDM-interventions may target patients or health care professionals, individually, or together.[10] SDM-interventions for health care professionals include distribution of printed materials, educational meetings, audit and feedback, educational outreach, and skills training.[11,12] Patient focused SDM-interventions include patient decision aids, which may be provided to patients before or during the consultation, and provide information on the diagnosis, health condition, and treatment options, while also helping patients clarify their preferences.[13] SDM-interventions may also be designed to target both patients and health care professionals. A Cochrane review of SDM-interventions found that "interventions targeting patients and healthcare professionals together show more promise than those targeting only one or the other."[10] Patient decision aids are the SDM-interventions most supported by evidence and most widely used. To date there have been over 500 decision aids developed, and over 100 randomized controlled trials (RCTs) evaluating their effectiveness.[13] Evidence from the Cochrane review of patient decision aids, published in 2017, suggested that SDM-interventions may improve patient knowledge and patient-provider communication, reduce decisional conflict, and result in patients choosing

treatments that are more congruent with their values.[13] While SDM-interventions may provide benefit, they may also impact health care costs.

A highly-cited 2013 editorial in the *New England Journal of Medicine* argued that SDM may result in lower health care costs and improve patient outcomes.[15] However two systematic reviews found that the evidence to support this claim is lacking.[16,17] Patient targeted interventions, such as patient decision aids, often require printed materials or DVDs, or incur costs for internet hosting or periodic updates.[16] SDM-interventions that target health care providers may require additional clinician or administrative staff time for training in SDM skills.[18,19] In many cases, additional staff time may be required to identify eligible patients or disseminate materials. SDM-interventions may also involve new consultations with other health care staff or increase the length of consultation with the health care professional. The 2017 Cochrane review of patient decision aids found a median 2.6-minute increase in consultation length when a decision aid was used.[13]

The perception that SDM requires additional time has resulted in organizations exploring dedicated SDM billing codes as a means of encouraging SDM. For example, the Centers for Medicare and Medicaid Services (CMS) in the United States aimed to begin a SDM pilot program in 2018 which would pay health care professionals $50 for each instance of providing a patient decision aid and having a dedicated SDM consultation. While the potential upfront investment for these programs is small at the individual level,

they target highly prevalent conditions, meaning that the overall investment is significant.

Since SDM-interventions may require investments, and all health systems operate in an environment of resource scarcity, it is critical to investigate whether SDM-interventions provide added value. This dissertation aims to quantify the economic value of interventions to support SDM in health care. Chapter One defines and describes the core foundational concepts underpinning this dissertation (Section 1.2), including introducing SDM and SDM-interventions, and outlining a case study: SDM in the context of treatment decision-making for patients with advanced osteoarthritis (OA). The definition and role of economic evaluation, and established methods to value aspects of health and healthcare are then described. Chapter One concludes by outlining the objectives of the individual studies that comprise this dissertation (Section 1.3) and provide an overview of the topics and methods used in each subsequent chapter (Section 1.4).

1.2 Foundational concepts

The aim of this section is to define and describe foundational concepts that underpin this dissertation. It begins by describing the emergence of SDM as a stated objective in international, national, and regional health care organizations. The case-study used throughout this dissertation is then described: treatment decision-making for advanced OA. This section provides an overview of the disease, available treatment

options, outlines how SDM may play an important role in improving care, describes why this area has become a key policy priority in multiple jurisdictions, and reviews past trials and studies of SDM-interventions in this context. Given the potential resource implications of providing SDM-interventions in this context, the role health economics plays in resource allocation decisions, and the specific economic evaluation and valuation methods available to inform these decisions, are described. This section concludes by reviewing past economic evaluation of SDM-interventions and highlights the key challenges that will be addressed in this dissertation.

1.2.1 Shared decision-making

Shared decision-making (SDM) is widely supported by health policy globally, as evidenced by its emergence as a key priority in policy documents. The World Health Organization quality framework states that health care systems should ensure autonomy for individuals to make choices about their own health.[20] SDM features prominently in the 2001 report entitled, "Crossing the Quality Chasm" from the United States Institute of Medicine. This report outlined "10 rules for redesign," which recommended that care should be "customized according to patients needs and values," and "patients should be given the necessary information and opportunity to exercise the degree of control they choose over health care decisions that affect them."[2] Toward this aim, the 2010 United

States Affordable Care Act (ACA) included financial provisions to encourage uptake of SDM in routine clinical practice.[15]

In the United Kingdom (UK), SDM is embedded in the National Health Service (NHS) Constitution and Mandate,[21] and is represented in the NHS quality standards for patient experience, which include "giving patients opportunities to discuss their health beliefs and preferences," "supporting patient choice," and "tailoring health services to the individual."[22] In Canada, the British Columbia Ministry of Health uses the terminology "patient-centered care," which is stated as the number one priority for the health system, with the aim of "empowering staff working with patients and residents to individualize the experience of care."[23] In defining patient-centered care, the British Columbia framework specifically cites "shared and informed decision-making," "improved information and understanding," and "an enhanced experience of health care."[23] It also highlights the focus on patients, families, and their caregivers, who should be supported and encouraged to participate in: their own care, decision-making about that care, and choosing their level of participation in decision-making.[23]

A growing body of literature has identified barriers and facilitators to the uptake of SDM by patients and providers.[7,24] The most commonly cited barrier from the provider's perspective is the belief that engaging in SDM will take additional time.[7] Additional barriers from the providers' perspective include a belief that SDM is not relevant for the specific clinical context, or the individual patient.[7] Providers may feel that

patients lack the ability to make an informed choice, though evidence suggests that patients have the potential to benefit regardless of their age or educational background.[8]

One clinical area that has witnessed rapid development of SDM-interventions is for OA of the hip and knee, where options include a range of surgical and non-surgical treatments. The treatment decision is preference-sensitive, requiring patients to consider the balance (or trade-off) between benefits and harms. This dissertation uses the case study of treatment decision-making for patients with advanced OA. This is a key policy priority where SDM, and SDM-interventions such as patient decision aids, may play an important role in ensuring that treatment decisions reflect what matters most to patients.[25]

1.2.2 A case-study: Shared decision-making in advanced osteoarthritis

1.2.2.1 Epidemiology of osteoarthritis

Osteoarthritis (OA) is a progressive, chronic, condition. The American College of Rheumatology defines OA as: "A heterogeneous group of conditions that lead to joint symptoms and signs which are associated with defective integrity of articular cartilage, in addition to related chances in the underlying bone at the joint margins."[26] The estimated prevalence of OA varies depending upon the definition used, and the joints investigated.[27] OA can be defined pathologically, radiographically, or clinically,[27] and

may affect large joints such as the hip or knee, or small joints of the hands or feet.[28] In many cases OA affects multiple joints in the same individual.

Trends suggest that the prevalence of OA is increasing, based on population aging and increasing rates of obesity.[29] In the United States, the prevalence of OA in adults aged 25 years and older increased from 21 million in 1995 to 27 million over a ten-year period.[29] In British Columbia, Canada, a 2007 population-based study using administrative data from 1991 to 2002 estimated the prevalence of OA at 11%,[30] while a separate study estimated the prevalence of diagnosed OA at 14% in Canadians over 30 years of age.[31] Of all Canadians aged 20 and older with OA, approximately 17% report having exclusively hip OA, 28% report having exclusively knee OA, and 29% report both.[32] As a result, hip and knee OA account for approximately three out of every four cases of OA in Canada. The etiology of OA is multifactorial, including a host of systemic and local risk factors. Systemic risk factors for OA include age, gender and hormones, race/ethnicity, genetics, and diet/nutritional factors.[27] Local risk factors for OA include obesity, injury, occupation, physical activity/sports, and mechanical factors.[27]

1.2.2.2 Treatment options

Given the progressive nature of OA, and the lack of therapies available to prevent or reverse disease progression, treatment is directed at symptom relief, such as reducing joint pain, aching, and stiffness. [32] There are a range of treatment options available to help

manage these symptoms. Guidelines for the non-surgical management of knee OA from the OA Research Society International indicate that core treatments for all patients include land-based exercise, weight management, strength training, water-based exercise, and self-management and education.[33] They also recommend different management strategies depending on whether an individual has knee-only OA or multi-joint OA, and co-morbid health concerns such as diabetes, advanced age, or depression. These include non-pharmacologic (e.g., biomechanical interventions, walking canes) and pharmacologic (e.g., acetaminophen, oral and topical nonsteroidal anti-inflammatory drugs, tramadol, intra-articular corticosteroids) treatments.[33] The same complement of treatments are recommended by the American College of Rheumatologists for individuals with both hip and knee OA.[34] Evidence suggests that 74% of Canadians manage hip and knee OA with non-prescription medications, while 52% use prescription medications.[32]

For individuals with advanced OA, for whom the pharmacological and non-pharmacological treatments listed above have either failed or become ineffective at managing symptoms, total joint arthroplasty (TJA) is a recommended treatment. TJA is a procedure where parts of the damaged joint are removed and replaced with prosthesis, which can be metal, plastic, or ceramic.[35–37] The aim is to replicate the movement of a healthy joint and improve the functioning and quality-of-life of patients. TJA has been shown to be highly effective at restoring joint function, reducing pain, and improving

overall quality of life, and highly cost-effective.[38–41] As a result, total hip arthroplasty (THA) and total knee arthroplasty (TKA) are among the most common elective surgical procedures. Data from the Canadian Joint Replacement Registry estimates that there were over 53,000 THAs and 64,000 TKAs performed in Canada in 2015/16, representing an 18% and 16% 5-year increase, respectively.[42] The increase in rates of TJA may be attributed to two trends: 1) the increasing prevalence of OA due to a rise in risk factors such as obesity and an aging population, and 2) an increase in the rates of TJA in younger patients with milder disease which reflects a shift from using TJA to "manage disability" to proactively "prevent disability."[43]

1.2.2.3 Role of SDM in treatment decision-making

SDM may help overcome two issues in patients deciding whether to undergo TJA. These include the propensity for patients to: 1) have unrealistic expectations, including overestimating the potential benefits of surgical treatment and underestimating the harms, and 2) fail to have an adequate trial of non-surgical treatment options. This is reflected in six criteria developed jointly by surgeons and patients in Canada which aim to identify when TJA is appropriate.[43]

For example, the one criterion states that patients are appropriate for surgery if both "the patient and surgeon agree that the potential benefits to the patient of joint replacement surgery outweigh the potential surgical risks." Having an informed

discussion about the pros and cons of treatment while considering the values and preferences of the patient is the crux of SDM.[4] While TJA is both highly effective and cost-effective for the 'average' patient, there is evidence that as many as 15-30% of patients do not improve and/or report dissatisfaction with the results.[44–46] A 2013 population-based cohort study found that only half of the included patients (n=202) achieve a good outcome, defined as meeting a minimally important improvement in pain and disability.[47] Furthermore undergoing TJA requires accepting an increased risk of complications, including thromboembolism (1-2%), infection (0.2-2.5%), periprosthetic fracture (0.5-10%), myocardial infarction (0.2%), congestive heart failure (0.6%), neurovascular injury (0.1-2.0%), dislocation (0.3-10%), and mortality (0.06-0.16%).[48] An additional consideration for patients is the potential for revision surgeries, which is estimated at 5% and 12% after five and ten years, respectively.[49] When compared with primary surgeries, revisions tend to have a higher rate of complications and patients are less likely to benefit.[50] SDM offers a mechanism by which patients and providers can discuss this trade-off between greater effectiveness but increased risk, to ensure that treatment decisions reflect patients' values and preferences. A second criterion states that patients are appropriate for surgery if their "expectations for total joint replacement surgery are achievable."[43] The most significant predictor of satisfaction post-TJA is expectations being met.[45] Evidence suggests that patients often have unrealistic expectations for medical treatments, including elective surgery, tending to overestimate

potential benefits and underestimate harms.[51] SDM provides an opportunity to align expectations with the best clinical evidence.

SDM may support patients and their providers in choosing whether to continue or intensify non-surgical treatment. For example, one of the Canadian criteria states that patients are appropriate if they have "had an adequate trial of nonsurgical arthritis treatment,"[43] reflecting evidence that suggests many patients considering TJA may benefit from more intensive non-surgical treatment. A 2015 RCT in Denmark found that both TKA and non-surgical treatment, which consisted of exercise, education, dietary advice, use of insoles, and pain medication, significantly improved outcomes at one-year follow-up.[52] While non-surgical treatment was only half as effective as TKA, it was associated with a significantly lower risk of serious adverse events.[52]

1.2.2.4 Policies to encourage SDM in advanced OA

In recent years, encouraging SDM in the context of treatment decision-making for advanced OA has become an important policy priority. As noted in Section 1.1, the CMS in the United States planned to begin two Beneficiary Engagement and Incentives Programs in 2018, including a Shared Decision-making Model and a Direct Decision Support Model. The aim of these two models was to encourage SDM for six preference-sensitive conditions, two of which are hip and knee OA.[25] The SDM Model would pay providers $50 for each service furnished, which includes: 1) identifying SDM eligible

beneficiaries; 2) distributing a patient decision aid; 3) providing a SDM consultation; and 4) tracking and reporting. The Direct Decision Support Model would pay decision support organizations to provide web-based patient decision aids, telephone decision support, and mobile e-health applications directly to the Medicare population. Despite an interest at the policy level, the start of this pilot program was delayed in late 2017 because "an insufficient number of [ACOs] were interested in participating in the model."[53] Nevertheless, health management organizations and insurers continue to make investments in programs to encourage SDM.

From a policy perspective, programs to support SDM have two aims: cost-containment and quality-improvement.[15,54] Given these aims, Ibrahim (2017) noted that TJA is an ideal target condition.[54] With respect to cost-containment, TJA accounts for a substantial portion of overall costs of surgical care, and demand is rising rapidly. The use of SDM-interventions may decrease the uptake of surgery, thereby mitigating some health care costs. For example, the Cochrane review of patient decision aids found that across 18 studies in elective surgery, which included over 3,000 patients, the use of decision aids was associated with a 14% reduction in the uptake of surgery (RR=0.86, 95% CI 0.75 to 1.00), but this was not statistically significant.[13] The potential for patient decision aids to reduce the uptake of surgery was cited as a motivating factor in implementing the SDM program at Group Health, with researchers noting that "leaders recognized strong evidence that decision aids for preference-sensitive health conditions

can improve decision quality and patient satisfaction and may reduce rates of elective surgical procedures."[55]

With respect to improving the quality of care, the decision about whether to undergo TJA features a clear trade-off between potential benefits and harms, meaning the appropriate treatment depends on patient preferences. In addition, Ibrahim noted that there is substantial variation in the rates of TJA among racial and ethnic groups despite similar prevalence of OA and access to treatment.[54] A Canadian study has suggested that the odds of an orthopedic surgeon recommending TKA to a male patient is 22 times that for a female patient, which may explain disparities in the uptake of surgery.[56] The differential rates of TJA based on characteristics such as race/ethnicity and gender may reflect overuse in some populations, and/or underuse in others, and the use of SDM-interventions may reduce disparities in treatment and outcomes.

1.2.2.5 Previous studies of SDM-interventions for advanced OA

To date, three randomized controlled trials and two observational studies evaluated the impact of patient decision aids in the context of advanced OA globally. All five studies evaluated patient decision aids developed by the Informed Medical Decisions Foundation and Health Dialoge.[55,57–60] These patient decision aids were developed for either THA or TKA and consist of a 50-minute video and accompanying booklet.

Arterburn et al. evaluated the impact of both THA and TKA patient decision aids that were integrated into standard clinical practice at Group Health, a health system that provides coverage for over 660,000 individuals in Idaho and Washington State in the United States.[55] The study used an observational pre- post-design, with the pre-period running from January 2007 through July 2008, and the post-period running from January 2009 to July 2010. Introducing patient decision aids was associated with a 26% and 38% reduction in the uptake of TJA and TKA, respectively, over the subsequent 6-months.[55] This translated to a 12 to 21% reduction in health care costs. One important contextual factor is that surgeons at Group Health are salaried, thus there is no financial incentive to perform surgery. Limitations of this study include the observational design, which did not include a concurrent control population, and the relatively short time horizon of the analysis. This led the authors to note that "we cannot exclude the possibility that the decision aid implementation has only delayed the timing of joint replacement surgery. It is entirely possible, given the natural history of osteoarthritis, that patients who choose to forgo joint replacement will reverse their decision later."[55]

Bozic et al. evaluated the impact of patient decision aids on informed decision making, and rates of surgery, and the quality of communication during the consultation, in patients considered medically appropriate for THA and TKA.[59] The randomized controlled trial was based in two academic medical centres in California (University of California, San Francisco and Stanford University), and included 123 patients recruited

between September 2011 and May 2012. Overall, a higher proportion of patients in the intervention group reached an informed decision, defined as scoring above 50% on a validated knowledge survey and reporting "having already chosen" on a validated decision-making instrument, compared with controls (58% vs. 33%, p=0.005).[59] Patients reported higher confidence in knowing what questions to ask their doctor (p=0.0034), and surgeons reported higher satisfaction with the quality (p<0.0001) and efficiency (p<0.001) of visits with intervention group participants and rated the appropriateness of their questions higher (p<0.0001).[59] A lower proportion of patients in the intervention (62.3%) compared with the control group (69.4%) chose surgical treatment, though this difference was not statistically significant (p=0.48).[59] However, the authors noted that this study was not sufficiently powered to detect statistical differences in rates of surgery.

Ibrahim et al. evaluated the influence of a decision aid on rates of TKA in black patients using a randomized design.[60] The motivation for this trial was recognition of significant racial variation in the use of TKA, where black patients are significantly less likely to undergo TKA compared to white patients.[61,62] A total of 336 participants were recruited from three university health systems in Pittsburgh between 2010 and 2014. The RCT found that 7.7% of controls and 14.9% of intervention patients underwent TKA within 12 months, a statistically significant increase of 70% (p=0.04).[60] The study authors noted several limitations, including the relatively short follow-up given the "long-term trajectory" of OA.

Stacey et al. evaluated the impact of patient decision aids and a preference report in patients considering THA and TKA, which summarized the patients' knowledge, values, preferred treatment choice, decisional conflict, and clinical assessment results in one page. A pilot RCT recruited 142 patients and found that patient decision aids and a summary report resulted in patients being more knowledgeable (71% versus 47%, $p<0.0001$), and increased the proportion achieving a high quality decision, defined as being both knowledgeable and making a treatment choice that was consistent with their values (56.4% vs. 25.0%, $p<0.001$).[63] The subsequent RCT evaluated the impact on wait times, decision quality, and rates of THA and TKA. The RCT recruited 343 patients between May 2008 and October 2009 from two orthopedic screening clinics in Ottawa, Ontario, Canada, and followed participants for two years. The intervention was associated with a trend towards a reduced waiting time (HR: 1.25, $p=0.065$), and resulted in a greater proportion of patients making a good quality decision (RR 1.25, $p=0.05$). Overall, fewer intervention participants underwent TJA (73.2% vs. 80.5%) though this was not statistically significant ($p=0.12$). As with Bozic et al. this trial was not powered to detect statistical differences in the rates of surgery.

Sepucha et al. evaluated the impact of a quality-improvement effort on use of patient decision aids in routine orthopedic care using a prospective cohort design.[57] This study consisted of a usual care cohort, where patient decision aids were available to be ordered through the electronic medical record (December 2013 to May 2014) and an

intervention cohort, which came after a quality improvement effort which aimed to identify eligible patients and send them the decision aid in advance of the visit (June 2014 to February 2015). The sample included four orthopedic conditions, including hip and knee OA, lumbar spinal stenosis, and lumbar disc herniation, with knee and hip OA accounting for approximately 43% and 26% of the total, respectively. Results suggested that those exposed to patient decision aids were more knowledgeable in both cohorts, and those in the intervention cohort reported greater SDM with their surgeon. Furthermore, those exposed to the patient decision aids were less likely to undergo surgery in both the intervention (42.3% vs 58.8%, p = 0.023) and usual care cohorts (44.3% vs. 55.7%, p=0.45), though the latter was not statistically significant. Limitations include the observational design, a lack of blinding of surgeons, and an inability to determine the effectiveness in the subgroup of individuals with hip and knee OA.

Overall, these studies suggested that SDM-interventions in this context may improve the quality of care, by increasing patient knowledge, the quality of decisions, and reducing disparities in the uptake of TJA. However, questions remain. While current evidence suggests that SDM-interventions may improve the decision-making process, it is unclear whether they result in better health outcomes for patients. The impact on secondary outcomes, such as health care costs, is also unclear. Cost savings have been demonstrated in one observational trial from the United States,[55] however this trial did not account for the cost of providing patient decision aids. Furthermore, physicians in

that trial were salaried, and the findings may not be generalizable to jurisdictions with fee-for-service payment models. Importantly, no analysis has simultaneously considered costs and health outcomes to determine whether SDM-interventions in this context provide value. This is a critical gap in the literature and is necessary to inform implementation and resource allocation decisions for SDM.

In this next section, the economic methods for assessing the value of health care interventions, including SDM-interventions, are described. The term 'value' can be defined in many ways. For instance, value can be defined as "principles or standards of behaviour; one's judgement of what is important in life," or "the numeric amount denoted by an algebraic term."[64] Throughout this dissertation, the term 'value' is used in the economic sense, which defines the value of a good or services as "what would people be willing to trade (i.e., to receive or to give up) so they would be equally satisfied or happy with or without the change."[65] Economic evaluation is a method that can be used as an important source of evidence to quantify value and guide resource allocation decisions in the context of finite health care budgets. In Canada, and other jurisdictions such as Australia and the UK, these techniques have become embedded into decision-making processes to inform how to best spend health care budgets.

1.2.3 Health economics and economic evaluation

Health care resources, including people, time, facilities, and knowledge, are scarce.[66] Consequently, decisions need to be made about how to allocate these finite resources. Within a health care system with a finite budget, allocating resources towards any drug, technology, or service creates an opportunity cost, defined as "the value of the benefits achievable in some other programme that has been forgone by committing the resources in question the first program."[66] Economic evaluation can be used to inform resource allocation decisions.

Economic evaluation is "the comparative analysis of alternative courses of action in terms of both their costs and consequences."[66] The aim of economic evaluation is to minimize opportunity costs by "…ensuring that the value of what is gained from an activity outweighs the value of what is sacrificed."[67] Defining the objectives of health care interventions is necessary to determine whether the value of one course of action is greater than another. There are two dominant perspectives of economic evaluation: the welfarist perspective, and the extra-welfarist perspective.

The welfarist perspective is based in welfare economics and asserts that "social welfare is a function only of individual welfare (or utility) and judgements about the superiority of one state of the world over another are made irrespective of the non-utility aspects of each state."[68] In effect, a welfarist perspective has the aim of maximising individual utility (or satisfaction). By contrast, the extra-welfarist perspective was

developed because governments and decision-makers may wish to consider elements other than (or in addition to) utility in evaluating the impact of health care programs and interventions.[69] It is viewed as a pragmatic approach that "focuses on relevant outcomes contingent on the policy problem at hand,"[69] and in Canada and the UK, decision-making bodies including the Canadian Agency for Drugs and Technologies in Health (CADTH) and National Institute for Health and Care Excellence (NICE) have adopted an extra-welfarist perspective.

There are four types of economic evaluations that differ in how consequences are considered. In a cost-minimization analysis (CMA), it is assumed that the consequences of the alternative interventions are identical, therefore only costs are considered. In reality, this assumption is rarely tenable,[70] leading some commentators to note "that CMA is not only dead but should also be buried."[71] Cost-benefit analysis (CBA) is based on a welfarist-perspective, and considers one or more relevant consequences of interventions, all of which are valued in monetary terms.[72] Monetary values may be derived from the markets where available, and in cases where a functioning market does not exist, can be elicited through hypothetical willingness-to-pay estimates. Cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) are two types of economic evaluation that take an extra-welfarist perspective. CEA values a single consequence, which is measured and valued in natural units, such as life years gained, disability days saved, or cancers detected.[66] CUA considers one or more consequences of interest (e.g., length and quality

of life) which are measured and valued relative to healthy years.[66] Economic evaluation guidelines recommend using CUA, with consequences of health care interventions and technologies measured using quality-adjusted life-years (QALYs).[73] In theory, this 'quality-adjustment' may apply to a broad variety of consequences.[74] In practice, Canadian guidelines suggest using generic measures of quality of life that are valued using societal preferences. One example is the EuroQol 5-dimension (EQ-5D) descriptive system which focuses on health status covered by five domains (mobility; pain; self-care; usual activities; anxiety/depression) that have published preference weights.[73]

1.2.3.1 The QALY

The QALY combines length and quality of life into a single measure, thereby accounting for the impact of health care interventions on both mortality and morbidity.[75] To calculate QALYs, the length of time in a health state is weighted by the quality of life in that state. The quality-of-life weight is called a "health utility" or "health state utility value," and is measured on a scale from zero, which is equivalent to dead, to one, which is equivalent to full health.[76] Negative health state utility values correspond to states considered worse than death.[77] One strength of QALYs is that they can, in theory, be used to evaluate any health care intervention, in any population. As a result, QALYs enable decision-makers to compare the relative value of very different types of programs or interventions.[75]

1.2.3.2 Valuation

Valuation is an important component of economic evaluation. For example, in a CBA, incorporating consequences requires that they be valued relative to money. Performing a CUA requires health state utility values, which are generated by eliciting the trade off between health states and life years. Valuation methods can be divided into two broad categories of methods: revealed preference (RP) or stated preference (SP).[78] RP methods involve "the exploration of people's preferences as (indirectly) revealed through their actions (choices) in markets specifically related to the value of interest."[79] There are several RP methods that have been applied to health, including the travel cost method,[80] hedonic pricing,[81] and averting behaviour.[81,82] While RP data are generally viewed as a robust indicator of preferences, there are several limitations in applying RP methods in health. For instance, many aspects of health care are not traded in markets or decision-makers may require information on new aspects of care, for which there is no market and thus no RP data. [79] In addition, in health care there is often asymmetric information between the patient and their provider and uncertainty about the outcomes of care, meaning that RP data may not reflect patient preferences.[79]

SP methods offer several potential advantages which can either help supplement RP data or estimate preferences in cases where no data exists. SP methods are often called 'preference elicitation techniques' or 'preference-based valuation methods.'[78,83] There has been a rapid increase in the number of SP studies in health.[84,85] There are several SP

methods available, including non-choice methods, such as the visual analogue scale, and trade-off-based methods which are either designed to value a whole good, including contingent valuation, standard gamble, and the time-trade-off, or attributes of a good, including conjoint analysis, standard gamble, and best-worst scaling.[78] In this next section the principles, strengths, and limitations of these methods are described.

1.2.3.2.1 The visual analogue scale

The visual analogue scale (VAS) is a non-choice method to value consequences. It is most often used to estimate health state utility values that can be used to generate QALYs for CUA. The VAS is a variation on a rating scale approach, consisting of a line on a page, with clearly defined end points. For example, the VAS for the EQ-5D, a quality of life scale, includes a range from 0, corresponding to the "worst imaginable health state," to 100, corresponding to the "best imaginable health state."[86] In valuing health states, researchers present respondents with a description of a health state, and ask them to indicate on the scale where they feel that health state fits.

The primary criticism of the VAS and other non-choice methods is that respondents are not asked to trade anything as an indicator of value and therefore do not take account of opportunity cost. This contrasts with trade-off-based methods, where value is determined based on what individuals are willing to exchange for a good or service. For example, using a trade-off-based method might involve asking respondents

how much money, time, risk, health, or another attribute of value, they would be willing to forego for that benefit.

1.2.3.2.2 Contingent valuation

Contingent valuation (CV) is a trade-off-based method that values consequences in monetary terms. In a CV task, respondents are asked to indicate their willingness-to-pay (WTP) for a specific good or service. In health care, CV has been used to value consequences in monetary terms, often for incorporation within a CBA. There are several limitations to CV, but perhaps most notably is that WTP is influenced by ability to pay, which may result in equity concerns.[87]

1.2.3.2.3 Standard gamble

Standard gamble (SG) is a trade-off-based method that values consequences based on how much risk of an undesirable outcome participants are willing to accept.[66] It is widely used to value health states for CUA, though it has also been used to value aspects of the process of care. Generally, a SG task that aims to value a health state involves presenting respondents with two options: A) where they would live in a health state for 10 years (the state being valued), or B) a gamble, described as a probability, with two possible outcomes: immediate death, or being returned to full health. The probability in option B is varied to determine the point at which the respondent is indifferent between the two options. The probability at which the respondent is indifferent corresponds to

the health state utility value, which is bounded between zero and one, and can be used to weight life-years to generate QALYs.[66] The SG method does change when attempting to value health states that are considered worse than death and may be modified when valuing temporary states, or aspects of care that are shorter in duration (e.g., chained-SG).[88]

SG is considered the most methodologically robust valuation method, as it incorporates uncertainty and thus conforms to the fundamental axioms of expected utility theory.[89] As a result, the SG is the only valuation method that elicits true von Neumann-Morgenstern 'utilities.' Throughout this dissertation, the term 'health state utility value' is used, and refers to any value, measured on a scale anchored at one (equivalent to full health) and zero (equivalent to dead), with some negative values (reflecting states worse than dead). This definition includes values estimated through the SG, and those that do not conform to the fundamental axioms of expected utility theory (e.g., VAS).

1.2.3.2.4 Time-trade-off

The time-trade-off (TTO) is a trade-off-based method that values consequences based on the number of life-years respondents are willing to give up.[76] It is used to value health states for CUA and was initially designed as a simpler alternative to the SG. In a TTO task, respondents are presented with two options: A) live for *x* years in a health state (the state being valued), or B) live for less than *x* years in perfect health. The number of

years for option B is varied to determine the point at which the respondent is indifferent between the two options. The number of years at which the respondent is indifferent, divided by 10 years, corresponds to the health state utility value, which is bounded between zero and one, and can be used to weight life-years to generate QALYs.[66] As with the SG, the TTO can be modified to value states considered worse than dead, and modified approaches have been developed to value temporary health states or aspects of care that are shorter in duration (e.g., chained-TTO).

1.2.3.2.5 Conjoint analysis

Conjoint analysis (CA) is a trade-off-based method that values consequences either through ranking or rating hypothetical alternatives which are described by attributes with varying levels.[90] In ranking CA, respondents are asked to order the hypothetical alternatives as a way of representing their preference, whereas rating CA asks participants to consider both the order and strength of their preferences. CA has been used to value preferences for risks and benefits of treatments,[91] and to generate health state utility values for QALYs.[90] However, CA has important limitations. For example, CA is not consistent with economic theory. Secondly, the task asks respondents to rank or rate goods, which is not consistent with conventional decision-making processes which rely on a discrete choice.[92]

1.2.3.2.6 Discrete choice experiments

A discrete choice experiment (DCE) is a trade-off-based method that asks respondents to make a choice between two or more alternatives, which are described using attributes and levels.[79] The use of DCEs in health care is growing rapidly.[84] DCEs are often used to value attributes relative to money using a cost attribute (i.e. willingness-to-pay), which provides information on the relative value of included attributes, and can also be incorporated within a CBA. However, more recently there has been a rise in the number of DCEs estimating the trade-off between health outcomes and experience factors, and valuing outcomes in terms of utility.[84] For example, including life-years within the DCE can be used to estimate health state utility values, which can be used to generate QALYs for a CUA.[93]

1.2.4 Economic evaluation of shared decision-making interventions

To date there have been eight trial-based and three model-based economic evaluations of SDM-interventions published. The trial-based economic evaluations are presented in Table 1.1.

In 2001, Murray et al. published two separate trial-based analyses on the economic impact of decision aids for individuals considering hormone replacement therapy (HRT) [94] and treatment for benign prostatic hyperplasia (BPH).[95] Patient decision aids were associated with increased per patient costs compared to usual care in those considering

HRT (£309 vs. £91, p<0.001) and treatment for BPH (£594 vs. £189, p<0.001). In both cases, much of the incremental cost was related to the video disc technology used for the patient decision aids. With respect to outcomes in the HRT trial, the decision aid was associated with a decrease in the uptake of HRT at three-month follow-up and a decrease in decisional conflict, but it had no impact on anxiety or general health outcomes. In the BPH trial, the decision aid was associated with lower decisional conflict, but had no impact on anxiety or general health outcomes.

This first full economic evaluation of a decision aid was published in 2002. This trial-based CUA explored the impact of patient decision aids in women with menorrhagia.[96,97] Women were randomized to one of three arms, including usual care, a patient decision aid, and a patient decision aid plus an interview-based values clarification exercise. At two-years follow-up, women in the patient decision aid plus interview arm were less likely to have undergone hysterectomy, an elective surgical procedure, compared to those in the usual care arm (adjusted odds-ratio = 0.60; 95% CI: 0.38 to 0.96) and those in the patient decision aid only arm (adjusted odds-ratio = 0.52; 95% CI: 0.33 to 0.82). While neither of the interventions had a statistically significant impact on health outcomes compared to usual care, the patient decision aid plus interview arm was dominant with lower mean costs and higher QALYs (£1,030, 1.582) than the patient decision aid arm (£1,333, 1.567) and usual care arm (£1,810, 1.574).

Vuorma et al. evaluated the impact of a decision aid for menorrhagia, compared to usual care, on health-related quality-of-life and costs. The RCT recruited 363 women and followed them for 12-months. The trial found no significant difference in total costs between the intervention and control (€3,760 and €3,094, respectively, p=0.10) or health outcomes, with just a single significant difference on one of the eight concepts of the RAND-36 (role functioning/emotional).

Hollinghurst et al. evaluated the influence of patient decision aids to inform mode of delivery among women with a previous caesarean section and found that rates of repeat caesarean section were lower in the decision aid arm (0.60, 95% CI: 0.53-0.66) compared with usual care (0.69, 95% CI: 0.62-0.75) though this was not statistically significant.[98] The authors found that the use of patient decision aids resulted in lower incremental costs compared with usual care (-£32, 95% CI: -£172 to £107).[98]

In 2014 Tubeuf et al. performed a within-trial CEA of a decision aid for parents deciding whether to vaccinate their child for measles, mumps, and rubella (MMR). Participants were randomized to one of three arms: an MMR decision aid, an MMR leaflet, and usual care.[99] MMR uptake was higher in the decision aid arm (42 out of 42, 100%) compared the leaflet arm (69 out of 75, 92%) or usual care arm (61 out of 623, 98%) and was associated with lower incremental cost compared to both the leaflet (-£7.17) and usual care (-£9.20), resulting in a high probability of being cost-effective.

In 2014, Patel et al. evaluated the impact of a SDM-intervention for treatment for low back pain in a trial-based CUA. The SDM-intervention consisted of an information booklet for patients and skills training for physiotherapists.[100] The authors found that the decision aid resulted in a lower proportion of patients being satisfied with their care (adjusted odds-ratio = 1.28; 95% CI: 0.79 to 2.09), lower incremental costs (£38 saving per patient), and poorer health outcomes (0.02 fewer QALYs). As a result, the probability of the decision aid being cost-effective at a willingness-to-pay threshold of £20,000 per QALY was just 16%.

The most recent trial-based economic evaluation was completed in 2018. This analysis was a RCT of a patient decision aid for women with breast cancer who were considering breast reconstruction surgery.[101] At six-month follow-up the patient decision aid arm had lower mean per patient health care costs (-$763) with a non-statistically significant increase in QALYs compared to control (0.01, 95% CI: -0.01 – 0.03). Overall the results suggested that the patient decision aid had an 87% chance of being cost-effective at a threshold of $60,000 per QALY.

In addition to trial-based economic evaluations, there have also been three model-based economic evaluations of SDM-interventions published. The first, published in 2015, performed a CUA of a decision aid for adults considering treatment for obstructive sleep apnea. The analysis found that patient decision aids could be cost-effective, provided that the decision aid increased adherence to treatment.[102] The second model-

based evaluation, published in 2015 performed a CEA of patient decision aids in the context of colorectal cancer screening and evaluated costs and life-years saved. The authors found that the decision aid strategy was more expensive ($3,249 vs. $3,023) and resulted in more life-years saved (18.20 vs 18.19) compared with the no decision aid strategy, resulting in an incremental cost-effectiveness ratio (ICER) of $36,126 per life-year-saved.[103] Lastly, in 2016, Penton et al. explored the potential cost-effectiveness of patient decision aids to guide osteoporosis treatment with oral bisphosphonates. The analysis suggested that patient decision aids could be cost-effective if they could improve treatment initiation or adherence by at least 20%.[104]

Table 1.1: Trial-based economic evaluations of SDM-interventions

First Author, Year	Participants, Setting	Strategies	Design and sample size	Cost	Outcome(s)	Cost-effectiveness
Murray, 2001 [95]	Patients with benign prostatic hypertrophy, GP office	(1) patient decision aid, and (2) control	RCT N=112 (60 patient decision aid, 52 control)	Significantly higher cost in intervention group (£594.10 vs. £188.80)	QoL: No significant differences	Not evaluated
Murray, 2001 [94]	Women considering hormone replacement therapy, GP office	(1) patient decision aid, and (2) control	RCT N=205 (103 patient decision aid, 102 control)	Significantly higher cost in intervention group (£306.50 vs. £90.90)	QoL: No significant differences	Not evaluated
Kennedy, 2003 [96]	Patients with menorrhagia, at-home prior to consultation	(1) patient decision aid, (2) patient decision aid + interview, and (3) control	RCT N=894 (296 patient decision aid, 300 patient decision aid + interview, 298 control)	Patient decision aid + interview had lower mean costs than patient decision aid alone and control (£1,030 vs. £1,333 and £1,810, respectively).	QoL: No significant differences QALYs: Patient decision aid + interview had higher mean QALYs than control and patient decision aid alone (1.582 vs. 1.574 and 1.567, respectively)	Patient decision aid + interview was dominant with lower costs and greater mean QALYs.
Vuorma, 2004 [105]	Gynecology patients, clinic	(1) patient decision aid and (2) control	RCT N=363 (184 patient decision aid, 179 control)	No significant difference (€4,607 decision aid vs. €5,164 usual care)	QoL: Statistically significant improvement in the decision aid group on one of the eight concepts of the RAND-36 (role functioning/emotional)	Not evaluated
Hollinghurst, 2010 [98]	Pregnant women, researcher home visit	(1) information, (2) patient decision aid, and (3) control	RCT N= 742 (n)	No significant differences (£2,069 in the information group, compared to £2,019 in the patient decision aid group and £2,033 in the control group)	Decisional conflict: Lower in the decision aid group. Rate of repeat caesarean section: Non-statistically significant reduction in the decision aid arm (0.60, 95% CI: 0.53-0.66) compared with control (0.69, 95% CI: 0.62-0.75)	Not evaluated

Tubeuf, 2015 [99]	Parents considering MMR vaccination, primary care	(1) patient decision aid, (2) leaflet, and (3) control	RCT N=179 (42 patient decision aid, 75 leaflet, 62)	Patient decision aid had lower incremental costs compared to leaflet (-£7.17) and control (-£9.20).	MMR vaccine uptake: patient decision aid had a higher rate (42 out of 42, 100%) compared to leaflet (69 out of 75, 92%) and usual care (61 out of 62, 98%).	Patient decision aid had a probability of being cost-effective ranging from 72% to 88% across a range of monetary values for an additional vaccination from £0 to £100.
Patel, 2016 [100]	Adults with non-specific low-back pain, community physiotherapy service	(1) SDM skills training for physiotherapists and information booklet for patients, (2) control	Cluster RCT N=148 (85 SDM-intervention, 63 control)	SDM-intervention had lower mean, per patient costs (-£38).	QALYs: SDM-intervention had lower mean, per patient QALYs (-0.02).	Incremental cost-effectiveness ratio for control, compared to SDM-intervention, was £1,900/QALY.
Parkinson, 2018 [101]	Women with breast cancer considering breast reconstruction surgery.	(1) patient decision aid, (2) control	RCT N=224 (106 patient decision aid, 116 control)	Patient decision aid had lower mean health care costs (-$763).	QALYs: Patient decision aid had higher mean, per patient QALYs though this was not statistically significant (0.41 vs 0.40)	Patient decision aid was dominant with lower costs and higher QALYs, and had an 87% probability of being cost-effective at a threshold of $60,000 per QALY.

1.2.5 Challenges in the economic evaluation of shared decision-making interventions

Several of the economic evaluations described in Section 1.2.4 have noted issues in applying conventional economic evaluation methods to SDM-interventions. For example, the model-based evaluation by Trenaman et al. stated that:

> "patients using a [decision aid] may legitimately choose a less effective treatment option ... From a patient preference perspective, this may be an appropriate choice. But it is at odds with current economic evaluation methods, which use societal weights for health states that would have assigned fewer QALYs for the worse health outcome. This conflict stems from current QALY measurement techniques that fail to capture some of the known benefits of [decision aids], including the satisfaction a patient might get from receiving the option that is most congruent with his or her values and preferences. New techniques such as discrete choice experiments provide an avenue for valuing these benefits in the future. However, until then, we must assume that current evaluation techniques are underestimating the benefit of [decision aids]."[102]

Cantor et al. also discussed the conflict, noting that making higher quality decisions is the objective of SDM-interventions, but that accounting for this in the analysis is challenging. Specifically, they highlighted issues around determining the value of higher quality decisions, noting that:

> "the present analysis could have revealed that improving decision quality by one point on the scale would cost an additional $10, but no

existing standard is available to help determine whether that increase in cost would be acceptable to policy makers and healthcare providers."[103]

1.3 Aim and objectives of the dissertation

In the context of this literature, the overarching aim of this dissertation is to quantify the economic value of interventions to support SDM in health care. Chapter-specific aims and research objectives are outlined in Table 1.2.

Table 1.2: Chapter-specific aims and objectives

Chapter	Title	Aim	Research Objective(s)
2	Decision aids for patients considering total joint arthroplasty: A cost-effectiveness analysis alongside a randomized controlled trial	To determine whether SDM-interventions provide value.	a) To estimate the impact of patient decision aids plus a surgeon preference report, compared to usual care, on costs, health outcomes, and cost-effectiveness, in adults considering total joint arthroplasty.
3	Long-term impact of a patient decision aid and surgeon preference report on total joint arthroplasty and health care costs	To determine whether SDM-interventions continue to provide value over the long-term.	a) To estimate the long-term impact of patient decision aids plus a surgeon preference report, compared with usual care, on uptake of total joint arthroplasty and osteoarthritis-related health care costs, using administrative data.
4	Capturing the consequences of shared decision-making interventions in economic evaluations	To consider the most appropriate way of evaluating SDM-interventions from an economic perspective.	a) To evaluate the appropriateness of conventional CEA in evaluating SDM-interventions. b) To identify techniques available to value the process of SDM, and ways of incorporating this evidence into economic evaluations of SDM-interventions.
5	What value on elements of shared decision-making? A systematic review of discrete choice experiments	To determine how much the process of SDM is valued based on previous studies.	a) To systematically review studies that have valued SDM using a discrete choice experiment. b) To determine how much SDM is valued relative to money, waiting time, and health outcomes.
6	Incorporating the value of the process of shared decision-making in knee osteoarthritis within the QALY: a discrete choice experiment	To value the process of SDM in a manner that can be incorporated within the QALY.	a) To estimate the health state utility value of the process of SDM in the context of treatment decision-making for advanced knee osteoarthritis.

1.4 Dissertation outline

The dissertation is structured using seven chapters. This introductory chapter has introduced foundational concepts, methods, and the objectives of this dissertation. Chapter Two begins by describing a CEA of a SDM-intervention, which consisted of a decision aid and surgeon preference report, which summarizes the patients' preference for the surgeon, compared with usual care among patients considering TJA. Using data from a RCT with two-year follow-up, this analysis found that decision aids resulted in lower health care costs, driven largely through a reduction in the rate of TJA, and better outcomes, measured in QALYs.

Chapter Three builds on Chapter Two, by evaluating the long-term impact of the decision aids and a surgeon preference report on rates of total joint arthroplasty and health system costs at seven-years follow-up. This was accomplished by linking trial and administrative data using the provincial health numbers of trial participants. This analysis found similar results to the two-year analysis: a smaller proportion of patients in the decision aid arm underwent TJA, which resulted in lower health care costs over the follow-up period.

Chapter Four outlines the challenges associated with using conventional CEA to assess the consequences of interventions to support SDM. Specifically, it highlights how using QALYs that focus on health outcomes may undervalue SDM-interventions, by failing to capture process (e.g., increase knowledge and involvement in decision-making)

or non-health outcomes (e.g., reassurance). It discusses instruments available to measure SDM and the merits of different valuation techniques, including recommendations from CADTH that the value of non-health benefits be measured through the trade-off with health outcomes using societal preferences. Lastly, this chapter discusses different ways of incorporating the value of SDM within an economic evaluation.

Chapter Five builds on Chapter Four, by systematically reviewing studies that have valued SDM using a DCE (n=25). Definitions of SDM vary widely, including both the number of attributes and levels used to describe SDM, and the essential elements covered in attribute and level descriptions. In total, 11 of the included studies valued SDM relative to money, waiting time, or health outcomes. The analysis suggested that respondents, primarily patients, were willing to pay, wait longer, and forego health for greater SDM. However, no studies have valued SDM in the context of treatment decision-making for advanced OA, and none have valued SDM following CADTH guidelines.

Chapter Six builds on Chapters Four and Five, by valuing the process of SDM in the context of advanced knee OA using a DCE. A web-based survey of 1,456 Canadians aged 60 years and older found that respondents were willing to forego potential health improvements for greater SDM in this context. Furthermore, this value was quantified in a manner that can be incorporated with a CEA that uses QALYs as the measure of benefit.

Lastly, Chapter Seven concludes by discussing this program of research, identifying strengths and limitations, implications for practice, and areas for future research.

2 Decision aids for patients considering total joint arthroplasty: A cost-effectiveness analysis alongside a randomized controlled trial

2.1 Introduction

As discussed in Chapter One, there is a paucity of evidence quantifying the economic implications of SDM-interventions. To date, there have been four trials evaluating the influence of decision aids for patients considering TJA, however there are no published CEA. The overarching aim of this chapter is to determine whether SDM-interventions provide value.

2.2 Background

Many health systems and providers across Canada have made patient-centred care a priority.[23,106,107] While the precise definition and meaning of patient-centred care varies, ensuring that care reflects patients' values is often a key component. Central to this goal are efforts that aim to encourage greater shared decision-making (SDM) between patients and providers.[108] One context where SDM may play an important role is for patients considering total joint arthroplasty (TJA) for OA. There are over 100,000 TJAs performed annually in Canada,[70] and for patients with advanced hip or knee OA it has been shown to be both effective and cost-effective.[109] Despite this, not all patients benefit, and undergoing surgery carries risks. Canadian appropriateness criteria state that a patient is an appropriate candidate for TJA when "the patient and surgeon agree that the

potential benefits to the patient of joint replacement surgery outweigh potential surgical risks."[43] However, during brief clinical encounters it can be challenging to ensure that patients are truly informed about the different treatment options, and that their values and preferences are established and communicated to their health care professional. Consequently, SDM interventions have been developed and evaluated for this clinical decision. Patient decision aids may be associated with greater patient knowledge, improved patient provider communication, and higher quality decision-making.[13] Such patient decision aids will require financial and time investments.[110] A previous US study (Arterburn et al.) garnered considerable attention in finding that the provision of patient decision aids resulted in 12% to 21% lower costs over 6-months for patients considering hip and knee arthroplasty.[55] However this study had several important limitations, including an observational design and no evaluation of patient outcomes. No studies have evaluated formally the cost-effectiveness of a patient decision aid intervention for patients considering TJA, and none have conducted a CEA of a patient decision aid in Canada. The objective of this chapter is to estimate the impact of patient decision aids plus a surgeon preference report, compared to usual care, on costs, health outcomes, and cost-effectiveness, in adults considering total joint arthroplasty.

2.3 Methods

2.3.1 Overview

We conducted a CEA using patient level data collected from a RCT designed to quantify health system costs, including all those related to each participant's affected joint, and health-related quality-of-life (HRQoL) in terms of QALYs.[58,66] All analyses were completed from a health systems perspective. The RCT followed up patients for two-years. Costs and outcomes in year two were discounted at 5%, per CADTH guidelines at the time of analysis.[111] This study was approved by the University of British Columbia Behavioural Research Ethics Board and The University of Ottawa Research Ethics Board. The analysis followed methods outlined in the study protocol (Appendix 2.1) and reporting followed Consolidated Health Economic Evaluation Reporting Standards (CHEERS) (Appendix 2.2).

2.3.2 Data

The full results of this RCT have been reported elsewhere.[58] Briefly, patients were recruited from May 2008 to October 2009 at one of two orthopedic screening clinics in the Ottawa area: The Ottawa Hospital (TOH), or Queensway-Carleton Hospital (QCH). Those consenting to participate in the study were randomly assigned to receive a decision aid plus preference report (decision aid arm) or usual care. Clinical history was taken at baseline, with follow-ups at 6, 12, 18, and 24 months. The decision aid arm consisted of a patient decision aid developed by the Informed Medical Decisions Foundation, which

included a video (hip or knee) and accompanying booklet, and a one-page surgeon preference report. The surgeon preference report was compiled by a research assistant, and included information on patient knowledge, values, preferred treatment choice, and decisional conflict. This information was added to standard information from the clinical assessment. Usual care consisted of a standard information pamphlet that outlines preparation for surgery, recovery after surgery, and discharge plans.

In total, 343 individuals were randomized to either the decision aid (n=174) or usual care (n=169) arms.[58] Baseline data were available for 167 participants in each group and are summarized in Table 2.1. Knowledge and decision quality were measured using the validated hip and knee OA decision quality instrument.(109) Patients randomized to the decision aid arm were found to be more knowledgeable and more likely to make a quality decision, defined as scoring ≥66% on the 5-item hip and knee decision-quality instrument knowledge test and making a treatment decision that was congruent with their values (RR=1.25, 95% CI 1.00 to 1.56, p=0.05).[58] The study found that initially, twelve fewer participants in the decision aid arm went on the waiting list and underwent surgery (n=120) compared with the usual care arm (n=132).[58] This trended towards statistical significance though the trial was not powered to detect this difference (RR=0.91, 95% CI: 0.81 to 1.03, p=0.12).[58] During the follow-up period, twelve participants in the decision aid arm and eight in the usual care arm returned to the surgical wait, with three additional decision aid participants undergoing surgery.

Table 2.1: Baseline characteristics of trial participants

		Decision aid arm (n=167)	Usual care arm (n=167)
Age (yrs), mean (SD)		66.1 (9.8)	66.9 (9.1)
Joint (n)	Hip	47	45
	Knee	120	122
HKPT* (total 80), mean (SD)		45.6 (13.8)	45.5 (13.2)
WOMAC* (total 96), mean (SD)		56.7 (17.3)	53.9 (16.0)
Sex (n)	Men	78	64
	Women	89	103
BMI, mean (SD)		31.0 (6.5)	31.8 (6.1)
Language (n)	English	163	164
	Other	4	3
Education (n)	< HS	11	13
	HS/TS	76	70
	College	32	24
	University	48	60
Living arrangement (n)	Alone	39	44
	With someone	128	123
Employment	full time(n)	33	31
	part time (n)	18	12
	retired (n)	106	105
	other(n)	18	11
Household income	<$20,000	14	11
	to $39,999	27	35
	to $59,999	40	35
	to $79,999	34	22
	to $99,999	16	16
	>$100,000	27	32
	no response	9	16

* HKPT: Hip Knee Priority Tool; WOMAC: Western Ontario McMaster Universities Osteoarthritis Index

2.3.3 Costs

Chart review was used to determine whether individuals had undergone surgery. Data on health care resource use for the problem joint (knee/hip) were collected prospectively through paper-based patient diaries at six, twelve, eighteen, and twenty-four months. Patients self-reported whether they had undergone TJA, attended doctor visits or physiotherapy, or filled prescriptions, and the dates of these events. In all cases, it was specified that resource utilization should be related to their "joint problem."

Participants were contacted by phone at each follow-up point to determine their resource utilization. Three attempts were made to contact participants before classifying the follow-up point as 'missing.' The cost of the intervention was calculated based on the time required to compile the surgeon preference report, the cost of the patient decision aid (DVD and booklet), and a surgical consultation. Costs were calculated by multiplying the resource use by average Ontario unit costs (Table 2.2). For physiotherapy visits and medications, it was difficult to determine whether costs were borne by patients or the health care system. Thus, all physiotherapy and medication resource utilization were included in the analysis. Costs were adjusted to 2014 Canadian Dollars using the health care component of the consumer price index. Incremental mean costs between the two arms were estimated with adjustment for baseline utility, using ordinary least squares (OLS) regression.

2.3.4 Quality-adjusted life-years

Societal health state utility values were not collected in the trial. Thus they were estimated using an established mapping algorithm that links WOMAC scores with EQ-5D-3L health state utility values.[115] QALYs for each patient were calculated as the area under the curve following the trapezium rule, which assumes linear interpolation between follow-up points.[117] Incremental mean QALYs between the two arms were estimated with adjustment for baseline utility and clinic site, using OLS regression.[118]

Table 2.2: Average Ontario unit costs for health care resource use

	Cost (2014 CAD$)	Source
Consultations		
GP	$77	[112]
Surgeon	$83	[112]
Specialist	$157	[112]
Procedures		
Hip Surgery	$8882	[113]
Knee Surgery	$7856	[113]
MRI	$63	[112]
X-ray	$32	[112]
Ultrasound	$44	[112]
Allied Health		
Nurse	$44	[114]
Physiotherapy, Massage	$65	*
Acupuncture, Chiropractor	$40	*
Intervention		
Decision Aid	$10	
Research Assistant†	$6	
Surgeon Consultation	$52	[112]

Abbreviations: GP = general practitioner, MRI = magnetic resonance imaging
*assumption based on review of websites; †calculation in Appendix 2.3

2.3.5 Missing data and uncertainty

Missing data were assumed to be missing at random (MAR). The R-package MICE (multiple imputation with chain equations) was used to impute missing data.[119] Predictive mean matching (PMM) was chosen for imputation, and is a method that imputes an observed value from an individual that is similar based on the predictor characteristics.[120] Predictor characteristics are described in Appendix 2.3. A total of ten multiple-imputed data sets were generated, with mean values averaged to provide point estimates for the outcomes of interest.[121] Given the extent of missing data (Appendix 2.3), the multiple imputed data is presented as the base case, with the complete-case analysis explored in a sensitivity analysis. Uncertainty in the outcome estimates were estimated by bootstrapping the data (n=1500).[122]

2.3.6 Cost-effectiveness

The cost-effectiveness of the decision aid arm was evaluated by comparing the costs and QALYs achieved with the usual care arm at two-years of follow-up, using conventional decision rules and estimating ICERs as appropriate. If one intervention resulted in greater mean QALYs and lower mean costs it was deemed cost-effective using the rule of dominance. The ICER is calculated if either treatment arm does not dominate.[123] Uncertainty in the cost-effectiveness estimates were presented using a cost-effectiveness acceptability curve (CEAC).[124]

2.3.7 Sensitivity analyses

Four sensitivity analyses were performed. They included: 1) varying the cost of the intervention, with one assuming there was no cost and the second assuming that no additional surgical consultation was required; 2) varying the discount rate: 0% and 3%; 3) using two different mapping algorithms to link the WOMAC with the EQ5D;[125,126] and 4) excluding individuals with missing data (complete case analysis).

2.3.8 Subgroup analysis

One subgroup analysis was undertaken that looked at only those individuals with knee OA.

2.4 Results

2.4.1 Costs

Mean two-year per patient costs in the decision aid arm were $7,530 (95% CI: $6,876 to $8,114), compared with $8,033 (95% CI: $7,360 to $8,557) in the usual care arm (Table 2.3). The number of surgeries was the main driver of costs in both arms, accounting for approximately 80% of total costs. Cost savings in the decision aid arm were driven primarily by fewer surgeries.

Table 2.3: Mean per patient costs and QALYs, by treatment arm

	Decision aid arm (n=167)	Usual care arm (n=167)	Incremental
Cost (per patient) (2014 CAD$)	$7,530 ($6,876 to $8,114)	$8,033 ($7,360 to $8,557)	-$560 (-$1,358 to $426) †
Intervention	$68	$0	
Surgery	$5,999 n=123	$6,356 n=132	
Surgeon	$360	$376	
GP	$133	$164	
Other physician services	$38	$34	
Allied Health	$886	$1017	
Prescription Drugs	$77	$116	
QALYs (per patient)	1.23 (1.16 to 1.30)	1.21 (1.15 to 1.28)	0.05 (-0.04 to 0.13) ‡
† controlling for baseline utility; ‡ controlling for baseline utility and clinic site			

2.4.2 Quality-adjusted life-years

Over the two-year trial, the mean number of QALYs per patient in the decision aid arm were 1.23 (95% CI: 1.16 to 1.30), compared with 1.21 (95% CI: 1.15 to 1.28) in the usual care arm (Table 2.3). EQ-5D health state utility values and WOMAC scores by follow-up point are presented in Figures 2.1 and 2.2, respectively. Analysis by treatment arm found that both undergoing surgery and delaying were associated with increased quality of life from baseline, however gains were smaller than those who underwent surgery.

Figure 2.1: EQ-5D health state utility values by treatment arm

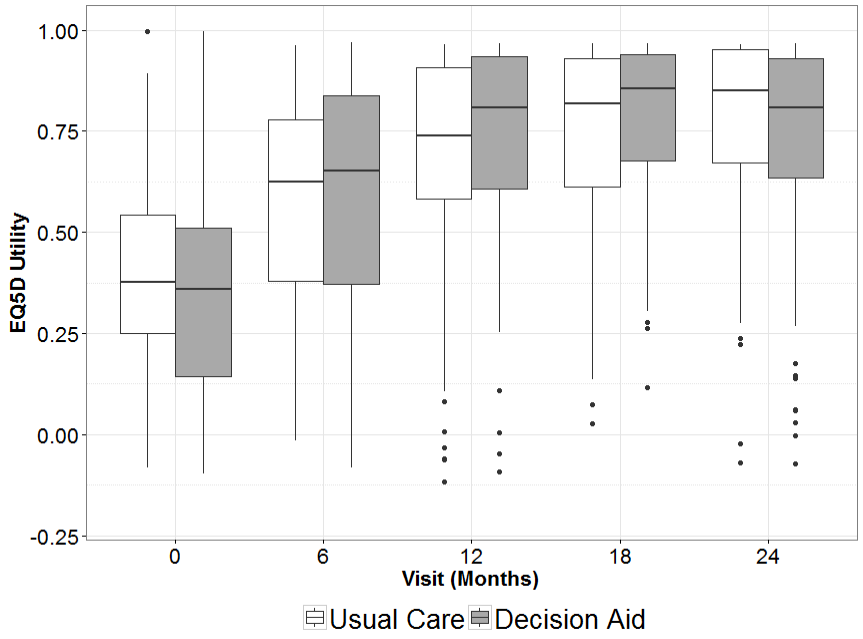
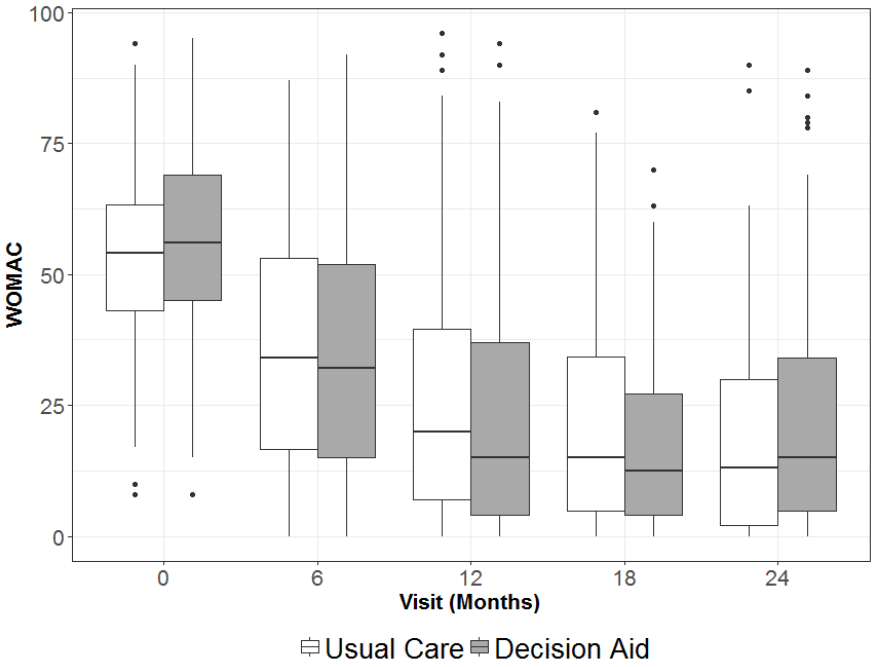


Figure 2.2: WOMAC scores by treatment arm

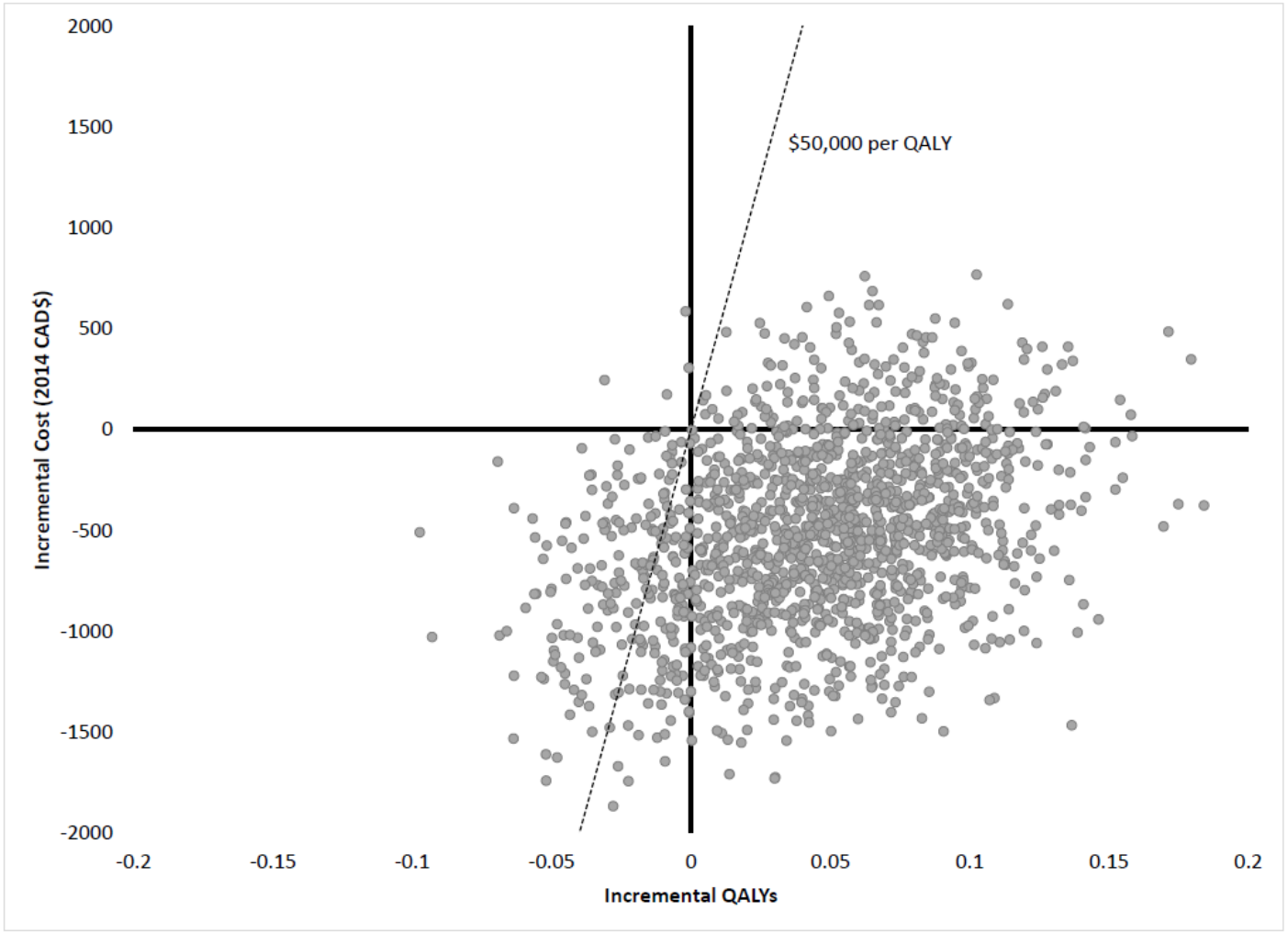


Abbreviations: EQ-5D = EuroQol 5-Dimension; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index

2.4.3 Cost-effectiveness

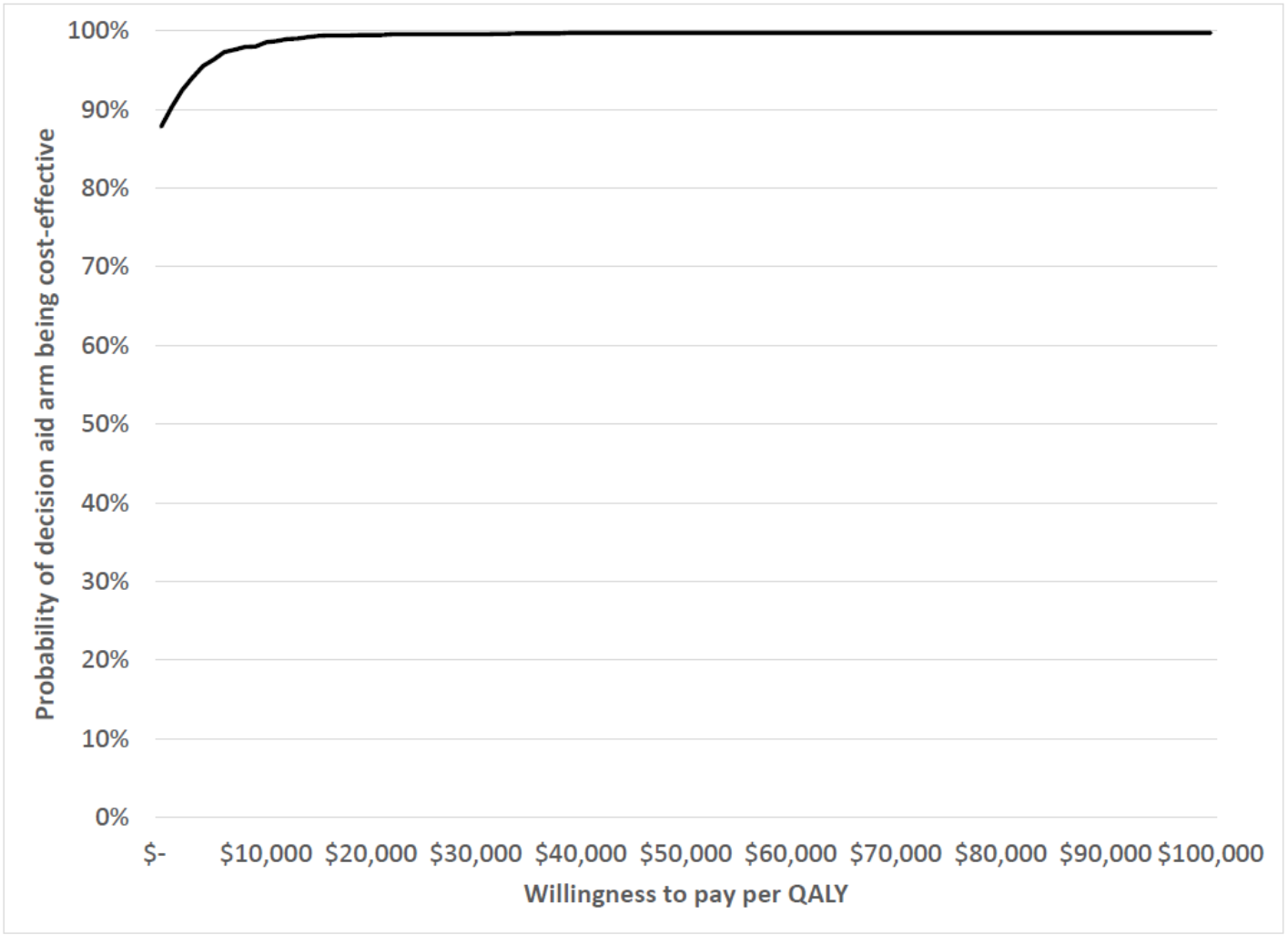
From a health system perspective, the decision aid arm was dominant, providing greater QALYs per patient (0.05, 95% CI: -0.04 to 0.13) at a lower cost (-$560, 95% CI: -$1,358 to $426) than the usual care arm. The cost-effectiveness plane (Figure 2.3) shows that the majority (73%) of bootstrap replications fall in the southeast quadrant, indicating lower costs and greater QALYs. The CEAC indicates that the decision aid arm has a high probability of being cost-effective, ranging from 88% to 99% across willingness-to-pay values of $0 to $100,000 per QALY (Figure 2.4)

Figure 2.3: Cost-effectiveness plane



Abbreviations: QALY = quality-adjusted life-year

Figure 2.4: Cost-effectiveness acceptability curve



Abbreviations: QALY = quality-adjusted life-year

Table 2.4: Sensitivity analyses

	Decision aid arm		Usual care arm		Incremental Cost	Incremental QALYs	ICER
	Cost, $	QALYs	Cost, $	QALYs	$		$
Base case	7,530 (6,876 to 8,114)	1.23 (1.16 to 1.30)	8,033 (7,360 to 8,557)	1.21 (1.15 to 1.28)	-503 (-1,358 to 426)	0.05 (-0.04 to 0.13)	Dominant
Discount rate: 3%	7,547 (6,888 to 8,138)	1.25 (1.18 to 1.31)	8,033 (7375 to 8,593)	1.23 (1.16 to 1.29)	-541 (-1,342 to 362)	0.04 (-0.04 to 0.13)	Dominant
Discount rate: 0%	7,576 (6,930 to 8,157)	1.27 (1.20 to 1.33)	8,099 (7,461 to 8,664)	1.25 (1.17 to 1.31)	-579 (-1,438 to 309)	0.04 (-0.04 to 0.13)	Dominant
Intervention: no cost	7,439 (6,854 to 8,065)	1.23 (1.16 to 1.30)	8,000 (7,407 to 8,581)	1.21 (1.15 to 1.28)	-615 (-1,427 to 229)	0.05 (-0.04 to 0.13)	Dominant
Intervention: no surgeon consult	7,465 (6,831 to 8,067)	1.23 (1.16 to 1.30)	8,015 (7,375 to 8,593)	1.21 (1.15 to 1.28)	-608 (-1,390 to 265)	0.05 (-0.04 to 0.13)	Dominant
Mapping: Barton et al.	7,526 (6,906 to 8,106)	1.15 (1.09 to 1.20)	8,004 (7,373 to 8,579)	1.16 (1.11 to 1.21)	-535 (-1,334 to 287)	0.01 (-0.06 to 0.07)	Dominant
Mapping: Grootendorst et al.	7,532 (6,861 to 8,126)	1.19 (1.14 to 1.24)	7,997 (7,366 to 8,513)	1.19 (1.14 to 1.24)	-519 (-1,309 to 359)	0.03 (-0.03 to 0.09)	Dominant
Complete case (n=158)	8,215 (7,451 to 8,977)	1.33 (1.26 to 1.40)	8,210 (7,466 to 8,924)	1.31 (1.24 to 1.38)	-113 (-1,146 to 900)	0.05 (-0.04 to 0.14)	Dominant

Abbreviations: QALY = quality-adjusted life-year, ICER = incremental cost-effectiveness ratio

2.4.4 Sensitivity analyses

Sensitivity analyses that varied the discount rate, intervention cost, mapping algorithm, and only considered participants with complete follow-up data (n=158) found that the decision aid arm was dominant, resulting in greater QALYs at lower cost (Table 2.4).

2.4.5 Subgroup analysis

The subgroup analyses that considered only participants with knee OA (n=242) found that the decision aid arm was dominant, with lower mean per patient costs (-$535, 95% CI: -$1,546 to $437) and greater QALYs (0.03, 95% CI: -0.06 to 0.13).

2.5 Discussion

This study aimed to estimate the impact of patient decision aids plus a surgeon preference report on costs, health outcomes, and cost-effectiveness, in adults considering TJA. The analysis suggested that a patient decision aid plus surgeon preference report was highly likely to be a cost-effective use of health care resources in a Canadian context. The results were robust to a series of alternative assumptions explored through sensitivity analyses. The primary driver for cost savings was through reduced surgeries.

This was the first study to evaluate formally the cost-effectiveness of a patient decision aid intervention for patients considering TJA, and the first CEA of a patient decision aid in Canada. Two previous studies from the United States evaluated the same

patient decision aid without the surgeon preference report.[55,59] Both found that the decision aid resulted in a reduction in the uptake of surgery, with one finding a statistically significant reduction [55] and the other being non-significant.[59] In the 2014 Cochrane systematic review of 115 randomized controlled trials investigating the effectiveness of patient decision aids, the authors found that in the context of elective surgery, patients exposed to a decision aid are less likely to choose surgery compared to those exposed to usual care.[127] Patients often overestimate the potential benefits of treatments and underestimate the harms,[51] thus this finding may indicate that patient decision aids result in more realistic expectations. While patient decision aids may result in patients delaying TJA, their primary goal is to ensure that treatments are provided in accordance with the values and preferences of patients, not change the uptake of services or health system costs. In the context of TJA, it is likely that patient decision aids will result in some patients choosing to delay surgery who would have not otherwise, and vice versa. A study in Ontario found that many good candidates for TJA are unwilling to undergo surgery, and that this decision is often based on incorrect assumptions, such as a belief that their pain/disability was not yet severe enough to warrant intervention.[128]

In this study, exposure to a patient decision aid resulted in slightly fewer patients undergoing TJA, which in turn resulted in decreased health care costs. Despite this, there was no evidence that delaying surgery had a detrimental impact on health outcomes. This finding could arise for a variety of reasons. The intervention may have encouraged

patients who were most appropriate, or likely to benefit, to undergo surgery. Patient decision aids may play an important role in ensuring that there is appropriate use of TJA. For instance, Canadian appropriateness criteria state that a patient is appropriate for surgery if "the patient and surgeon agree that the potential benefits to the patient of joint replacement surgery outweigh potential surgical risks" and "the patient's expectations for joint replacement surgery are achievable."[43] Patients systematically overestimate the potential benefits of treatment and underestimate the potential harms [51], however in this trial, patients in the decision aid arm were more knowledgeable, which may also explain why patients in this arm had better outcomes. Evidence suggests that patients who are appropriate candidates see greater improvement,[129] and that patients with more realistic expectations report greater HRQoL.[130] This may explain why, despite having fewer surgeries, patients in the decision aid arm reported better health outcomes.

2.5.1 Limitations

This study had several limitations that warrant consideration. Participants were enrolled from an orthopedic screening clinic, where patients with less severe OA were sent back to their referring physician.[131] The results may not be applicable in contexts where patients go directly to the surgeon. With regards to patient population, there is evidence that the rates of TJA are increasing in younger patients, representing a move from 'disability management' to 'disability prevention.'[43,132] The study cannot determine

whether there is a differential effect of the intervention in younger (vs. older) patients, or those with more (vs. less) severe OA. However, the risk/benefit trade-off may change based on age, as getting surgery at a younger age is associated with a greater risk of prosthesis infection, early revision surgery, and more routine placements that occur at 15 to 20 years.

In estimating EQ-5D health state utility values to generate QALYs, it was necessary to rely on a mapping (or 'cross walking') technique from the condition-specific WOMAC measure.[133] The mapping algorithm was developed using data from a registry of Spanish patients, and a value set from the UK population,[115] which may not be representative of the Canadian population. This mapping algorithm was chosen because it included a larger sample of patients than alternatives and individuals with both hip and knee OA. Sensitivity analyses using two other algorithms were conducted, with both revealing incremental QALYs were higher in the decision aid arm. The analysis also relied on self-reported health care resource utilization to derive costs. In some cases it was impossible to distinguish between costs borne by the health care system and those borne by the patient. The unit costs for surgery, which represent a significant proportion of total costs, did not explicitly capture variation in length of stay, in-hospital complications, and other factors that may influence total costs. Missing data were an issue. Data were assumed to be MAR and multiple imputation was used to overcome this

limitation. If the MAR assumption was violated this could lead to biased results, however the results were robust to sensitivity analysis.

OA is a progressive, chronic condition and our current analysis did not capture outcomes beyond the two-year time horizon. As a result, it is unclear whether patients who chose to not to have surgery during the trial simply delayed surgery or chose not to have surgery at all. However, delaying surgery, even for a few years, may have benefits for both health system costs and patient outcomes. From a patient perspective delaying surgery may decrease the probability of needing a revision surgery, thereby avoiding potential surgical risks. From a health system perspective, there is an advantage to delaying surgical costs or avoiding revision surgery altogether.

This analysis was conducted from an economic perspective and suggests with a high degree of confidence that the SDM-intervention was cost-effective. However, policy-makers often consider both economic and clinical evidence when making decisions.[134] The clinical trial upon which this analysis was based found no statistically-significant impact of exposure to a patient decision aid on uptake of TJA and was not powered on this outcome. This conflict between the economic and clinical evidence has been discussed elsewhere [135] and will be explored in greater detail in subsequent chapters.

2.6 Conclusion and future directions

This analysis suggested that the implementation of a patient decision aid and surgeon preference report intervention within the clinical care pathway for individuals

with moderate-to-severe OA could encourage greater patient-centred care at a reduced cost to the health care system, while producing similar health outcomes for patients. The two-year time horizon for the analysis raises questions about whether these results are maintained over the long-term.

3 Long-term impact of a patient decision aid and surgeon preference report on total joint arthroplasty and health care costs

3.1 Introduction

This chapter builds on the RCT reported in Chapter Two, in which the limitations of using a two-year time horizon were described. This limited time horizon may not capture the full economic implications of patient decision aids for individuals considering total joint arthroplasty (TJA). Upon enrollment to the trial, participants were asked to consent to having their trial and administrative data linked. This trial cohort provided the first opportunity to explore the long-term implications of patient decision aids. The overarching aim of this chapter is to determine whether SDM-interventions continue to provide value over the long-term.

3.2 Background

Trial-based economic evaluations have both advantages and disadvantages. For instance, they can provide an early opportunity to estimate the cost-effectiveness of an intervention, and access to person-level data which allows researchers to, among other things, explore differences in cost-effectiveness across subgroups.[136] However, trial-based economic evaluations also have important limitations. This includes limited generalizability, a failure to incorporate all available evidence, and a truncated time

horizon, which limits the ability of policy makers to make informed decisions based on long-term outcomes.[136–138]

Chapter Two reported the results of a trial-based CEA of patient decision aids and a surgeon preference report, compared with usual care, on health system costs and QALYs, for patients considering TJA. Over the two-year time horizon of the trial, fewer participants exposed to decision aids underwent TJA compared with those in usual care (RR= 0.91; 95% CI: 0.81 to 1.03, p=0.121),[58] which resulted in lower costs and improved outcomes. This is the only economic evaluation of patient decision aids in this context.[55,58–60]

While these findings suggested that patient decision aids are potentially cost-effective, the two–year horizon may not be sufficient to evaluate the full economic impact. Of concern from an economics perspective is the influence of patient decision aids on uptake of TJA, which is the most significant driver of health system costs and health outcomes. As stated by Arterburn et al. "it is entirely possible, given the natural history of osteoarthritis, that patients who choose to forgo joint replacement will reverse their decision later."[55] Delaying the cost associated with surgery, even for a short period, is beneficial, as is avoiding the cost associated with future revisions. However, delaying TJA may result in a more complicated or costly surgery later, or increased use of other health care resources, such as physician visits or pain medication. Ultimately, the concern

is that an analysis that considers a longer time horizon may reach a different conclusion than the CEA in Chapter Two.

Using linked data this analysis explored the long-term impact of patient decision aids in the context of advanced OA on resource use and costs. However, administrative data does not contain a measure of health status, so it is difficult to accurately estimate QALYs and conduct a CEA. Thus, the objective of this study was to estimate the long-term impact of patient decision aids plus a surgeon preference report, compared to usual care, on the uptake of TJA and OA-related health care costs, using administrative data.

3.3 Methods

3.3.1 Study design

This study used secondary analysis of linked randomized controlled trial data and administrative health care data.

3.3.2 Setting and participants

Detailed methods for the recruitment of trial participants were reported in Chapter Two. Briefly, 343 patients with moderate-to-severe hip or knee OA were recruited from two orthopedic screening clinics in Ottawa, Ontario, Canada. Participants were randomized to receive either a decision aid plus surgeon preference report or usual care.

3.3.3 Intervention and comparator

As described in Chapter Two, the intervention consisted of a patient decision aid and surgeon preference report for individuals considering THA and TKA. The decision aids were developed by the Informed Medical Decisions Foundation and consisted of a 50-minute video and accompanying booklet. There were unique videos and booklets for those considering THA and TKA. Approximately two-weeks after the clinical visit a study coordinator contacted participants to collect data on outcome measures, including their knowledge, values, preferred treatment choice, and decisional conflict. For participants in the decision aid arm, these measures were combined with patients' clinical assessment to create a one-page preference report that was placed in the patient's file for the surgeon. Usual care consisted of a standard information pamphlet prepared by the hospital which outlines preparation for surgery, recovery after surgery, and discharge plans.

3.3.4 Outcomes

There were two outcomes of interest in this analysis: 1) the proportion of patients undergoing total joint arthroplasty (TJA) at two- and seven-years follow-up, and 2) OA-related health system costs, in 2016 Canadian dollars.

3.3.5 Data sources

Upon entering the trial, participants consented to have their personal health number linked to administrative databases for follow-up. Administrative data included: a) hospital discharge abstracts from the Canadian Institute for Health Information Discharge Abstract Database (DAD); b) physician billings from the Ontario Health Insurance Plan (OHIP); c) inpatient rehabilitation data from the National Rehabilitation Reporting System (NRS); d) prescription medication data from the Ontario Drug Benefit (ODB) database which covers all individuals aged 65 and older, day surgery, outpatient and community-based clinic care; e) emergency department admissions from the National Ambulatory Care Reporting System (NACRS); and f) basic demographic information from the Registered Persons Database (RPDB). These datasets were linked using unique encoded identifiers and analyzed at the Institute for Clinical Evaluative Sciences (ICES). Data were available from trial enrollment until March 31, 2016, resulting in an average follow-up of approximately seven-years.

3.3.6 Analysis

3.3.6.1 Proportion undergoing TJA

Total hip arthroplasties (THA) and total knee arthroplasties (TKA) were identified in the DAD using procedure and diagnostic codes from the International Classification

of Disease, Canadian Classification of Health Interventions (ICD-10-CA/CCI). This included procedure codes starting with "1VA53" for THA, and "1VG53" for TKA. Location of the procedure (i.e., left, right, bilateral) was identified using a supplementary status attribute, "inatloc," corresponding to "L," "R," and "B." Revision THA and TKA were identified using the "incode" status attribute, where "R" indicated the surgery was a revision. Deaths amongst study participants were identified by the "dthdate" field from the RPDB. Participants were censored if their "endofelig" date was to the end of the study follow-up, indicating that they had moved from the province prior to completion of the study.

The proportion of patients undergoing TJA was calculated at two follow-up points: two-years and seven-years. These points were chosen to allow the proportion of patients undergoing TJA over the long-term to be compared with the two-year time horizon of the trial described in Chapter Two. The proportion undergoing TJA at each time point was compared using Cochrane-Mantel-Haenszel chi-squared tests. The proportion came from cumulative incidence plots that account for competing risk of death,[139,140] which were estimated using the cuminc function (R Software). In addition, cumulative risk regression was undertaken using the crr function (R software), to control for site (Ottawa Hospital vs. Queensway-Carleton) and joint (knee vs. hip) which may impact uptake of TJA. Exponentiated regression coefficients from the regression were interpreted as the instantaneous rate of surgery in subjects who had not experienced the

event or who had experienced death.[141] These coefficients and their associated significant tests indicated the direction of association (i.e. whether the rate of surgery was higher or lower) but did not directly quantify the magnitude of this association.[141]

3.3.6.2 Health care system costs

The analysis focused on OA-related resource utilization and costs. For the base case analysis, all resource utilization was included. This included initial and subsequent surgeries, surgical complications, analgesic medications, visits to the GP, and hospital inpatient, outpatient, and day visits. The DAD, SDS, NACRS, and NRS databases were searched with relevant ICD-10 codes to identify surgeries, admissions for complications related to TJA (e.g.,, deep vein thrombosis, acute myocardial infarction), and rehabilitation costs attributable to TJA. Additional billing codes were used to exclude those undergoing TJA with a primary diagnosis other than OA (e.g.,, cancer, motor vehicle accident). The OHIP database was searched to identify physician billings for TJA procedures, in addition to any additional billings with a primary diagnosis of OA. OA-related drugs were identified in the ODB by Anatomical Therapeutic Chemical (ATC) category. All drugs in ATC categories 'M' (Musculoskeletal) and N02 (Nervous system – analgesics) were identified using their drug identification number. The ODB includes all prescription medication costs for individuals aged 65 and older.

Person-level resource utilization and costing followed guidelines on person-level costing using administrative databases in Ontario developed by the Health System Performance Research Network.[142] Costs from the DAD, NACRS, and NRS databases considered variability in resource utilization based on factors such as age and clinical severity. For the DAD, SDS, and NACRS databases, this was accomplished by using the Resource Intensity Weight (RIW) present for each admission. The RIW is a measure used in Canadian acute care hospitals that represents "the average amount of hospital resources used by individuals with a particular condition relative to average resources consumed by other persons."[142] In cases where a RIW was not present, it was assumed to be the mean RIW for all admissions in that database for our sample. For the NRS database, a unique rehabilitation cost weight (RCW) was calculated.[142] For the DAD, SDS, NACRS, and NRS, person-level costs were calculated by multiplying their cost weight by the cost of a standard hospital stay (CSHS) for Ontario, as reported by the Canadian Institute for Health Information (CIHI). A CSHS was available for 2010 to 2014,[143] with costs for 2008-9 and 2015-16 adjusted based on the health and personal care component of the consumer price index. Cost of physician and laboratory services (OHIP) and prescription drugs (ODB) were taken directly from their respective databases. Costs from the NRS were calculated by multiplying RIW weights by provincial cost per CACS weighted case. Data were available for 2008-11, with costs for 2012-16 adjusted using the Health and Personal Care Component of the Consumer Price Index. All costs were adjusted to 2016 Canadian

dollars, and discounted at 1.5% based on CADTH guidelines.[144] This discount rate is different from the rate used in Chapter Two (5%), which reflects an update in the guidelines. Total costs and mean per-patient costs were calculated by database and year. Mean per-patient costs between arms were compared using Welch two sample t-tests. The t-test assumes normality, an assumption that is unlikely to hold for cost data. However, with sufficient sample size, it has been demonstrated that the t-test is valid even when data do not follow a normal distribution.[145] In the base case analysis, costs could have been missing because the patient died, or censored because the patient moved from Ontario, resulting in their health resource not being captured. Costs in the ODB database could have been missing or censored as above but also could have been missing because the ODB does not include people under 65 years of age.

Subgroup analysis was undertaken to evaluate costs in patients considering TKA, and sensitivity analysis was undertaken to evaluate the impact of censoring resource utilization and costs at: 1) a second primary surgery (i.e. THA or TKA, regardless of their initial surgery), and 2) at a second primary surgery on a different joint (i.e. THA if their initial surgery was TKA). These two sensitivity analyses were completed for all participants and the subgroup considering TKA.

3.4 Results

3.4.1 Sample characteristics

Of the 343 trial participants, 324 provided a personal health number for Ontario that enabled their trial and administrative data to be linked (see Figure 3.1). Length of follow-up was similar between the two arms. On average, patients in the decision aid arm were followed up for 6.8 years (SD=1.1 years) compared to 6.7 years (SD=1.0 years) in the usual care arm. Sample characteristics are presented in Appendix 3.1. Overall, a greater proportion of patients were considering TKA (n=236, 73%) compared to THA (n=88, n=27%).

3.4.2 Proportion undergoing TJA

The number of initial, second primary, and revision TJAs during the follow-up period are presented in Table 3.1. At two-years follow-up 119 of 161 (73.9%) patients in the intervention and 129 of 163 (79.1%) patients in the usual care arm had undergone TJA (X^2 =1.23, p=0.27). At seven-years follow-up, 136 of 161 (84.4%) patients in the intervention and 146 of 163 (89.6%) patients in the usual care arm had undergone TJA (X^2 =1.86, p=0.17). Cumulative incidence plots are presented in Figure 3.2. Competing risk regression found that the rate of undergoing initial TJA was not statistically significantly different between the decision aid and usual care arms when controlling for joint and site (HR=0.92, 95% CI: 0.73 – 1.17, p=0.49). The same was true for those in the subset of patients considering TKA (HR=0.85, 95% CI: 0.64 -1.12, p=0.24).

Figure 3.1: Consort flow diagram

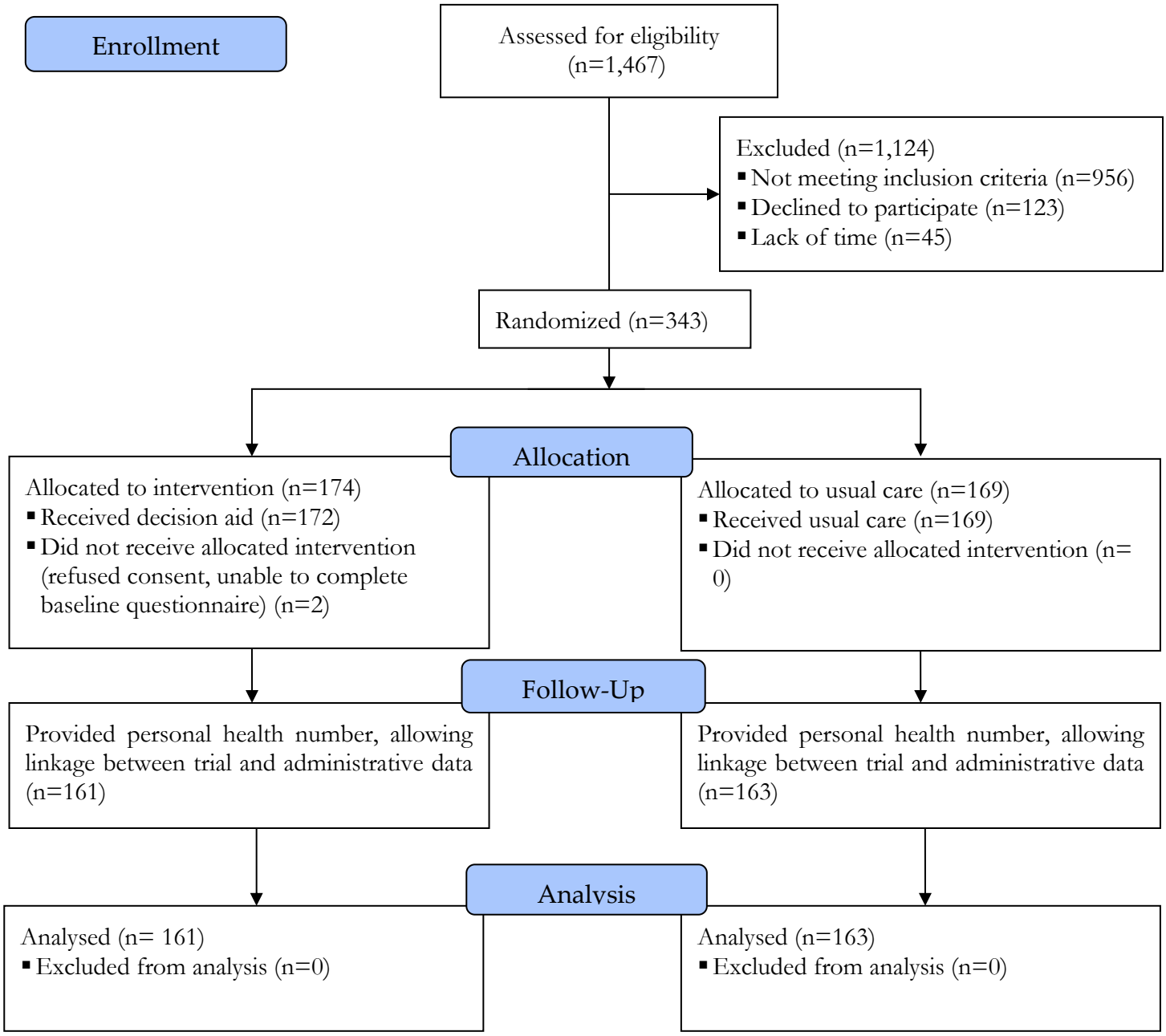
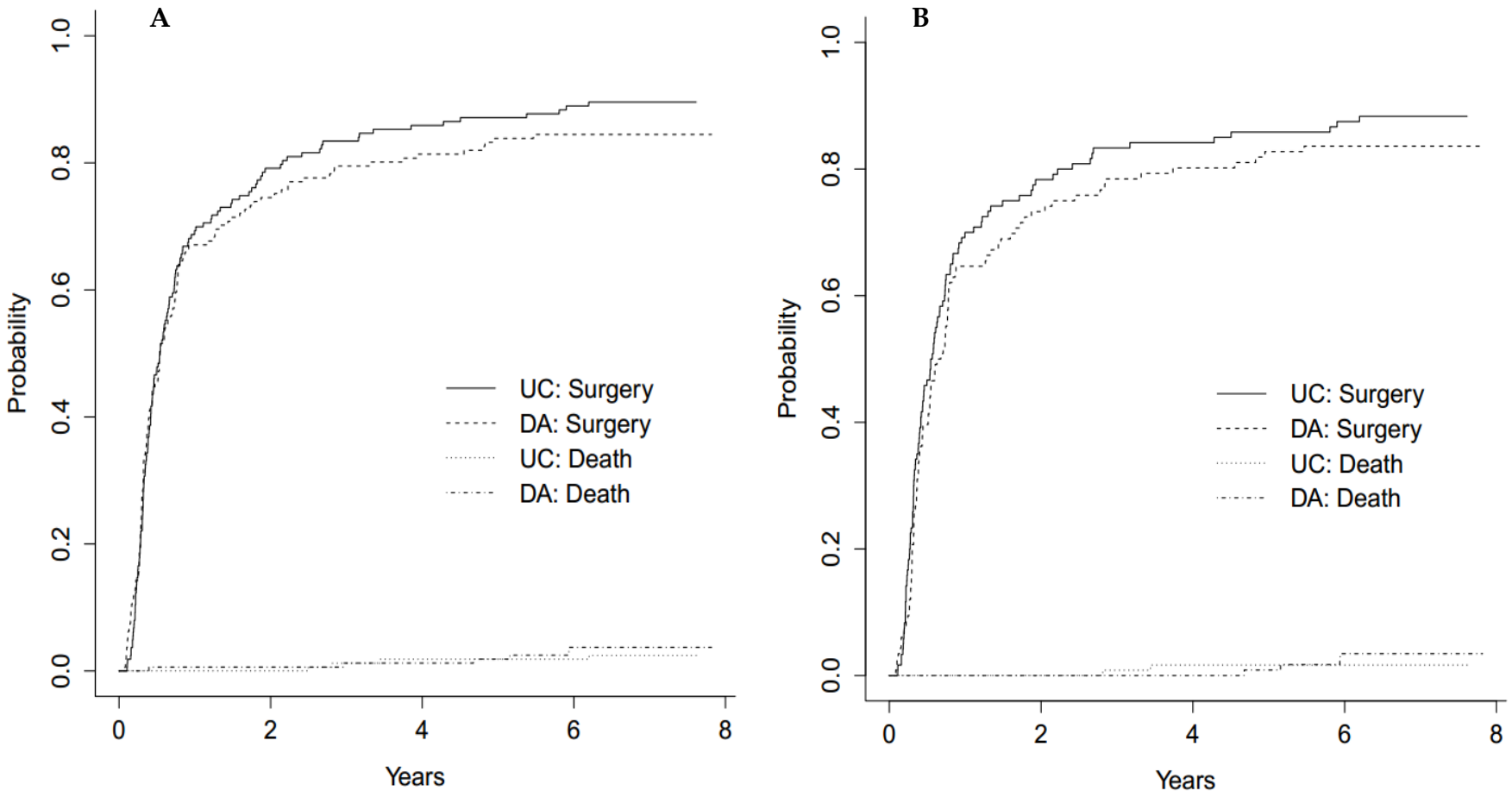


Table 3.1: Number of initial, second primary, and revision surgeries by arm during follow-up

	Decision Aid Arm (n=161)	Usual Care Arm (n=163)
Initial Surgeries	136	146
THA	39	40
TKA	97	106
Second Primary Surgery	20	14
Revision Surgery	8	10

THA = total hip arthroplasty; TKA = total knee arthroplasty

Figure 3.2: Cumulative incidence of surgery by treatment arm for (a) all participants and (b) participants considering TKA



UC: Usual Care Arm; DA: Decision Aid Arm

3.4.3 Health care system costs

Mean and the standard deviation (SD) of costs per patient are reported by database and in Table 3.2. Comparing mean total costs using Welch Two Sample t-test found no significant differences between the two arms (Table 3.2). Total costs by database are reported in Table 3.3. Overall, the decision aid arm had fewer costs than the usual care arm, driven largely by fewer inpatient hospitalizations for surgery captured in the DAD, and fewer rehabilitation costs as captured by the NRS. Subgroup analysis of participants considering TKA found similar results, a non-statistically significant reduction in mean per patient costs in the decision aid compared to usual care arm ($21,043 vs. $23,932, p=0.22).

Sensitivity analysis exploring four alternative scenarios, including different censoring criteria for all participants and only those considering TKA, found similar results: a statistically insignificant decrease in average per-patient costs in the decision aid arm (See Table 3.4).

Table 3.2: Per patient mean, SD costs (2016 CAD$), by database

	Decision Aid Arm		Usual Care Arm		p
Database	Mean	SD	Mean	SD	
DAD: Inpatient Hospitalization	$ 12,755	$ 10,010	$ 13,804	$ 12,110	
DAD: Same Day Surgery	$ 87	$ 452	$ 44	$ 285	
NACRS: Emergency Department	$ 55	$ 188	$ 60	$ 212	
ODB: Medications	$ 1,272	$ 6,522	$ 1,110	$ 3,750	
OHIP: Physician Services	$ 6,471	$ 5,438	$ 6,693	$ 5,362	
NRS: Rehabilitation Services	$ 1,324	$ 2,914	$ 1,970	$ 3,868	
Total	$ 21,965	$ 17,633	$ 23,681	$ 18,178	0.39

Table 3.3: Total costs (2016 CAD$), and 95% CI, by database

	Decision Aid Arm			Usual Care Arm		
Database	Total Cost	2.5%	97.5%	Total Cost	2.5%	97.5%
DAD: Inpatient Hospitalization	$ 2,053,488	$ 1,835,171	$ 2,576,060	$ 2,249,972	$ 1,974,200	$ 2,576,060
DAD: Same Day Surgery	$ 14,057	$ 4,412	$ 25,176	$ 7,115	$ 1,249	$ 14,883
NACRS: Emergency Department	$ 8,820	$ 4,623	$ 13,747	$ 9,833	$ 5,225	$ 15,363
ODB: Medications	$ 204,839	$ 97,657	$ 388,840	$ 181,009	$ 107,519	$ 291,582
OHIP: Physician Services	$ 1,041,904	$ 917,282	$ 1,190,320	$ 1,090,986	$ 973,370	$ 1,223,393
NRS: Rehabilitation Services	$ 213,189	$ 146,899	$ 291,446	$ 321,102	$ 233,500	$ 422,800
Total	$ 3,536,298	$ 3,132,944	$ 3,995,131	$ 3,860,017	$ 3,451,949	$ 4,321,187

Table 3.4: Subgroup and sensitivity analyses, mean per patient costs (2016 CAD$)

Censoring	Patients considering…	Decision Aid Arm	Usual Care Arm	p
None (Base case)	THA or TKA	$ 21,965	$ 23,681	0.39
Second Primary Surgery - Different than initial joint	THA or TKA	$ 21,332	$ 23,316	0.30
Second Primary Surgery - Same as initial joint	THA or TKA	$ 19,170	$ 20,928	0.30
Second Primary Surgery - Different than initial joint	TKA	$ 20,803	$ 23,776	0.20
Second Primary Surgery - Same as initial joint	TKA	$ 18,876	$ 21,017	0.30

3.5 Discussion

Using administrative data, the proportion of patients undergoing TJA at seven-years follow-up was identified to be lower in patients exposed to a decision aid but this difference was not statistically significant. This finding mirrored the results observed during the two-year time horizon of the trial and translated into a non-statistically significant reduction in average per-patient health care costs. While these results are not conclusive, they do address a gap in knowledge, by suggesting that the trend of a fewer patients choosing TJA when exposed to the decision aid may be maintained at seven-years. As was observed in Chapter Two, SDM-interventions may be highly cost-effective despite a non-statistically-significant reduction in costs.

This study population included individuals considering both THA and TKA. Subgroup analysis suggested that the trend of fewer patients exposed to the decision aid undergoing surgery may only be present in those considering TKA, which accounts for approximately 70% of our sample. One explanation for this finding is that the trade-off between potential benefits and harms is less favorable for those considering TKA. On the benefits side, evidence suggests that THA is more effective at improving function, and results in greater satisfaction.[38,146,147] A systematic review of longitudinal studies found that 10% of patients who undergo THA showed no clinically or statistically significant improvement, compared to 30% of those undergoing TKA.[148] With respect to harms, evidence suggests that TKA has a significantly higher rate of infection than THA.[149]

Several limitations need to be considered. First, unlike the two-year trial that included a CEA, this analysis did not quantify patient outcomes. From a resource use perspective, the proportion of patients who underwent TJA and associated health care costs at seven-years follow up mirrored the findings from the two-year CEA. However, the quality-of-life of some individuals who chose not to have surgery may have deteriorated over time. This analysis did not capture this directly. It could have manifested through increased resource utilization, but no significant differences in costs were observed between the two arms. Delaying surgery could be beneficial for patients by reducing the need for future revisions,[55] but could also have a detrimental impact if patients deteriorate and gain less function post-operatively.[150] Another limitation is that the ODB database did not capture drug costs for individuals under age 65 years. In addition, this analysis did not capture cost and outcomes with a lifetime time horizon. This could be addressed by modelling patient outcomes and costs beyond the trial period. This would allow for the full impact of the SDM-intervention on costs and outcomes to be estimated while considering the uncertainty.

An important factor in contextualizing these findings is that the proportion of patients undergoing TJA was not the primary outcome of the original clinical trial, and the trial was not powered to detect a difference on this outcome. Post-hoc analysis of the original trial suggests that this analysis had only 22% power to detect a difference in the uptake of TJA. Thus, these analyses do not provide a definitive answer on whether

patient decision aids in this context reduce the uptake of TJA. Despite this limitation, these results provide useful information to inform future research in this area. For example, in a population with moderate-to-severe OA who are considering TJA, very few individuals may undergo primary TJA after the first five years, suggesting that a trial with a five-year time horizon may be sufficient to evaluate the impact of patient decision aids on proportion of patients undergoing TJA. The results also suggest that the influence of patient decision aids may be present for those considering TKA but not THA. Lastly, these results provide an estimate of uptake of TJA in those exposed to patient decision aids and usual care, which could be used to determine the required sample size for a trial. For example, a trial evaluating the impact of decision aids for patients considering TJA at 7-years follow-up, with a 95% confidence level and 80% power would require a sample size of approximately 650 in each group.

3.6 Conclusions

This analysis suggested that the trend observed in Chapter Two, where fewer patients exposed to decision aids underwent TJA during the two-year trial, may persist for up to seven years. However, these results were not statistically significant, and post hoc analysis suggested that the trial had less than 30% power to detect this result. While these results do not provide a definitive answer on the impact of patient decision aids on uptake of TJA, this is the first evidence on their long-term impact which can inform future research.

4 Capturing the consequences of shared decision-making interventions in economic evaluations

4.1 Introduction

In Chapter Two, concerns were raised about whether CEA, which focuses on the impact on health outcomes, may fail to capture all the relevant benefits of SDM-interventions. The overarching aim of this chapter is to consider the most appropriate way of evaluating SDM-interventions from an economic perspective. The information reported in Chapter Four provides the conceptual basis for subsequent empirical research reported in Chapters Five and Six.

4.2 Background

Performing economic evaluation of SDM-interventions within a conventional CEA poses a challenge. Conventional CEA evaluates the impact of interventions on health outcomes. This focus is not completely consistent with the aim of SDM-interventions to encourage better decision-making, which ideally consists of the patient being well-informed, clear about their personal values and preferences, and making a decision that is congruent with those informed preferences.[151] It is possible that a SDM-intervention could result in patients achieving that primary outcome - making informed, value congruent decisions - but it could nonetheless appear suboptimal in a conventional CEA. This result could stem from a combination of two factors, which are now described.

First, individual values may differ substantially from aggregate societal preferences that are used in a CEA. This assumption has been discussed for health outcomes, where Brazier et al. noted that preferences for many states of the EQ-5D have 'enormous variation' reflecting the heterogeneity in the sample.'[152] Further, the difference in values for "health states has to be as large as 0.20 for 70% of respondents to agree with the ordinal ranking of the health states alone."[152] In addition, conventional CEA assumes risk and time neutrality. If, for example, an individual who is risk averse chooses an option that maximizes their personal expected utility, this may appear suboptimal from a societal CEA perspective.[152] Thus, SDM-interventions may result in patients choosing an option which provides the expected health outcomes that they value most, but this may not be reflected in a CEA that uses societal values.

In addition to the conflict between societal and individual preferences, individuals may consider more than just the potential impact on health outcomes when choosing between treatments. For example, a SDM-intervention may result in a patient choosing a treatment that provides less health benefit because it has, say, a more convenient mode of administration. Within a conventional CEA, the value of a more convenient mode of administration will not be captured. Commentators have raised concerns that the focus on health status, as currently quantified using multi-attribute measures like the EQ-5D, may not be adequate to capture all the relevant consequences of some health care interventions, including SDM-interventions.[102,153,154] SDM-interventions may influence

health status, as measured using societal preferences, but patients, providers, and payers may also value the process of SDM and/or non-health outcomes that arise from it. Failing to consider the broader consequences in CEA may result in SDM-interventions being valued incorrectly.[102,153]

Evidence suggests that decision-makers are willing to consider paying for improvements in the process of care related to SDM. For instance, the 2013 NICE Methods Guide for Technology Appraisal suggests, "if characteristics of healthcare technologies have a value to people independent of any direct effect on health, the nature of these characteristics should be clearly explained and if possible the value of the additional benefit should be quantified. These characteristics may include convenience and the level of information available for patients."[73] While this recommendation suggests that the value of process and non-health outcomes may be considered, it is unclear how these consequences might be incorporated into resource allocation decisions.

The objectives of this chapter are to (a) evaluate the appropriateness of conventional CEA in evaluating SDM-interventions, and (b) identify techniques available to value the process SDM, and ways of incorporating this evidence into economic evaluations of SDM-interventions. This chapter discusses the identification, measurement, and valuation of consequences of SDM-interventions,[66] and takes a health care perspective which defines process and non-health consequences as those "which are an integral part of the types of healthcare evaluated,"[155] but are not necessarily captured

by preference-based health status measures such as the EQ-5D. Notably this definition does not include aspects relevant from the societal perspective such as education outcomes, labor participation, and criminal behavior.

4.3 Identifying the consequences of shared decision-making interventions for incorporation in an economic evaluation

This section aims to address the two questions: *What consequences of SDM-interventions might be valued?* and, *To whom are the consequences relevant?* In an attempt to fully consider which consequences are relevant and to whom, the Donabedian model (see Table 4.1) for evaluating the quality of health care is used.[156] This model defined three consequences that are relevant in health care: 1) structures, 2) processes, and 3) outcomes.

Structures correspond to the conditions under which care is provided. *Processes* refer to the activities that constitute health care. *Outcomes* evaluate the changes (whether desirable or undesirable) in individuals and populations attributable to health care.[157] Notably, the Donabedian model also differentiates between two types of processes; interpersonal processes and technical processes. More detailed definitions for each component of the model are presented in Table 4.1.[156] In addition, health and non-health outcomes are distinguished between each other. The rationale for making this distinction is that it allows an explicit consideration of the consequences that are not captured by current preference-based measures that are used to generate QALYs but are of interest within a health system perspective.

Table 4.1: The Donabedian model and definitions

Model Concept	Definition
Structure	Structure denotes the attributes of the settings in which care occurs. This includes the attributes of material resources (such as facilities, equipment, and money), of human resources (such as the number and qualifications of personnel), and of organizational structure (such as medical staff organization, methods of peer review, and methods of reimbursement).
Process	Process denotes what is actually done in giving and receiving care. It includes the patient's activities in seeking care and carrying it out as well as the practitioner's activities in making a diagnosis and recommending or implementing treatment.
Interpersonal Process	The vehicle by which technical care is implemented and on which its success depends. Through the interpersonal exchange, the patient communicates information necessary for arriving at a diagnosis, as well as preferences necessary for selecting the most appropriate methods of care. Through this exchange, the physician provides information about the nature of the illness and its management and motivates the patient to actively collaborate in care.
Technical Process	Depends on the knowledge and judgment used in arriving at the appropriate strategies of care and on skill in implementing those strategies. The goodness of technical performance is judged in comparison with best practice.
Outcome	Outcome denotes the effects of care on the health status of patients and populations. Improvements in the patient's knowledge and salutary changes in the patient's behavior are included under a broad definition of health status, and so is the degree of the patient's satisfaction with care.

Even with clear definitions it can be challenging to distinguish between processes and outcomes. This has been noted by Donabedian [158] and prominent health economists.[155,159] In this case, the desire to sub-define the consequences as structures, processes, and outcomes is meant solely to ensure that relevant consequences are considered. As stated by Mooney, "for those who want to argue that processes are in fact outcomes and that an outcome is for example 'being autonomous', then I have no real quarrel. I do not think in the end it matters terribly much what we call this phenomenon. My desire to call this something different arises primarily because I believe that these other arguments have been neglected."[160]

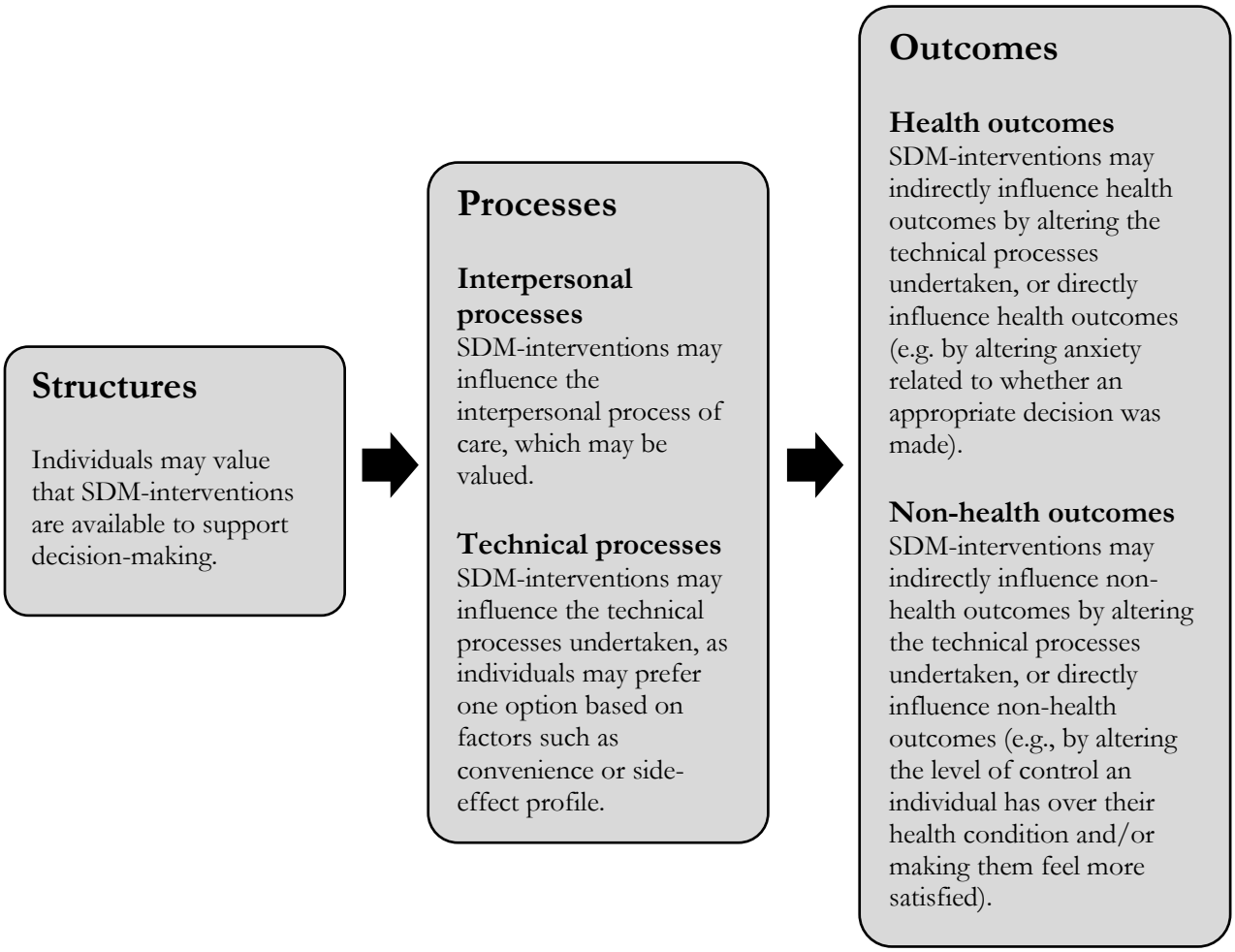
The consequences of SDM-interventions may be relevant for many actors in the health care system. For example, the individual who is the focus of the SDM-intervention may incur consequences because of the process of treatment or subsequent outcomes. In addition, there may be consequences for family members, friends, carers or health care professionals.[161] These consequences are often called 'spillovers.' There may also be consequences that fall on society (or citizens) which are termed 'externalities.'[162] Whether to consider the consequences relevant to these different groups in a CEA will depend on the context and perspective of the analysis. The potential consequences of SDM-interventions and on whom they may impact are now described.

4.3.1 Structures

Figure 4.1 outlines how SDM-interventions fit within the Donabedian model, and the consequences that may arise in terms of structures, processes, and outcomes. SDM-interventions can be classified as structures because they are supported by the institutions that deliver care, such as the integration of patient decision aids within electronic health systems or training practitioners in SDM skills.[10] As structures, SDM-interventions may be valued in and of themselves, and may also impact health care processes and outcomes. For example, a SDM-intervention that involved training general practitioners in SDM (structure) increased patient participation in decision-making (interpersonal process), led to a reduction in use of antibiotics (technical process) without

any negative impact on patient outcomes (health outcomes).[11] SDM-interventions may also provide 'option value,' defined as a "willingness to pay for something for the option to consume the commodity in the future."[163] In the context of SDM, this may refer to value that members of society place on a SDM-intervention on the basis that they may one day be faced with that specific clinical decision. While conceptually valid, this value is likely small.

Figure 4.1: The Donabedian model applied to the evaluation of consequences of SDM-interventions



4.3.2 Processes

The Donabedian model identifies information exchange between the patient and provider and patient engagement in decision-making as interpersonal processes. These are key components of SDM, and goals of SDM-interventions.[6] There is some evidence that SDM-interventions result in patients being more knowledgeable, experiencing increased participation in the decision-making process, and having reduced decisional conflict.[13] Importantly, these consequences may be valued independently of their impact on health outcomes.[162]

Notably, the interpersonal consequences of SDM-interventions are not necessarily beneficial. Provision of SDM-interventions in advance of a consultation may result in patients making a treatment choice prior to meeting with their health care professional, which in turn may inhibit SDM rather than promote it. Patients may also derive disutility from receiving too much or conflicting information or feeling obligated to participate in decision-making. As stated by McGuire et al, "making decisions where such adverse outcomes are possible may involve disutility in itself. Additional to the outcome disutility associated with the adverse outcomes, there is a 'process' disutility associated with having to make (or having made) decisions where the adverse outcome is possible (or has occurred)."[164] The consequences of SDM-interventions on interpersonal processes are likely to vary depending on the personal characteristics of the individual and the clinical context.

SDM-interventions may influence technical processes by changing treatment choice. As stated by Donabedian, interpersonal processes are "the vehicle by which technical care is implemented and on which its success depends."[157] For example, the 2017 Cochrane review of patient decision aids found that across 15 studies, patient decision aids led to a reduction in the number of patients choosing major elective surgery compared with usual care (RR=0.86, 95% CI: 0.75-1.00), though this was not statistically significant.[13] A patient choosing an alternative treatment may reflect a patient preference for different aspects of the treatment itself, such as a more convenient mode of administration, but may also impact downstream health and non-health outcomes.

Process spillovers may result from SDM-interventions. For example, interpersonal spillovers may arise when SDM-interventions result in relatives of nursing home patients being more informed and involved in the decision-making around end-of-life care.[165] Technical process spillovers may arise if a SDM-intervention results in a patient choosing a treatment that requires greater informal care, which can result in either utility gains [166] or disutility [167] for carers. Lastly, positive externalities may arise if members of society care that SDM-interventions result in patients being treated in accordance with their preferences.

4.3.3 Outcomes

SDM-interventions may impact both health and non-health outcomes. For example, SDM-interventions may result in improved health outcomes if patients choose treatments with greater expected benefit, are more adherent to their chosen treatment,[168] or experience a reduction in anxiety due to reduced decisional conflict. A recent study in cancer patients found that participants had lower anxiety and depression scores immediately following, and three months post-consultation with physicians trained in SDM.[169] SDM-interventions may result in poorer health outcomes if patients choose less effective treatments. For example, a patient may be very risk averse, and thus prefer conservative management to elective surgery, despite the elective surgery providing greater expected health benefit. This is a 'good' decision for the patient as it corresponds with their personal values and preferences; but is not currently reflected in the QALY valuation. SDM-interventions may impact non-health outcomes by encouraging patients to choose a treatment with which they are more comfortable, which in turn may result in feeling treated with dignity and respect, empowered to manage their health condition, and/or satisfied with their care.[169,170]

SDM-interventions may result in health or non-health outcome spillovers for friends, family members, and carers. Spillovers of health outcomes have been considered in economic evaluations for a wide range of health conditions,[171] but have not been investigated in economic evaluations of SDM-interventions. SDM-interventions may also

result in health-related externalities.[68] For instance, a SDM-intervention that resulted in greater uptake of a vaccine could generate positive health externalities.[99]

4.3.4 A case study: advanced knee osteoarthritis

Not all the consequences outlined in Section 4.3.3 will be relevant for each SDM-intervention. Using a specific case study can provide an illustrative approach to identify which consequences may be worth considering in an economic evaluation that takes an extra-welfarist perspective. In applying the Donabedian model to the knee OA context, it is clear that implementing a SDM-intervention into practice may result in a variety of consequences (see Table 4.2). There is evidence that patients with knee OA value SDM (interpersonal processes), and that the implementation of these interventions change the treatment choice (technical processes), resulting in fewer patients choosing to undergo surgery.[55,58,59] This choice may impact the patients' health outcomes, or have an influence on non-health outcomes, such as regret if patients are dissatisfied post-surgery.

Table 4.2: Potential influence of a SDM-intervention on structures, process, and outcomes in advanced knee osteoarthritis

		Individual	Friends, Family Members, Society
Structure		Patients may value that a SDM-intervention is available to support decision-making.	Friends and family members may value that a SDM-intervention is available to support decision-making for their loved ones.
Process	Interpersonal	Patients may value the process of SDM in choosing between undergoing surgery and delaying/avoiding surgery. [172]	Family members may value knowing patients have made an informed decision about whether to undergo surgery.
	Technical	SDM may result in more patients delaying or avoiding surgery, [55,59] which in turn may decrease waiting times.[58]	Family members or friends may prefer patients avoid surgery on the basis that it carries a small risk of death or other serious complications.
Outcome	Health	Delaying or avoiding surgery may result in improved health outcomes over the short-term (no post-op recovery period or surgical complications) but poorer health outcomes in the long-term.[173] Risk averse patients may choose to avoid surgery due to potential surgical complications, resulting in poorer expected health outcomes.	The patients' partners may experience stress if their partner experiences a worsening of pain or mobility as a result of delaying surgery.[174]
	Non-Health	SDM may increase patients confidence in their decisions,[175] or may result in regret if patients choose a treatment and are dissatisfied.	

4.4 Measuring and valuing the consequences of shared decision-making interventions for incorporation in an economic evaluation

In Section 4.3 the potential consequences of SDM-interventions that may be relevant in an economic evaluation were identified. Conventional CEA using generic HRQoL measures will capture health outcomes arising from treatment choices. However, this may fail to capture process (e.g., increased knowledge, choice, autonomy, participation in decision-making) and non-health outcomes (e.g., being treated with dignity and respect, satisfaction) that are valued.

Incorporating the value of the process and non-health consequences of SDM-interventions into an economic evaluation requires that they are measured. There is considerable heterogeneity in the definitions of SDM, akin to the different health state descriptive systems used to measure health outcomes. Different measures of SDM will capture different consequences, and to date over 28 different instruments to measure SDM have undergone psychometric testing.[176] Scholl et al. describe three categories of measures: decision antecedents, decision process, and decision outcomes.[177] Decision antecedents measure "features of the patient, provider, or organization."[176] Measures of the decision-making process assess the features of behaviours during the consultation,[176] and measures of decision outcomes might include patient knowledge, decision quality,[175] and decision regret.[178] Lastly, measures may assess the patient's perspective,[179] provider's perspective,[180] or that of a third-party observer.[181]

Once measured, it is possible for the consequences of SDM to be considered alongside evidence from an economic evaluation. However, guidelines from CADTH suggest going further, and valuing these broader consequences (what they term 'non-health effects'). Specifically, CADTH argues that "the value of non-health effects should be based on being traded off against health."[144] The guidance from CADTH highlighted above reflects that considering broader consequences in the economic evaluation of SDM-interventions may impose a cost to the health care system. As noted by Culyer, a departure from health maximization "costs lives, or at least the quality of lives."[182] In the

current context, the trade-off to consider is whether decision-makers allocating health care resources are willing to forego potential improvements in health status for process or non-health benefits generated by SDM-interventions. In effect, the CADTH guidelines aim to inform this decision by requiring evidence that individuals are willing to forego health to achieve better process or non-health outcomes.

While guidance from CADTH does not suggest a specific valuation technique, the guidance does imply that trade-off-based methods are preferred to non-trade-off methods (e.g., rating scale, visual analogue scale). There are several trade-off methods available to value the process of SDM, including the SG, TTO, CV, DCEs, BWS, and CA.[78] Additional methods, including swing weighting, measure of value, analytical hierarchical process, allocation of points/budget pie, and person trade-off are available, but a previous review found almost no examples of their use for valuing health care processes.[183]

The SG and TTO are the most widely used methods to value health outcomes for economic evaluation.[76] As a result, valuing process and non-health consequences using these techniques has the advantage of being consistent with the evidence used in conventional CEA. However, one challenge in using these methods is that they ask respondents whether they would be willing to trade life-years, or a small risk of instantaneous death, for the benefits being considered. In the case of health outcomes, this trade-off is realistic (e.g., *Would you accept a small risk of instantaneous death to have*

better overall quality-of-life for X years?), however this may be unrealistic for consequences that have a short duration, such as temporary health states and health care processes. For example, in valuing SDM, a SG question would effectively be asking: *Would you be willing to accept a small risk of instantaneous death to have greater SDM with your doctor?* Higgins et al. reviewed studies that valued convenience-based aspects of process, and noted that "the relatively small proportion of studies using traditional forms of utility assessment is to be expected, given their inherently longer-term scope: attempting to value convenience in the context of potential instantaneous death (as per the SG) is very likely to induce ceiling effects, given the difference in magnitude between the seriousness of the two concepts."[184] To overcome this limitation, researchers developed chained valuation approaches, including the chained-TTO and chained-SG. These techniques are routinely used when valuing temporary health states,[185] or processes relative with the aim of incorporating the value within the QALY.[159]

Techniques such as DCEs and BWS studies allow researchers to quantify the trade-off between attributes of a good. When valuing SDM, trade-off methods would allow researchers to value components of SDM relative to each other (e.g., information vs. decision-making) and/or relative to other attributes of interest (e.g., health outcomes). While traditionally attributes have been valued relative to cost by including a monetary attribute (e.g., willingness-to-pay), a previous systematic review has shown an increase in the number of studies aiming to estimate health state utility values.[84]

Regardless of the valuation method used, an additional consideration is: *Whose preferences should be considered?* Economic evaluation guidelines recommend using societal preferences for health outcomes when calculating QALYs, as opposed to patient or provider preferences.[73] For consistency, CADTH guidelines also suggest using societal preferences to value non-health effects, such as SDM.[144] A fulsome debate of the potential merits of each approach is beyond the scope of this chapter, and has been discussed elsewhere.[186] However, it is worth highlighting that value may vary depending upon whose preferences are used. For example, on average, patients report a smaller impact of health impairment than is expected by members of society.[187–191] While there is little evidence on whether societal and patient preferences vary systematically with respect to processes or non-health outcomes, recent evidence has shown that patients value processes more, and outcomes less, than providers.[192]

4.5 Incorporating the results into economic evaluation

An important consideration is how the results of valuation studies should be incorporated into an economic evaluation. A 2014 systematic review evaluating process utility included 27 studies, and concluded that "a preference for convenience-related process utility exists, independent of health outcomes … however it is difficult to assess how large such a preference might be, or how it may be effectively incorporated into an economic evaluation."[184] The impact of process utility may not be inconsequential. In evaluating a SDM-intervention for cardiovascular prevention, researchers found that

over a quarter of subjects reported that the disutility of a daily preventative tablet was greater than the expected health outcomes, even among those at high-risk.[193]

Broadly speaking, the value of process and non-health consequences can be aggregated within a single outcome, such as the QALY or net-monetary benefit, or disaggregated. Each approach has merit. A single outcome may be easier for decision-makers to act upon, although it does have the potential to mask important considerations. Alternatively, "the use of multiple outcome measures presents decision makers (such as NICE) with the problem of how to use such measures to make comparisons across sectors or how to combine them to provide an overall measure of benefit whilst avoiding double counting."[194] As mentioned in Section 4.4, several studies have aimed to value processes in a manner that can be incorporated within the QALY.[159] There have also been efforts to capture consequences using multiple outcomes, an example being a recent trial-based economic evaluation of pharmacy services that reported the results of a DCE alongside a cost per QALY analysis.[195]

With respect to guidance, the Second US Panel on Cost-effectiveness in Health and Medicine stated that "It would be helpful to inform decision makers through the quantification and valuation of all health and non-health effects of interventions, and to summarize those effects in a single quantitative measure ... however, there are no widely agreed on methods for quantifying and valuing some of these broader effects in cost

effectiveness analyses."[196] Thus, despite a desire for a single measure, the panel suggested presenting disaggregated consequences in an 'impact inventory.'[196]

4.6 Discussion

This chapter has summarised the issues related to the economic evaluation of SDM-interventions. It has demonstrated that conventional CEA using QALYs captures the impact of SDM-interventions on health outcomes but may fail to capture process and non-health outcomes. The different types of instruments available to measure SDM and methods available to value SDM were described. In addition, the approach to incorporate the resulting value into an economic evaluation was described.

Many topics described in this chapter, including process utility, spillovers, option value, and externalities, are not unique to SDM-interventions and there has been methodological work aimed at incorporating this value into economic evaluations.[159,171,184,197] However the environment appears to have changed in recent years, suggesting that there is a desire for a broader perspective for the QALY. Considering the value of process and non-health consequences in economic evaluations of interventions that span health and related sectors has been discussed.[194] For instance, Wildman et al. outlined a number of challenges in the economic evaluation of assisted living technologies which span health and social care.[198] Greco et al. made a similar case for public health interventions in low- and middle- income countries.[154] Even with a focus on the health sector a clear desire to consider consequences beyond health outcomes has

been demonstrated. This desire is best evidenced in strategic planning documents that identify greater patient engagement in health care decisions as a key objective of health systems.[22,23]

In identifying techniques to value the consequences of SDM, economic evaluation guidelines that recommend valuing non-health consequences through the trade-off with health outcomes were highlighted.[144] It is unclear whether individuals are willing to make this trade-off. For example, SDM may be viewed as a basic standard of care or ethical imperative that is not tradable.[199,200] Future research should explore whether this trade-off is acceptable, and in which contexts.

The value of these process and non-health consequences of SDM are likely to vary based on factors such as the magnitude of the decision and the extent to which the decision is preference sensitive.[201] Demographic characteristics are also associated with preferences for information and involvement in decision-making. Evidence suggests that younger, more well-educated, and female patients prefer a more active role in decision-making.[202] Maximising the potential benefit of SDM-interventions will require tailoring the level of information and decision-making involvement to the preferences of each individual patient.

The impact of changing the evaluative space to include consequences beyond health status has implications for decision-making. Recent work comparing health to sufficient capability in treatments for drug addiction concluded that "different evaluative

spaces and decision-making rules have the potential to offer opposing treatment recommendations."[203] Thus while health systems appear to be willing to consider additional consequences, how best to provide this evidence is less clear. Methodological work is needed to determine how to incorporate additional consequences into economic evaluation, and the subsequent impact on decision-making.

4.6.1 Future research

Future research could begin by exploring whether SDM is valued. Relevant questions include: *In which contexts has SDM been valued? How was SDM described? Which valuation techniques have been used? Whose preferences were elicited? What has SDM been valued against? Was the aim to supplement traditional CEA? And if so, was the aim to aggregate the value within a single measure, such as the QALY, or present it in a disaggregated format*? More methodological work is needed to understand how such evidence can best be integrated in an economic evaluation to allow policy makers to make informed decisions. Chapters Five and Six explore these questions in more detail, while maintaining a focus on the case study of treatment decision-making for advanced OA.

5 How much is shared decision-making valued? A systematic review of discrete choice experiments

5.1 Introduction

This chapter builds on Chapter Four, which demonstrated that process and non-health consequences of SDM-interventions may not be captured in a CEA using QALYs. In Chapter Four, methods for valuing SDM, such as discrete choice experiments (DCEs), and ways of incorporating this evidence into an economic evaluation were described. The overarching aim of this chapter is to determine how much the process of SDM is valued based on previous studies.

5.2 Background

Given that health care resources are limited, investing in SDM may require sacrificing other health system objectives. Quantifying the value of SDM can help ensure that investments are justified given other priorities. SDM-interventions may provide value by improving patient outcomes or reducing health care costs, as was demonstrated in the CEA from Chapter Two.[15] However, as detailed in Chapter Four, SDM-interventions may result in additional process and non-health consequences, such as being more informed, involved in decision-making, and having greater satisfaction with care, which are not captured in CEA.

A DCE consists of a series of hypothetical questions, where respondents are asked to make a choice between two or more alternatives, where each is described using distinct attributes (or 'characteristics').[204] Data from a DCE quantifies the trade-off between the included attributes, and thus can be used to determine the degree to which attributes are valued relative to each other.[183] Given that DCEs are arguably the most widely used method to value aspects of health and health care, the objectives of this chapter are to (a) systematically review studies that have valued SDM using a discrete choice experiment, and (b) determine how much SDM is valued relative to money, waiting time, and health outcomes.

5.3 Methods

A systematic review, produced in accordance with York guidance [205] and PRISMA reporting criteria.[206]

5.3.1 Eligibility criteria

Any empirical (i.e., no reviews, instrument development, or guidelines), peer-reviewed, English language study that used a DCE to value SDM in a health care context was eligible for inclusion. There were no exclusions based on population or clinical context. A DCE was defined as a choice-based stated-preference survey that described goods/services in terms of attributes. SDM was defined based on the nine 'essential elements' suggested by Makoul.[6] These nine elements include: 1) define and explain the

health care problem; 2) present options; 3) discuss pros and cons (benefits, risks, costs); 4) clarify patient values and preferences; 5) discuss patient ability and self-efficacy; 6) present what is known and make recommendations; 7) check and clarify the patient's understanding; 8) make or explicitly defer a decision; and 9) arrange follow-up. Given the variability in definitions of SDM, a minimal criterion was that the element 'make or explicitly defer a decision' should be present in one of the attribute or level descriptions. This criterion was meant to ensure that elements of SDM were being valued in a context where it was clear that there was a decision to be made.

5.3.2 Information sources and search strategy

A previously published review that identified all DCEs in health care served as a starting point for this review.[207] The comprehensive search strategy included search terms used in previously published systematic reviews,[84,208] and has been used to develop a database of DCEs completed in health care (Table 5.1). The previous reviews covered the period from 1990 to 2015. The search was run in MEDLINE (OvidSP) and aimed to update the available database by identifying newly published health care DCEs between January 1, 2015 and February 8, 2016.

Table 5.1: Search terms

discrete choice experiment$
discrete choice model$
stated preference$
part-worth utilit$
functional measurement
paired comparison$
pairwise choice$
conjoint analysis
conjoint measurement
conjoint stud$
conjoint choice experiment$
1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
limit 12 to yr="2015-2016"

5.3.3 Study selection

A two-stage study selection procedure was undertaken to: 1) identify any empirical studies using a DCE in a health care context and 2) identify DCEs that included one or more SDM attributes. In stage-one, two reviewers with expertise in DCEs independently reviewed titles and abstracts to identify health DCE studies that met the inclusion criteria, with discrepancies resolved through discussion. Relevant publications were added to the database of health care DCEs. In stage-two, all studies in the DCE database were reviewed in full-text by one study member (LT). Any study that included one or more of the nine essential elements of SDM in attribute or level descriptions was retained. Full-text studies were then reviewed in duplicate (LT, NB) to identify those that included the essential element 'make or explicitly defer a decision,' with discrepancies resolved through discussion.

5.3.4 Data extraction

Data extraction included elements derived from a checklist designed to assess the quality of DCEs. This checklist has been used previously [207] and focuses on: 1) choice question format, attributes, and levels; 2) experimental design criteria; 3) sample and survey administration; and 4) analysis procedure and statistical tests. In addition, the quality of included studies was assessed using the PREFS checklist, which includes five items and results in an aggregate score ranging from 0 to 5, with higher scores indicating greater quality.[209]

5.3.4.1 Attribute classification

Qualitative information from attribute and level descriptions were extracted, and all attributes were classified according to the Donabedian model for evaluating quality in health care.[156] This model outlines three dimensions in health care, including structures, processes, and outcomes. Structures are defined as "the setting in which care occurs," which includes material resources, human resources, and organizational structure. Processes "denotes what is actually done in giving and receiving care," and can include interpersonal processes (i.e. the interpersonal exchange between the patient and provider) and technical processes (i.e., appropriate diagnoses and technical skill in delivering care). Outcomes "denote the effects of care on the health status of patients and

populations." Outcomes were subclassified as either 'health outcomes' or 'non-health outcomes.' Two study members (LT, SM) independently classified attributes with discrepancies resolved through discussion. According to the Donabedian model, SDM is an interpersonal process. All attributes classified as 'interpersonal processes' were analyzed against the nine essential elements to determine whether the attribute was related to SDM, and the number of elements of SDM present in each study.[6] Coding was completed in duplicate (LT, SM) with discussion to resolve discrepancies.

5.3.4.2 Value of SDM

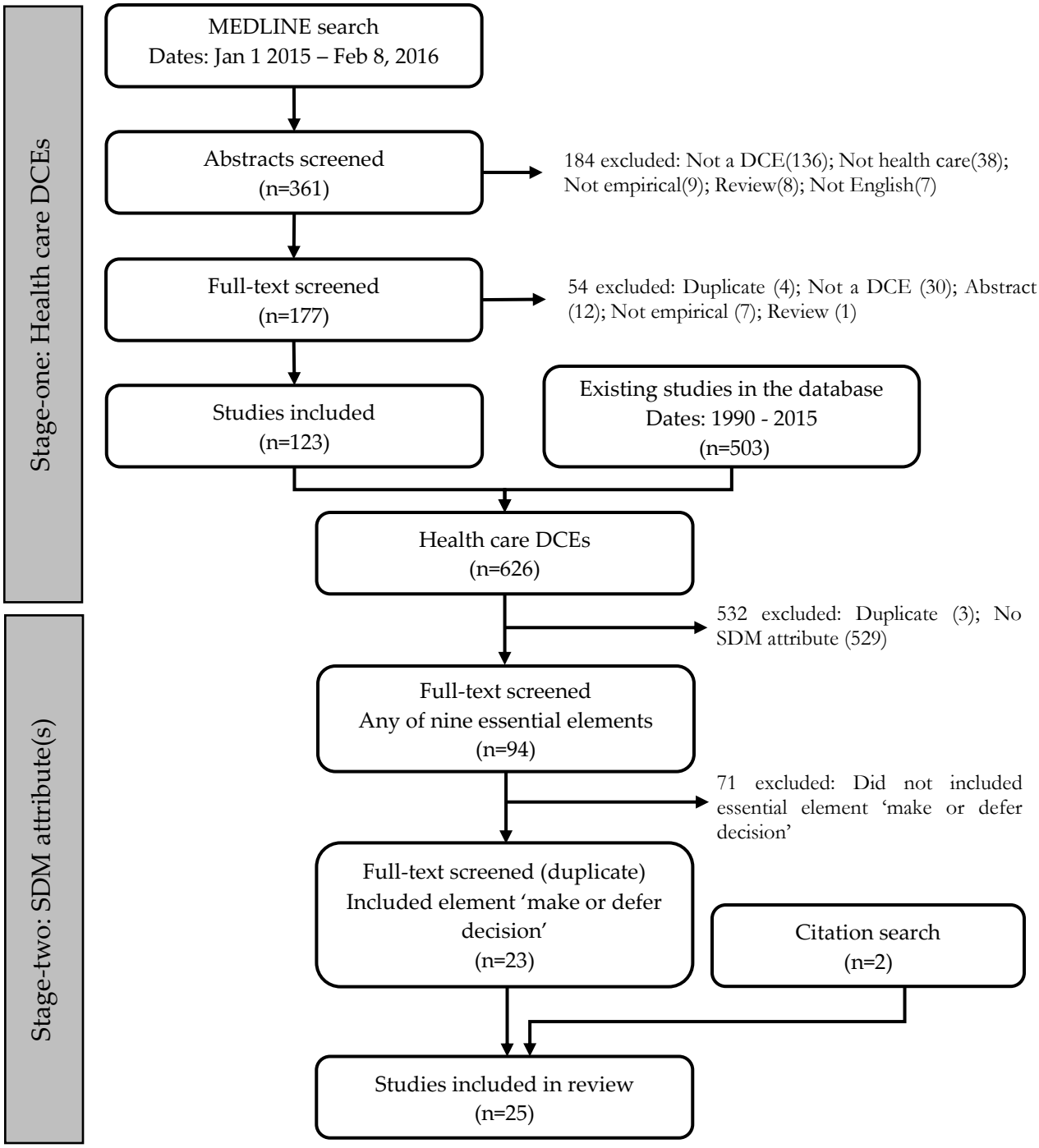
The value of SDM can be defined in terms of what respondents are willing to sacrifice for it. In a DCE, it is assumed that respondents trade between included attributes, with the results indicating how much of one attribute (e.g., money) respondents would be willing to sacrifice to obtain another (e.g., SDM). The rate at which participants trade between attributes is known as the marginal rate of substitution (MRS). A MRS can be obtained between any included attributes. However, given the heterogeneity in attributes across studies, this analysis explored the value of SDM based on three common 'payment vehicles' used in DCEs: 1) money; 2) waiting time; and 3) health outcomes.[84] Studies with these payment vehicles had the MRS with included attributes extracted. In cases where the MRS was not reported it was calculated from model coefficients. If coefficients for payment vehicles were reported categorically, they were converted to be continuous by

assuming a linear relationship. In all cases, the value of attributes was calculated based on the range of included levels (e.g., most relative to least valued). Across studies, the aim was to summarize value by elements of SDM, to allow the value to be compared against each other (e.g., information vs. decision-making).

5.4 Results

A total of 25 DCEs were included in this review. A total of 361 papers were identified in through the electronic search. Following abstract and full text screening, 123 health care DCE studies were added to the 503 studies in the existing database of DCEs, resulting in a total of 626 studies. In stage two, these papers were reviewed in full-text by one study member (LT), who identified 94 that included attributes related to one of the nine essential elements of SDM. Review by two study team members (LT, NB) identified 23 papers that included an attribute related to the essential element of 'make or explicitly defer a decision.'[210–232] Two additional papers were identified during a citation search.[233,234] A PRISMA diagram that outlines the study selection process and the primary reasons for exclusion is provided in Figure 5.1.

Figure 5.1: PRISMA diagram



5.4.1 Study characteristics

Detailed information on the included 25 studies is available in Table 5.2. Over half of the studies came from three countries, with seven from the Netherlands,[215,216,222,226,229,230,233] four from England,[212,221,227,228] and three from Australia.[223,225,227] Studies were divided between a generic context (e.g., health system, primary care, hospital care) and a specific clinical decision (e.g., maternity care). No included studies valued SDM in the context of advanced OA, which is the focus of this dissertation. With respect to the choice question, all DCEs were generic and the majority did not include the option to opt out (92%, n=23). Over half of included studies had five or six attributes (56%, n=14) and the most common method to identify attributes and levels was a literature review. Experimental designs were primarily fractional factorial and aimed for orthogonality. The majority elicited patient preferences (76%, n=19), with a minority eliciting provider (28%, n=7) and societal (8%, n=2) preferences. With respect to quality, all studies scored three or four out of five on the PREFS checklist (mean 3.44), though quality varied across categories. For instance, all studies described the purpose of the study, explained the preference assessment methods, and undertook significance testing, but just three included studies compared characteristics of responders and non-responders, and eight clearly stated that all responders were included in the analysis, or provided evidence of how exclusions impacted the results.

Table 5.2: Characteristics of included studies

Author (year), Country	Sample: Survey	Context	Choice Question	Choice Set	Design	Analysis
Akkazevia et al. (2006), Hungary [210]	Patients: Mail survey (n=19), 53% RR; rheumatology clinic survey (n=49) 98% RR.	Health system	Which [proposed healthcare system], (A) or (B) do you prefer to have?	Generic choice set with no opt out.	Main effects, balanced, minimum orthogonal array with a fold-over design and eight choice sets.	Three models: simple OLS, random effects OLS, mixed effects OLS
Berhane and Enquselassie (2015), Ethiopia [211]	Patients: Survey at 9 public hospitals (n=1,054) 95% RR.	Hospital care	Which hospital do you prefer?	Generic choice set with no opt out.	Fractional factorial design with orthogonal main effects using SPSS, with checks for level-balance and minimal overlap. 16 sets were blocked into 2 sets of eight choices.	Random effects probit model, with no exclusions reported.
Cheraghi-Sohi et al. (2008), England [212]	Patients: Six family practices (n= 1,193), 53% RR.	Primary care	Given this medical scenario, if you were offered options A and B which one would you choose?	Generic choice set with no opt out.	Fractional factorial using D-optimality criterion, sixteen choice sets using CHOICEFF SAS macro, 16 sets were blocked to 2 sets of eight choices.	Random effects probit model, with analysis for nonresponse bias, and no exclusions in the main analysis.
Davison et al. (2010), Canada [213]	Patients/providers: Mail survey (n=351), 75% RR.	Management of chronic kidney disease, procurement and allocation of organs for transplantation, end of life care discussions and decision-making	Which program do you prefer?	Generic choice set with no opt out.	Orthogonal main effects designs, resulting in 48 choice sets which were blocked into four versions with 12-questions	Fixed effects multinomial logit regression

Author (year), Country	Sample: Survey	Context	Choice Question	Choice Set	Design	Analysis
Gidman et al. (2007), United Kingdom [214]	Family members/providers: Postal survey with parents of children who had been admitted to hospital under a surgical consultant (n=280), 29% RR, and anaesthesiologists (n=193), 54% RR.	Daycase surgery	Which service is preferable?	Generic choice set with no opt out.	SPEED v 2.1 was used, taking into account d-efficiency criteria to create eight pairwise choices	Random effects probit model
Groenewould et al. (2015), Netherlands [215]	Patients: Internet panel with depression sufferers (n= 368), 11% RR.	Primary care	Based on this information, I would choose provider…	Generic choice set with no opt out.	Orthogonal (main effects plan), fractional factorial design consisting of 27 choice sets which were blocked into three sets of nine scenarios.	Conditional logit model
Hendrix et al. (2010), Netherlands [216]	Patients/family members: Postal questionnaire for pregnant women (n =321), 77% RR, and their partners (n = 212), 73% RR, from randomly selected Midwifery practices (of 150, 100 agreed to participate).	Maternity care	Which profile do you prefer?	Generic choice set with no opt out.	Orthogonal main-effects fractional factorial design with eight profiles which formed seven choice sets with each choice-set containing a 'base' profile and an alternative.	Random effects binary probit
Hundley et al. (2001), Scotland [217]	Patients: In-person questionnaire with low-risk pregnant women (n=301), with a 40% response rate.	Maternity care	Which unit would you prefer?	Generic choice set with no opt out.	Speed 2.1 was used to reduce the number of scenarios to a manageable level	Random effects probit model

Author (year), Country	Sample: Survey	Context	Choice Question	Choice Set	Design	Analysis
Huppelschoten et al. (2014), Netherlands [233]	Patients/partners: postal questionnaire (n=540), 55% RR; Health insurers: postal questionnaire (n=45), 54% RR.	Fertility care	I would choose this clinic:	Generic choice set with no opt out.	Orthogonal main/interaction effects fractional factorial design 81 choice sets (patients: 5 versions with 16 or 17 choice sets; insurers: 4 versions with 20 or 21 choice sets).	Generalized estimating equations logistic regression model
Kessels et al, 2015, Canada, Europe, Oceania, United States [218]	Providers/policy makers: (n=547), 27% RR.	Health system	Which situation would you prefer as a change to your current healthcare system performance due to payment system effects?	Generic choice set with no opt out.	Bayesian D-optimal design with 54 choice sets blocked into 3 surveys with 18 choice sets with two profiles	Multinomial logit
Krucien et al. (2015), France [219]	Patients: Hospital survey for people with multiple chronic conditions (n=150), a 94% RR.	Primary care	Would you accept this GP care?	Generic choice with one set of attributes/levels.	Attributes were divided into two blocks using attribute block design (ABD) with each block containing two common attributes, with eight tasks per block (1 repeated to check consistency of choices).	Binary logit model account for multiple responses per individual
Longo et al. (2006), Wales [220]	Patients: Survey administered at twenty general practices (n=747), 78% RR.	Primary care	What kind of visit would you prefer?	Generic choice set with no opt out.	Fractional factorial design with 27 scenarios.	Multilevel logistic regression model

Author (year), Country	Sample: Survey	Context	Choice Question	Choice Set	Design	Analysis
Longworth et al. (2001), England [221]	Patients: Survey of women who planned a home birth (n=118) or hospital birth (n=139) at one of two maternity units, 55% RR.	Maternity care	Unclear	Generic choice set with no opt out.	Fractional factorial design with 16 scenarios (Speed 2.1) which were formulated into 4 surveys (with eight questions). One of the scenarios was a constant comparator.	Random effects probit model
Muhlbacher et al. (2016), United States [234]	Patients: Survey of members of Duke University Health System (n=3,900), unknown RR.	Health system	Which system would you choose?	Generic choice set with no opt out.	D-optimal fractional factorial design with 15 choice sets (1 repeated) and 10 versions.	Random effects logit model
Pavlova et al. (2009), Netherlands [222]	Patients: Survey of nulliparous, pregnant, women attending a consultation at a midwifery practice (n=78), 98% RR.	Maternity care	I prefer:	Generic choice set with no opt out.	Orthogonal main-effect fractional factorial which resulted in eight scenarios. One of the eight scenarios was a constant comparison.	Random effects probit model
Peacock et al. (2006), Australia [223]	Patients: Survey of Ashkenazi Jewish women who provided a blood sample for research (n=209), 76% RR.	Genetic counselling for cancer	Which appointment would you prefer?	Generic choice set with no opt out.	SPEED computer package was used to select the optimal number of scenarios.	Random effects probit model.
Rischatsch and Zweifel (2013), Switzerland [224]	Providers: Survey of ambulatory care physicians (n=1,088), 11% RR.	Health system	I am willing to sign the MC contract with these obligations; I would like to remain independent without obligations	Generic profile with option to sign contract, or remain independent.	D-optimal design to generate 40 scenarios which were blocked into 4 surveys with 10 scenarios each – each scenario was compared to non-managed care (constant comparator)	Random coefficient logit model

Author (year), Country	Sample: Survey	Context	Choice Question	Choice Set	Design	Analysis
Salkeld et al. (2005), Australia [225]	Patients: Survey administered to colorectal cancer patients at two teaching hospitals (n=103), with an unknown response rate.	Colorectal cancer care		Generic choice set with no opt out.	Fractional factorial design resulting in 18 pairwise choices	Probit model
Schellings et al. (2012), Netherlands [226]	Providers: Web-based survey administered to employees who were involved in the inspection of mental health care services (n=25), 76% RR.	Psychiatric care	Which hospital do you choose?	Generic choice set with no opt out.	Six attributes, each with two levels resulted in 10 choice-sets	Logistic regression model
Scuffham et al. (2010), Britain and Australia [227]	Public: Survey administered to convenience sample at two universities, on in the UK and one in Australia (n=100), unknown RR.	Health system	Which health system is most preferred?	Generic choice set with no opt out.	Fractional factorial design using an orthogonal main effects plain, resulting in 27 choice sets	Mixed logit model
Tinelli et al. (2015), Germany, Slovenia, England [228]	Patients: Survey administered at nine general practices (n=692), 75% RR.	Primary care	Which situation would you choose?	Generic choice set with 'current practice' opt out.	Fractional factorial design with 16 choice sets using D-optimality criterion, four questionnaires with 5 choices each (1 dominant as an internal validity check)	Multinomial conditional logit model
van Haaren-ten Haken et al. (2014), Netherlands [229]	Patients: Survey administered at Midwifery practices to low-risk nulliparous women at 16 weeks gestation (n=562), 78% RR.	Maternity care	Which scenario do you prefer?	Generic choice set with no opt out.	Orthogonal main-effect fractional factorial design (with a check for orthogonality and level balance), seven scenarios all compared against one basic scenario	Random effects binary probit

Author (year), Country	Sample: Survey	Context	Choice Question	Choice Set	Design	Analysis
van Helvoort-Postulart et al. (2008), Netherlands [230]	Providers: Survey administered to anesthesiologists, surgical oncologists and breast-care nurses (n= 174), 10% RR.	Breast cancer surgery	Which circumstances would you choose?	Generic choice set with opt out.	Orthogonal, main effect, fractional factorial, foldover design	Random effects logit model
Vick et al. (1998), Scotland [231]	Patients: Survey administered as a general practice (n=101), 63% RR.	Primary care	Which kind of visit would you prefer?	Generic choice set with no opt out.	Factional factorial design with two-way interaction terms, 26 choice sets with 2 unique questionnaires (13 per)	Random effects probit model
Watson et al. (2012), Scotland [232]	Citizens: Survey administered to members of the public (n=68), unknown RR.	Health system	Which service do you prefer?	Generic choice set with no opt out.	Fractional factorial design with orthogonal main effects plan	Random effects probit model

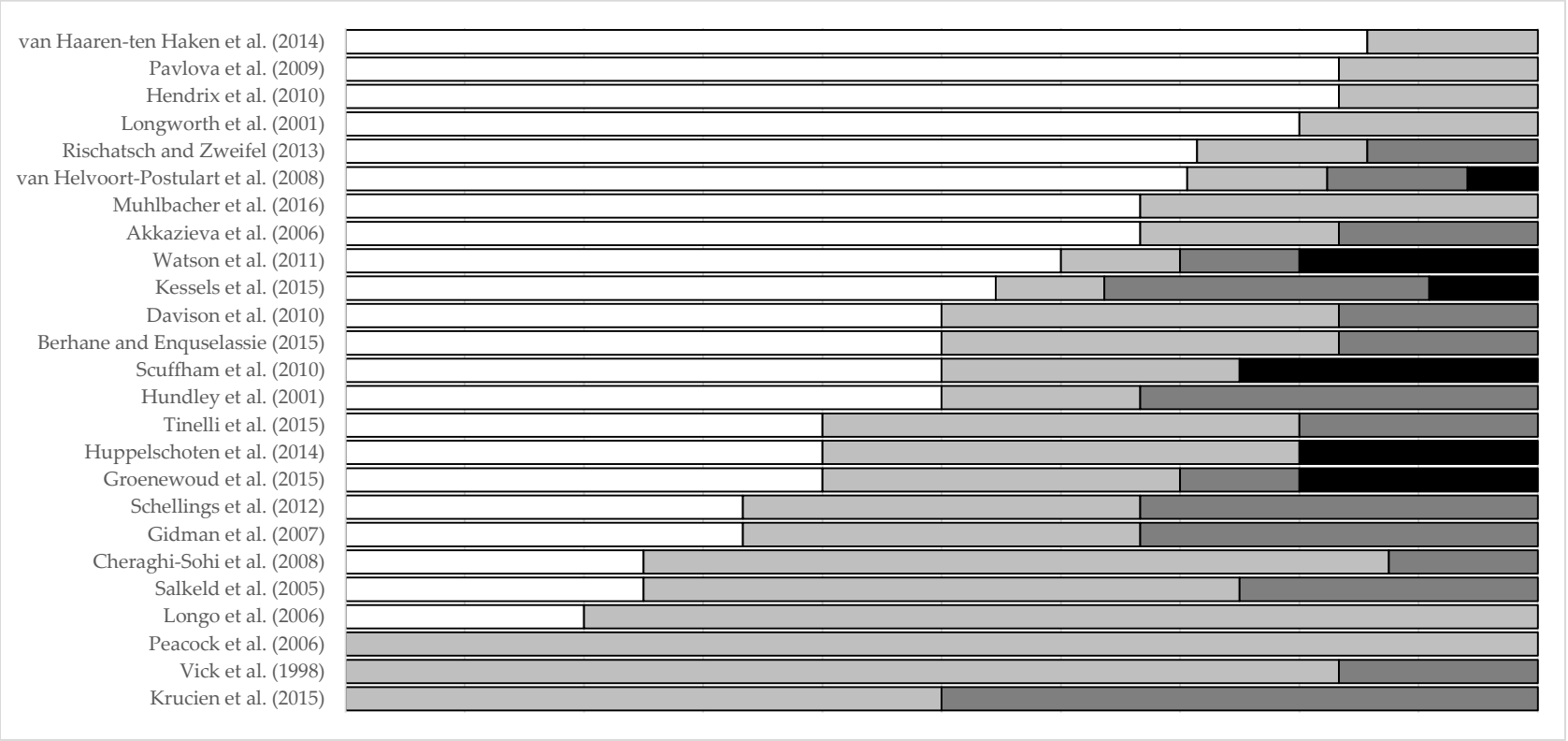
RR=Response Rate

5.4.2 Attribute classification

There was a total of 176 attributes across all 25 included studies. According to the Donabedian Model, attributes were predominantly related to health care structures (49%, n=86) and processes (46%, n=81). In all 25 studies, a total of nine attributes (5%) were related to outcomes. Of these nine attributes, seven reflected health outcomes and two non-health outcomes. The proportion of attributes by study, as classified by the Donabedian model, is presented in Figure 5.2.

In total, 55 attributes were classified as interpersonal processes, of which 51 included one or more of the nine essential elements of SDM. There was heterogeneity in how SDM was characterized, with studies using between one and five attributes to describe SDM and covering between one and six of the nine essential elements (see Table 5.3). Eligibility criteria required that the element 'make or explicitly defer a decision' be present, however, each of the other 8 elements were present in less than a third of the studies, with no studies including the element 'arrange follow-up' (Figure 5.3). Example descriptions for each of the nine essential elements covered by included studies are presented in Table 5.4.

Figure 5.2: Proportion of attributes as classified by the Donabedian Model



Structures
The attributes of the settings in which care occurs, including material resources, human resources, and organizational structure.

Interpersonal Processes
The patient communicates preferences and the information necessary for arriving at a diagnosis, and the physician provides information about the nature of illness and its management (e.g. shared decision-making).

Technical Processes
The skill in implementing the appropriate strategies of care. The goodness of technical performance is judged in comparison with best practice.

Outcomes
The effects of care on the health status of patients and populations.

Table 5.3: Elements of SDM present in included studies

	Akkazieva et al. (2006)	Berhane et al. (2015)	Cheraghi-Sohi et al. (2008)	Davison et al. (2010)	Gidman et al. (2007)	Groenewoud et al. (2015)	Hendrix et al. (2010)	Huppelschoten et al. (2014)	Kessels et al. (2015)	Krucien et al. (2015)	Longo et al. (2006)	Longworth et al. (2001)	Ratcliff and Longworth (2002)	Muhlbacher et al. (2016)	Pavlova et al. (2009)	Peacock et al. (2006)	Rischatsch and Zweifel (2013)	Sakeld et al. (2005)	Schellings et al. (2012)	Scuffham et al, (2010)	Tinelli et al. (2015)	van Haaren-ten Haken et al. (2014)	van Helvoort-Postulart et al. (2008)	Vick et al. (1998)	Watson et al. (2011)
Define/explain problem	✓	✓		✓							✓					✓			✓					✓	
Present options		✓									✓					✓			✓					✓	
Discuss pro/cons													✓			✓				✓	✓		✓		
Patient values/preferences			✓						✓	✓									✓		✓			✓	
Discuss patient ability/self efficacy										✓				✓		✓									
Doctor knowledge/recommendations			✓							✓						✓									
Check/clarify understanding											✓							✓						✓	
Make or defer decision	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Arrange follow-up																									

Figure 5.3: Proportion of studies that include essential SDM elements

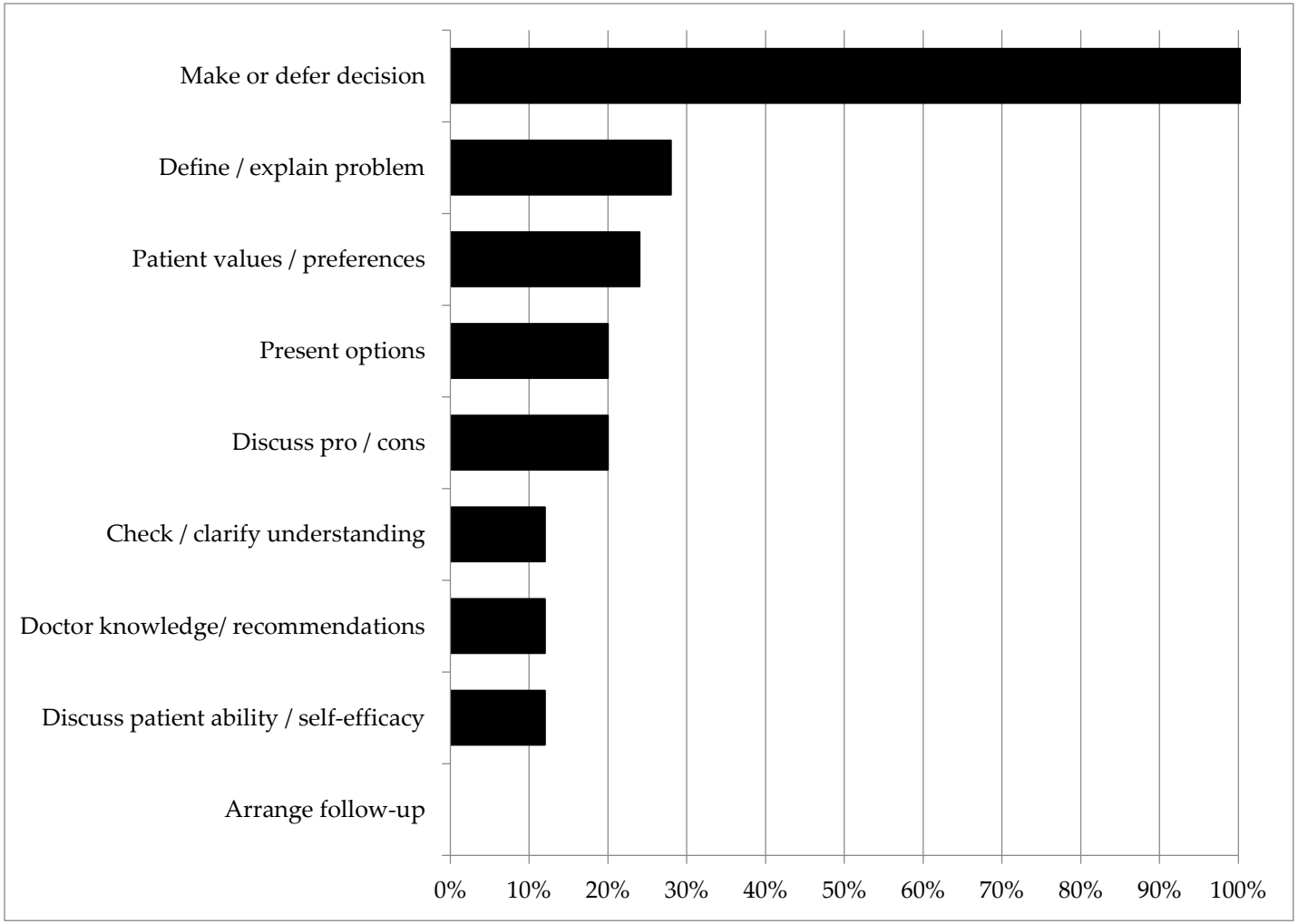


Table 5.4: Example descriptions of SDM elements from included studies

Define / explain problem	"Provides patient with information with understandable language about the illness, lab investigation, and treatment."[211]
Present options	"Information about the treatment of your health problem: The doctor gives you [little / a lot of] information."[231]
Patient values / preferences	"The degree to which care is respectful of and responsive to individual patient preferences and values, ensuring that patient preferences and values guide major clinical decisions."[218]
Discuss pros / cons	"Counsellors and doctors discuss the possible benefits and limitations of having a genetic test."[223]
Check / clarify understanding	"The doctor's words and explanations are [easy/difficult] to understand."[231]
Doctor knowledge / recommendations	"Counsellors and doctors talk to you about options for early detection of breast and ovarian cancer and provide recommendations for the frequency of mammography, ultrasound and other means of early detection appropriate for you."[223]
Discuss patient ability / self-efficacy	"The GP asks how the patient's daily life is modified by his chronic conditions and how he copes with them."[219]
Make or defer decision	"Treatment decision is made by you on the basis of sound medical advice and received information of your circumstance and health status."[210] "How should decisions to stop dialysis be made? [Personal decision / Shared decision with the medical team that combines personal preferences and medical facts]."[213] "Who finally decides your treatment? [Your surgeon alone / Your surgeon (after considering your opinion) / Shared (between you and your surgeon) / You (after considering your surgeon's opinion)]."[225]

5.4.3 Value of SDM

Eleven of the 25 studies valued SDM using either money, waiting time, or health outcomes, while the remaining 14 studies did not include one of these common payment vehicles. Value was summarized by payment vehicle, rather than by the elements of SDM for two reasons. First, as described in Section 5.4.1.2, eight of the nine essential elements of SDM appeared in less than a third of the included studies. Secondly, many attribute and level descriptions included more than one essential element, making it impossible to assess their value independently.

Seven studies estimated willingness-to-pay (WTP) by valuing SDM against money.[212,214–216,227,233,234] The value of SDM varied considerably depending on the context

(see Table 5.5). For instance, patients in the UK were willing to pay $12 to have a doctor that involved them in treatment decisions in primary care,[212] while citizens were willing to pay over £2,500 in additional tax revenue for a health system that had an 'adequate' level of patient choice in treatment decisions relative to 'none.'[227] Five studies estimated willingness-to-wait (WTW) by valuing SDM against waiting time.[211,212,215,227,228] The type of 'wait' varied across studies depending on the context. The types of wait included: minutes (in office wait for primary care consult); hours (hospital services); days (wait for primary care consult); weeks (specialist care consult); and months (surgical services) (see Table 5.6).

Five studies included an attribute which allowed SDM to be valued relative to health.[215,218,227,232,233] Groenewould et al. explored Dutch patients' preferences for choosing a health care provider in the context of treatment for chronic depression.[215] The study found that respondents valued a provider that allowed them to actively participate in decision-making (relative to hardly participate) as much as they valued a provider who had 80% of their patients report 'good results' (relative to 20%). Having preferences that matched with their provider (compared to there being no match) was valued the same amount. Huppelschoten et al. investigated Dutch patients' and insurers' preferences for choosing a clinic for fertility care.[233] Both patients and insurers valued the opportunity to 'always' be involved (compared to never) as much as a 5% (absolute) increase in the clinic's mean pregnancy rate. Kessels et al. investigated the value of potential health care

system reforms from the perspective of physicians, policy-makers, and executives in various countries, and found that maximizing health gain and minimizing harms was valued around three times as much as increasing the patient-centredness of care by respecting preferences and values.[218] Scuffham et al. invested the preferences of British and Australian citizens for health care system characteristics.[227] Adequate patient choice (compared to none) was valued as much as a three to five-year increase in life expectancy, and a 1 per 1000 decrease in infant mortality, and adequate patient information (relative to none) was valued as much as a three-year increase in life expectancy, and a 1 per 1000 decrease in infant mortality. Lastly, Watson et al. investigated societal preferences for health priority setting, and found that having patients share in treatment decisions (relative to 'health professionals making the decision') was valued half as much as having a 'large health gain to a large number of people' (relative to 'small health gain to a small number').[232]

Table 5.5: Marginal willingness-to-pay (WTP) for shared decision-making

Author	Context	SDM attributes: Description	Marginal WTP of SDM		Least Valued Attribute (WTP)	Most Valued Attribute (SDM)
Cheraghi-Sohi et al. (2008), England [212]	Primary care	Patient perspective: The doctor is interested in your own ideas about what is wrong (relative to 'not interested') Shared decision-making: The doctor involves you in decisions about treatment (relative to 'does not involve you') Physician's knowledge of the patient: The doctor has access to your medical notes and knows you well (relative to 'does not know you well')	$15 $12 $18		Shared decision-making ($12)	Physician's knowledge of the patient ($18)
Gidman et al. (2007), United Kingdom [214]	Child day surgery	Shared medical decision-making: Parents involved (relative to 'not involved)	£90 £65	(Provider) (Parents)	Doses of postoperative pain relief needed (£7) Doses of postoperative pain relief needed (£3)	Shared medical decision-making (£90) Immediate postoperative recovery (£91)
Groenewould et al. (2015), Netherlands [215]	Chronic depression	Participation: Client in control (relative to 'hardly any possibilities for participation') Vision on treatment: Vision matches with client (relative to 'no clear vision')	£58 £46		Expertise of clinician (£12)	Continuity of care (£70)
Hendrix et al. (2010), Netherlands [216]	Obstetric care	Possibility of influencing decision-making during birth: Possible (relative to 'Not possible')	£504 £347	(Patient) (Partner)	Possibility of pain relief treatment (£214) Type of birth setting (£ 156)	Possibility of influencing decision-making during birth (£504) Possibility of influencing decision-making during birth (£347)
Huppelschoten et al. (2014), Netherlands [233]	Fertility care	Patient involvement: Always (relative to 'never')	€389		Continuity of physicians (€287)	Clinic's mean ongoing pregnancy rate (€1,201)

Author	Context	SDM attributes: Description	Marginal WTP of SDM		Least Valued Attribute (WTP)	Most Valued Attribute (SDM)
Muhlbacher et al. (2016), United States [234]	Health system design	Shared decision making: …is when your care providers always involve you in treatment decisions by helping you find the best choice for your situation (relative to 'rarely or never together'). Proactive care: … providers contact you once every month to check on your progress, ask how you are doing and offer help and support (relative to 'once every 12 months').	$2,085 $869		Guidance within the facility ($454)	Shared decision making ($2,085)
Scuffham et al. (2010), Britain and Australia [227]	Health system design	Patient choice: The level of patient choice in their treatment decision. Adequate (relative to 'none') Patient information: 'Adequate' level of information available on the chosen treatment (relative to 'poor')	£2,389 £2,564 Not significant £1,819	(UK) (Australia) (UK) (Australia)	Patient information (£ns) Patient information (£1,819)	Infant mortality rate (£11,738) Infant mortality rate (£ 8,812)

ns: not statistically significant

Table 5.6: Marginal willingness-to-wait (WTW) for shared decision-making

Author	Context	SDM attribute: Description	Marginal WTW for SDM	Least Valued Attribute (WTW)	Most Valued Attribute (WTW)
Berhane and Enquselassie (2015), Ethiopia [211]	Hospital services	Physician communication: Likelihood that the physician has a friendly approach; provides patient information with understandable language about the illness, lab investigation, and treatment; reassures the patient; plus involves the patient in decisions. (relative to 'poor') Nursing communication: As above.	Not significant 2.7 hours	Diagnostic facilities (1.1 hours)	Drug availability in the hospital (3.3 hours)
Cheraghi-Sohi et al. (2008), England [212]	Primary care consultations	Patient perspective: The doctor is interested in your own ideas about what is wrong (relative to 'not interested') Shared decision-making: The doctor involves you in decisions about treatment (relative to 'does not involve you') Physician's knowledge of the patient: The doctor has access to your medical notes and knows you well (relative to 'does not know you well')	2.1 days 1.7 days 2.6 days	Shared decision-making (1.7 days)	Physician's knowledge of the patient (2.6 days)
Groenewould et al. (2015), Netherlands [215]	Chronic depression care	Participation: Client in control (relative to 'hardly any possibilities for participation') Vision on treatment: Vision matches with client (relative to 'no clear vision')	34.3 weeks 27.5 weeks	Expertise of clinician (7.3 weeks)	Continuity of care (41.2 weeks)

Author	Context	SDM attribute: Description	Marginal WTW for SDM	Least Valued Attribute (WTW)	Most Valued Attribute (WTW)
Scuffham et al. (2010), Britain and Australia [227]	Health system design	Patient choice: 'Adequate' level of patient choice in their treatment decision (relative to 'none') Patient information: 'Adequate' level of information available on the chosen treatment (relative to 'poor')	1.0 months (UK) 1.8 months (Australia) Not significant (UK) 1.0 months (Australia)	Patient information (Not significant) Patient information (1.0 months)	Infant mortality rate (5.4 months) Infant mortality rate (4.9 months)
Tinelli et al. (2015), Germany, Slovenia, England [228]	Primary care consultations	Information: 'Always' being able to receive all the information you want from the GP on your care (e.g., treatment, tests, test results, and referral to hospital) (relative to 'rarely') Listened to and involved in decision-making: 'Always' being listened to and involved in decision making about your care with the GP (relative to 'rarely')	42.2 minutes 35.1 minutes	Booking time (27.8 minutes)	Best care (43.2 minutes)

5.5 Discussion

This systematic review found 25 studies that have valued SDM using a DCE. Overall, there is evidence that respondents (primarily patients) were willing to pay more for greater SDM, though this value varies dramatically depending on the context. Within a publicly funded health care system, asking patients to pay for greater SDM is an unlikely policy option. However, this evidence could be used to assess the value of paying for additional consultations from a health system perspective. This review also found evidence that respondents would be willing to wait longer for greater SDM. One of most widely cited barriers to SDM is the perception that undergoing SDM takes too long.[7] Evidence does suggest that using tools to support SDM may increase the length of consultation,[13] and the implication of this finding may be that there are fewer consultations and thus longer waiting times. Many health systems collect and publicly report measures of patient outcomes, provider performance (e.g., waiting time), and patient experiences.[235] The evidence here suggested that respondents would be willing to consider waiting longer to see a provider that engages in greater SDM, however as of yet this data is not routinely reported, particularly at the physician level.[235]

As discussed in Chapter Four, Canadian guidelines for economic evaluation state that "the value of non-health effects should be based on being traded off against health" and that this valuation should assess societal, rather than patient or provider preferences.[144] This review found two studies that valued SDM relative to health

outcomes using societal preferences.[227,232] In one study, respondents were willing to accept lower life expectancy and higher infant mortality for greater SDM within a health system,[227] and in the other, willing to accept a smaller health gain to fewer people for greater SDM.[232] Neither of these two studies valued SDM relative to life-years which would allow the value to be incorporated within the QALY. In the context of the case study of this dissertation, no studies have valued SDM in the context of advanced OA.

There was substantial heterogeneity in how SDM was defined, both in terms of the number of attributes used to describe SDM, and the essential elements included in attribute and level descriptions. This heterogeneity likely reflected the different research questions stated in each of the included studies. For instance, some studies were interested in the relative value of different elements of SDM, such as involvement in decision-making and level of information, whereas others aimed to value them together. Regardless of the approach, our review highlights that most studies defined SDM using few of the nine essential elements, as defined by Makoul et al.[6]

By classifying included attributes according to the Donabedian model [156] our analysis highlights that SDM was largely valued relative to health care structures or other processes, such as waiting time, money, or other interpersonal characteristics. Only five included studies valued SDM relative to health outcomes. For studies that did not include a health outcome attribute, it was unclear whether the value ascribed by respondents related to the potential value of SDM to improve health outcomes or indicated that the

process of SDM was valued independently. For example, when Longworth et al. found that woman preferred greater decision-making autonomy during intrapartum care, it was unclear whether women valued autonomy in and of itself, or perceived that it would translate into better outcomes for themselves or their baby.[221]

5.5.1 Limitations

There are several limitations of this review. First, the search was restricted to identify DCEs, though there are other valuation techniques that can be used to value healthcare processes including CV, CA, analytical hierarchical process, SG, and TTO and person-trade-off.[183] This decision was made to limit the methodological heterogeneity in our sample that would have made it more challenging to draw conclusions. The search was restricted to the MEDLINE database, and may have missed relevant publications that were indexed elsewhere. Of the 25 studies included in this review, two were identified though a citation search. Our eligibility criteria required that DCEs include the essential element 'make or explicitly defer a decision' within attribute and level descriptions. Screening full-text studies for attributes related to SDM was only completed by one reviewer and may have resulted in some studies being missed. Another limitation is that one reviewer conducted data extraction, though it was checked by a second. However, this is consistent with previously published systematic reviews in this area.[84,207] Several studies did not report a MRS between SDM and the payment vehicle. For several

studies, calculating a MRS required assuming that categorical attributes, such as waiting time or money, could be treated as linear (i.e., there was a constant proportional trade-off). This assumption may not hold in all circumstances. While the aim of this study was to summarize value according to the essential elements of SDM, this proved challenging because many elements appeared in very few studies, and some attributes included multiple elements of SDM which made it impossible to assess their relative value. Thus, the results summarized value by payment vehicle. Lastly, as with all stated preference studies, the trade-offs and subsequent valuations are based on hypothetical scenarios and thus may differ from choices observed in real scenarios.

5.5.2 Conclusion

This review demonstrates that that there is heterogeneity in how SDM has been defined and valued using DCEs. The evidence suggests that SDM is valued from a patient's perspective, however the value varies dramatically depending upon the context. Furthermore, in most cases it is unclear whether patients value the process of SDM or the potential for SDM to improve outcomes. When assessing these findings through the lens of published economic evaluation guidelines, there is limited evidence that members of society are willing to forego health for the benefits of SDM. This review found no studies that valued SDM in the context of treatment decision-making in advanced OA. Furthermore, no studies valued SDM in a manner that could be incorporated within the

QALY. In Chapter Six an empirical study is described that used a DCE to value SDM relative to life-years to allow this health state utility value to be incorporated in a CEA (such as the one described in Chapter Two).

6 Incorporating the value of the process of shared decision-making in knee osteoarthritis within the QALY: a discrete choice experiment

6.1 Introduction

This chapter builds on Chapters Four and Five. Chapter Four outlined how conventional CEA may fail to capture the value of SDM-interventions and discussed techniques to value SDM. Chapter Five systematically reviewed studies that have valued SDM using a DCE. This review found only two few studies have valued SDM relative to health outcomes using societal preferences, and none that have done so in a manner that could be incorporated with the QALY. The overarching aim of this chapter is to value the process of SDM in a manner that can be incorporated within the QALY. This chapter focuses on advanced knee OA, since in comparison to hip OA, SDM is likely to play a more important role. Specifically, three times as many patients undergoing TKA experience no clinical or statistically significant improvement compared to those undergoing THA.

6.2 Background

Guidelines for the design and conduct of economic evaluations to inform national resource allocation decisions recommend using CEA where effectiveness is measured using QALYs.[73,144] QALYs focus on health status and reflect a specific policy goal: to maximize population health.[75] As described in Chapter Four, there are concerns that CEA

using QALYs may be insufficient to evaluate the full impact of SDM-interventions, which may result in improved health outcomes but have additional aims, such as improving patient-provider communication and increasing patient knowledge and involvement in decision-making.[236]

Within a single-payer health care system, a decision to allocate resources to support SDM through the implementation of SDM-interventions requires sacrificing health benefit elsewhere.[182] As described in Chapter Four, given that health maximization is a clear objective of health care, health outcomes can serve as a common currency with which to value other consequences. Doing so would reflect the health opportunity costs and may facilitate comparisons with evidence generated in a CEA.

The objective of this chapter is to estimate the health state utility value of the process of SDM in the context of treatment decision-making for advanced knee osteoarthritis. The study uses a two-step 'chained' valuation technique, which includes a DCE, to value the process of SDM relative to life-years. The results of the DCE are presented, and then the process for estimating the health state utility value of SDM is described.

6.3 Methods

Our study was approved by the University of British Columbia Behavioural Research Ethics Board.

6.3.1 Valuing the process of shared decision-making using a two-step 'chained' approach

As described in Chapter Four, valuing consequences on the QALY scale requires eliciting the trade-off with either life years (TTO, DCE with duration) or risk of death (SG). The trade-off between quality-of-life (health outcomes) and length-of-life is realistic, however, this may not be the case for process or non-health consequences. For instance, asking respondents to trade length of life or a risk of instantaneous death for greater SDM may be unrealistic due to the difference in both magnitude and seriousness of the consequences.[184]

The 'chained' approach has been used to overcome challenges posed by standard techniques in valuing aspects of health and care that are either temporary or have a small value relative to health outcomes. This is accomplished by separating the valuation into two-steps. Step one estimates a MRS between the good being valued and an intermediate good, and step two estimates the value of the intermediate good relative to the payment vehicle of interest. For example, McNameee et al. used a chained approach to value palliative health care states (the good being valued) relative to an anchor health state (intermediate good), and then valued the anchor state relative to life-years (the payment

vehicle) using a TTO and SG.[88] This study used a chained approach to estimate the health state utility value of SDM. Step one estimated the MRS between SDM and health outcomes using a DCE, and step two estimated the MRS between health outcomes and life-years using previously published values.

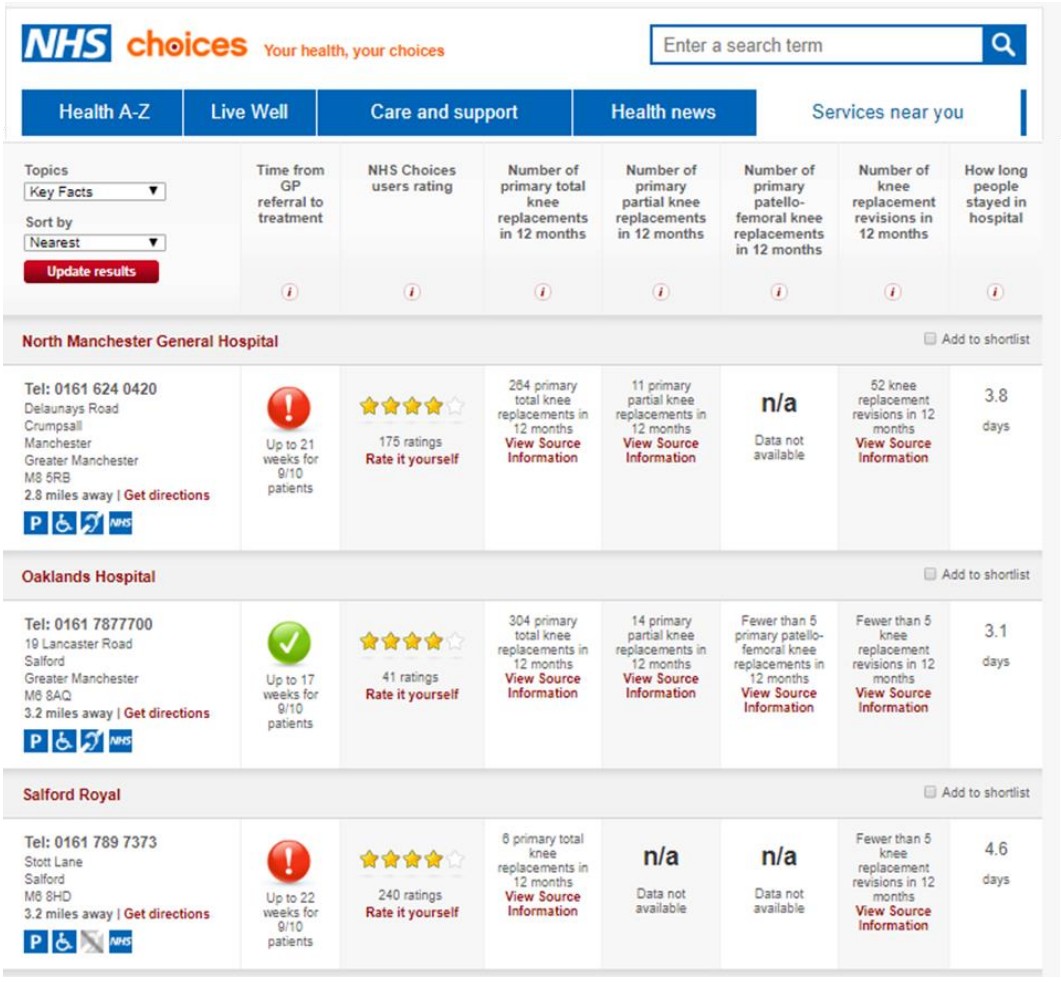
6.3.2 Step one: Estimating the marginal rate of substitution between shared decision-making and health outcomes

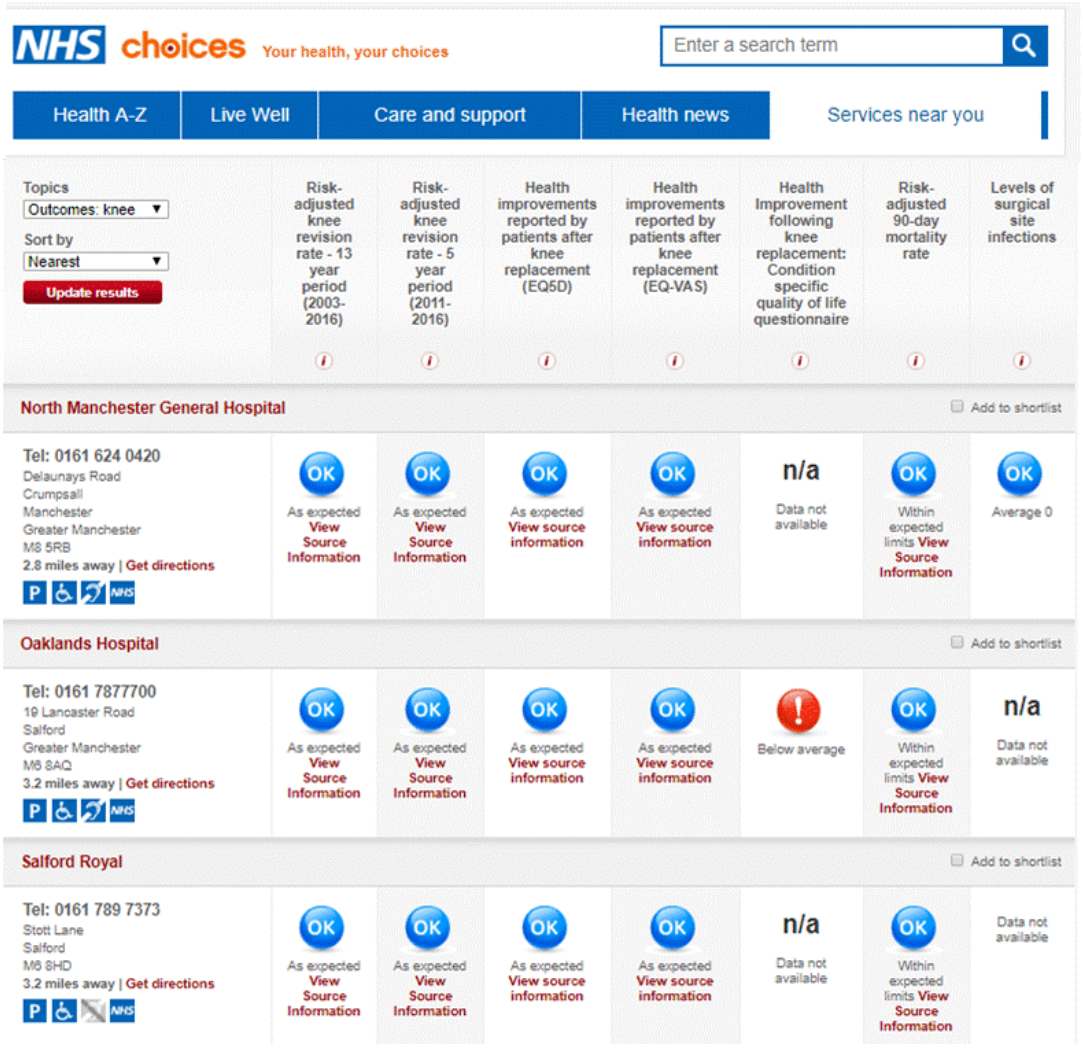
A DCE was chosen for step one. As described in Chapters Four and Five, DCEs are used widely to value health care processes,[208] but more importantly the DCE task can be designed to mimic a real decision faced by patients with advanced knee OA in choosing between care providers. Many health systems provide information to patients considering TJA to help them make an informed decision between providers.[237] For example, NHS Choices allows patients to choose between providers for TKA on the basis of waiting time, user rating, the number of surgeries performed, average length of stay, revision rate, health improvement based on the EQ-5D, mortality rate, and complications, and many other attributes (Figure 6.1). This information is comparable to the attribute and level format of a DCE.

DCEs are based on Lancaster's Theory,[238] which theorizes that goods are valued based on their attributes, and Random Utility Theory (RUT).[79] RUT is used to model choices between alternatives in a DCE, and assumes that respondents, when faced with

a DCE choice task, have some construct of utility for the different alternatives.[79] However, researchers are unable to observe all factors that contribute to utility. Therefore, utility can be separated into two components: 1) an explainable component specified as a function of the attributes (and levels) of the alternatives, and 2) a random (unexplainable) component that represents unmeasured variation in preferences.[79]

Figure 6.1: NHS Choices website for choosing a provider for knee arthroplasty





6.3.2.1 Choice question

Respondents were provided with information on OA, TJA, and non-surgical treatment options from a widely used patient decision aid.[239] Respondents were asked to imagine that they were diagnosed with OA and needed to choose between appointments with two arthritis specialists to choose whether to: have knee arthroplasty surgery, or use other treatments like exercise, weight loss, or medicines. Each choice set consisted of two

unlabelled alternatives ("Arthritis Specialist A" and "Arthritis Specialist B") with respondents required to choose one. The term "Arthritis Specialist" was chosen over "Surgeon" so that health outcomes were viewed as independent of the treatment chosen. Respondents were asked to choose between specialists, rather than hospitals, for two reasons. First, the process of SDM is dependent on the interpersonal manner of the individual physician. Second, this study aimed to elicit the trade-off between SDM and health outcomes and framing this as a choice between hospitals could introduce confounding factors (e.g., distance to provider).

6.3.2.2 Selection of attributes and levels

Attributes and levels were derived through a process consisting of: raw data collection; data reduction; removing inappropriate attributes; and wording.[240] Raw data collection used the systematic review of DCEs that included attributes related to SDM (see Chapter Five), hand searching of a health care DCE database to identify studies related to TKA,[215,241] reviewing studies that describe how patients choose health care providers,[242,243] and a gray literature search of public websites that present provider information to support this decision-making process (e.g., NHS Choices, ProPublica, BC Surgical Wait Times).

The data reduction stage aimed to reduce the number of candidate attributes. To achieve study aims, the DCE required attributes related to the process of SDM and health

outcomes. The CollaboRATE instrument was chosen to represent SDM [179,244] over other candidate measures (40) because it is a concise, valid, and reliable patient-reported outcome measure of SDM, designed for use following a consultation where a treatment decision was made. Levels were derived from the five-item response version to represent the range of SDM that could be experienced. Through discussions with team members, the EQ-5D descriptive system was chosen to represent health outcomes. This choice was made because the EQ-5D is routinely collected pre- and post-TKA and is reported to patients to inform provider choice in the context of TKA (NHS Choices). Of the 5-dimensions, 'pain or discomfort' was chosen to represent health outcomes because reducing pain or discomfort is considered the most important priority for individuals considering TKA.[245] Patients were asked to imagine they were experiencing either 'moderate' or 'extreme' pain or discomfort in their knee. The health outcome attribute was described as the proportion of patients who improved to have no pain or discomfort one-year following their appointment. Survey instructions stated that any improvement would be expected to 'last for at least two years.' Levels were derived from pre- and post-TKA EQ-5D data from the UK. Lastly, waiting time was chosen from the list of potential attributes, given evidence that wait time is important to patients.[246,247] Levels were derived based on wait times to see orthopedic specialists in British Columbia. Attributes and levels are presented in Table 6.1.

Table 6.1: Attributes and levels

Attribute	Description	Levels
Waiting time	This is how long you must wait to have an appointment with the arthritis specialist.	4 months (ref) 6 months 8 months
Shared decision-making	This is how much effort the arthritis specialist puts into: Helping you understand your health issues, Listening to the things that matter most to you about your health issues, and Including what matters most to you in choosing what to do next	No effort Some effort (ref) Every effort
Chance of improvement in pain or discomfort	This is the number of patients treated by the arthritis specialist who improve from moderate/extreme pain or discomfort to NO pain or discomfort 1 year after the appointment	50 out of 100 patients improve (50%) (ref) 60 out of 100 patients improve (60%) 70 out of 100 patients improve (70%)

6.3.2.3 Experimental design

Experimental design refers to "the process of generating specific combinations of attributes and levels that respondents evaluate in choice questions."[248] Designs can be 'full-factorial' meaning that they contain all possible combinations of attribute levels. However, full-factorial designs are generally large, and often require impractical sample sizes. For example, in the current context, a full-factorial, two-alternative design using three attributes, each with three levels results in 27 possible profiles (3^3) and 351 possible combinations of choice questions [3^3 x (3^3-1)/2]. As a result, most DCEs use 'fractional factorial' designs which use a fraction of the full-factorial design. This study used a fractional factorial D-efficient design which attempted to maintain level balance, orthogonality, minimal overlap, and utility balance where possible.[79]

Ngene v.1.1.2 was used to produce a D-efficient fractional factorial orthogonal design that considered main and interaction effects (between coefficients and pain scenario). No priors were specified, and the resulting design included 36 choice sets. Given time constraints for this component of the survey, choice sets were blocked into four versions with nine questions. A reliability check was added, resulting in ten choice sets per respondent: five where they were asked to imagine they were experiencing 'moderate' pain, and five with 'extreme' pain. The five scenarios always appeared together (e.g., five moderate scenarios followed by five extreme scenarios) though the order was random.

6.3.2.4 Pilot survey

A pilot DCE survey was developed. The survey included two sections. Section one began by asking the EQ-5D-3L,[249] followed by the choice questions. Section two asked respondents the Control Preferences Scale to measure preferred level of involvement in medical decision-making,[250] self-reported questions about whether the respondent, or their friends or family members, had been diagnosed with arthritis or undergone hip or knee arthroplasty, and demographic questions, included age, sex, and level of education.

The pilot study was conducted in two parts: online panel and 'think-aloud' interviews.[251,252] Consent documents are presented in Appendix 6.1. Pilot testing in an online panel sought to check the usability of the survey in different devices, operating

systems, and browsers, determine how long the survey took to complete, and to test whether respondents would trade across attributes and levels. Think-aloud interviews asked participants to verbalize their thoughts and decision-making process as they completed the survey. The aim of this stage of pilot testing was to assess the participants' understanding of the survey instructions, attribute descriptions, and the credibility of the hypothetical choice context. Pre-defined probes were used to determine whether respondents felt the attributes and levels were realistic, and to identify any factors (other than the attributes presented) that were influencing choices (Appendix 6.2).

In total, 151 individuals participated in pilot testing in an online panel. Results suggested that there were no usability issues, and that respondents were trading across levels. Eight English-speaking Canadians aged 60 and above participated in think-aloud interviews, lasting between 35 and 90 minutes. Suggestions were incorporated iteratively into future interviews. This process resulted in several changes to the DCE. For example, the waiting time attribute was changed from 'weeks,' which is routinely reported by health systems, to 'months' which was more intuitive for respondents. SDM, which was originally described using two separate attributes ('communication' and 'decision-making'), was changed to one attribute ('shared decision-making'). This change was made because respondents described focusing on the decision-making attribute as a proxy for both SDM attributes. In several cases respondents failed to recall the level of pain or discomfort they were asked to imagine experiencing. At their suggestion, 'pain

you are experiencing' was added to the choice sets. Additional changes suggested by respondents were to simplify attribute and level descriptions and survey instructions. For example, the term 'consultation' was changed to 'appointment' throughout. An example choice set is presented in Figure 6.2. The final online survey is presented in Appendix 6.3.

Figure 6.2: Example DCE question

Which arthritis specialist (A or B) would you choose for your appointment to make a treatment decision?

	Arthritis Specialist A	Arthritis Specialist B
Pain while you wait *Please imagine that you are experiencing -*	Moderate pain	Moderate pain
Waiting time *How long you must wait for an appointment.*	4 months	6 months
Shared decision-making *Effort made by the specialist to:* *-help you understand your health issues* *-listen to the things that matter most to you* *-include what matters most to you in choosing what to do next*	Some effort	Every effort
Chance of improvement in pain or discomfort *Number of patients who improve to NO pain or discomfort 1 year after the appointment*	50 out of 100 patients improve (50%)	60 out of 100 patients improve (60%)
I would choose:	○	○

6.3.2.5 Study sample and recruitment of respondents

Members of the Canadian population were recruited using an online market research panel. The sample was limited to English-speaking Canadians aged 60 years and above. This age restriction was applied to facilitate respondents placing themselves in the hypothetical choice context, which in this case involved a diagnosis of OA. The survey was anonymous online survey, which first asked respondents to provide online consent.

6.3.2.6 Analysis of data

The DCE data were analyzed using a random utility framework, where utility includes a systematic, observable component and a stochastic unobservable component. All data were analyzed in STATA (v14.2). The base analysis included all respondents except those who always chose the same alternative.

Analysis began with the conditional logit model (command *clogit*), which is the 'workhorse' model for analyzing DCE data. For the SDM attribute, 'some effort' was chosen as the reference level. Two attributes (wait and pain) were initially modelled as categorical, and subsequently modelled as linear if this assumption seemed reasonable. Previous research has demonstrated that the order of scenarios (e.g., moderate then extreme vs extreme then moderate) may impact model coefficients.[214] To determine whether order effects were present, interaction terms between model coefficients and an indicator for order, were included.

Given that this DCE included two alternative scenarios (moderate and extreme pain or discomfort), scale heterogeneity was explored using a heteroskedastic conditional logit model (command *clogithet*). In this context, exploring scale heterogeneity is important because the variance in the error term between the two scenarios may be different, which could lead to erroneous conclusions when attempting to compare them.[253]

Lastly, one of the main limitations of the conditional logit model is its inability to account for preference heterogeneity. While preference heterogeneity is not considered in QALYs, it is useful to understand the extent to which preferences vary. Preference heterogeneity was explored using the mixed logit model (command *mixlogit*) and latent class logit model (command *lclogit*). Latent class analysis explored between two and eight classes and the optimal number of classes was chosen based on two criteria. The first criteria was the Bayesian information criterion (BIC), which is considered "a good indicator of class enumeration."[254] Given evidence BIC may overfit resulting in more classes than suggested by theory,(262) the second criteria focused on 'common sense' interpretation of class coefficients. All models included main effects and interactions for the extreme pain scenario. The MRS of substitution between SDM and health outcomes were estimated using the *wtp* command (STATA v14.2), with confidence intervals calculated using the delta method.[256]

6.3.2.7 Robustness of results

Robustness of results was assessed by evaluating the impact of excluding respondents who may not have understood the DCE or chose not to engage with the survey. Our base case analysis included all respondents except those who always chose the same alternative. Additional analysis evaluated the results in alternative samples. This involved excluding those with a) lexicographic preferences, and those with

lexicographic preferences who also b) failed the consistency test, and c) spent less than five or d) ten seconds per DCE question, on average.

6.3.3 Step two: Estimating the marginal rate of substitution between health outcomes and life-years

Step two of the chained approach involved obtaining a MRS between health outcomes and life-years. In the DCE, health outcomes were described by the EQ-5D descriptive system, as the potential improvement from 'moderate' or 'extreme' pain or discomfort to 'no' pain or discomfort. In Canada, there is a published population value set for the EQ-5D, which used the TTO to estimate the value of these health improvements relative to life-years.[249]

6.3.4 Estimating the societal health state utility value of SDM

There is no 'gold standard' method for calculating societal value using the chained approach. Some researchers calculate the mean for each link in the chain and multiply them to estimate the societal value. Others calculate value at the respondent level and estimate the societal value by taking the mean of the sample. Wright et al. have noted that calculating "the mean of total utility, rather than the mean of each link in the chain … may make the result less precise."[185]

The societal health state utility value of SDM was estimated by multiplying the results from step one (MRS between *SDM* and *the potential improvement in pain or discomfort* obtained by the conditional logit model from the DCE) by the results from step two (MRS between *pain or discomfort* and *life-years* obtained from Canadian value set). Societal health state utility values were calculated for both moderate and extreme scenarios, and for two levels of SDM: 'no effort' and 'every effort.' Given the MRS for 'every effort' to engage in SDM is expected to be positive, this would correspond to a utility gain, whereas 'no effort' to engage in SDM is expected to correspond to a disutility.

6.4 Results

6.4.1 Step one: Estimating the marginal rate of substitution between shared decision-making and health outcomes

A total of 1,509 respondents completed the online survey. Twenty-seven respondents were excluded for reporting being less than 60 years of age, and twenty-six respondents were excluded in the base case analysis because they always chose the same alternative. This resulted in a final sample of 1,456. Respondent characteristics are presented in Table 6.2. Most of the sample was between the ages of 60 and 69 years (70%). Nearly half of the sample (47%) self-reported a diagnosis of arthritis, and approximately 5% had undergone a THA or TKA. Over half the sample (59%) reported experiencing moderate pain or discomfort as measured by the EQ-5D-3L.

Table 6.2: Characteristics of respondents (n=1,456)

Age group, n (%)		
60-64	601	41%
65-69	421	29%
70-74	274	19%
75-79	122	8%
80+	38	3%
Gender, n (%)		
Male	692	48%
Female	764	52%
Education, n (%)		
8th grade or less	10	1%
Some high school, but did not graduate	75	5%
High school or high school equivalency certificate	368	25%
College, CEGEP or other non-university certificate or diploma	466	32%
Undergraduate degree or some university	354	24%
Post-graduate degree or professional designation	183	13%
Preference for involvement in decision-making, n (%)		
…to make the final treatment decision.	78	5%
…to make the final treatment decision after seriously considering my doctor's opinion.	588	40%
…that my doctor and I share responsibility for deciding which treatment is best.	629	43%
…that my doctor makes the final treatment decision, but seriously considers my opinion.	131	9%
…to leave all treatment decisions to my doctor.	30	2%
Has been diagnosed with arthritis, n (%)	691	47%
Friend/family member has been diagnosed with arthritis, n (%)	1,094	75%
Has had a THA or TKA, n (%)	80	5%
Friend/family member has had a THA or TKA, n (%)	976	67%
Has 'moderate' pain or discomfort, n (%)	856	59%

6.4.1.1 Analysis of data

All main effects attribute coefficients for the conditional logit model were significant and in the expected direction (Table 6.3, Model 1). Both waiting time coefficients were negative and significant, indicating that respondents preferred an arthritis specialist with a shorter waiting time. The SDM coefficient for 'no effort' was negative and significant, and the coefficient for 'every effort' was positive and significant, indicating that respondents preferred an arthritis specialist who put more effort into helping them understand their health issues, listening to what mattered most to them

about their health issues, and including this in choosing what to do next. The pain coefficients were positive and significant, indicating that respondents preferred an arthritis specialist who had a greater number of patients improve to no pain or discomfort one-year after the consultation. Interaction effects found that when asked to imagine experiencing extreme pain or discomfort, respondents had a significantly stronger preference for a shorter wait, and a significantly weaker preference for their potential improvement in pain or discomfort.

Analysis suggested that the order in which respondents saw scenarios (moderate then extreme and vice-versa) effected results. A total of seven of the twelve interaction terms were statistically significant at the p=0.05 level (Appendix 6.4, Table A.1). Main and interaction effects coefficients for the conditional logit model accounting for order effects are presented in Table 6.3, Model 2. Given the presence of order effects, all subsequent models included interaction terms to account for this. Given evidence that the assumption of linearity for the wait and pain attribute was reasonable (Appendix 6.5) wait and pain were modelled linearly in all subsequent models.

The results of the conditional logit model with waiting and pain modelled linearly are presented in Table 6.4 (Model 3). Testing for scale heterogeneity between moderate and extreme pain scenarios was conducted using a heteroskedastic conditional logit model. The results suggested that the error variance did not differ between moderate and extreme pain scenarios (Appendix 6.4, Table A.2).

Preference heterogeneity was explored by using the mixed logit model (Table 6.4, Model 4). Coefficient signs and significance were consistent with Model 3. In addition to mean coefficient estimates, the mixed logit model provides a SD for coefficients which quantifies preference heterogeneity. The results from the mixed logit model find that for several coefficients, there is substantial heterogeneity. Both the wait attribute and 'no effort' level of the SDM attribute had a larger SD than the mean estimate.

To further explore preference heterogeneity, a latent class analysis was performed. Based on BIC, the best model fit was observed with six classes (Appendix 6.4, Table A.3), however this only represented a marginal improvement over the four-class model. After reviewing model coefficients for interpretability, a four-class model was retained for further analysis (Table 6.5). Class 1 could be described as "Balanced – prioritize outcomes less with severity." Members of this class traded across all attributes, with main effect coefficients significant and in the expected direction but had a weaker preference for a specialist with better outcomes when asked to imagine experiencing extreme (compared with moderate) pain or discomfort. Class 2 could be described as "Prioritize outcomes." Members of this class had a strong preference for specialists who had a higher proportion of patients improve to have no pain or discomfort, as this was the only main effect coefficient that was statistically significant. Class 3 could be described as "Balanced – prioritize outcomes more with severity." Like Class 1, members of this class traded across all attributes, with all main effects coefficients statistically significant and in the expected

direction. However, when asked to imagine experiencing extreme pain or discomfort, members of this class had a stronger preference for specialists who had a greater proportion of patients improve to no pain or discomfort. Lastly, members of Class 4 could be described as "Prioritize SDM." Members of this class had a strong preference for SDM, with both main effects coefficients statistically significant and in the expected direction. Respondent characteristics by latent class are reported in the Appendix 6.4, Table A.4. There was no clear relationship between Class membership and demographic or clinical characteristics.

Table 6.3: Conditional logit models

	Model 1 Conditional Logit			Model 2 Conditional Logit Controlling for order effects*		
	β	**SE**	**p**	**β**	**SE**	**p**
Main effects						
Wait						
4 months	Ref			Ref		
6 months	-0.37	0.04	0.00	-0.22	0.06	0.00
8 months	-0.84	0.04	0.00	-0.67	0.06	0.00
Shared decision-making						
No effort	-0.58	0.04	0.00	-0.77	0.06	0.00
Some effort	Ref			Ref		
Every effort	0.36	0.05	0.00	0.36	0.06	0.00
Pain						
50%	Ref			Ref		
60%	0.92	0.04	0.00	0.82	0.06	0.00
70%	1.56	0.05	0.00	1.51	0.06	0.00
Interaction terms						
Wait						
4 months * Extreme	Ref			Ref		
6 months * Extreme	-0.31	0.06	0.00	-0.62	0.10	0.00
8 months * Extreme	-0.59	0.07	0.00	-1.13	0.12	0.00
Shared decision-making						
No effort * Extreme	0.06	0.06	0.28	0.34	0.09	0.00
Some effort * Extreme	Ref			Ref		
Every effort * Extreme	-0.07	0.06	0.29	-0.08	0.10	0.44
Pain						
50% * Extreme	Ref			Ref		
60% * Extreme	-0.12	0.06	0.05	-0.11	0.10	0.30
70% * Extreme	-0.28	0.07	0.00	-0.21	0.10	0.04
Log-likelihood	-7313			-7256		
AIC	14,650			14,559		
BIC	14,748			14,756		
Observations	26,208			26,208		

*model includes interaction terms between each parameter and a dummy indicator for order (not displayed)

Table 6.4: Conditional logit and mixed logit*

	Model 3 (Conditional Logit)			Model 4 (Mixed Logit)		
	β	SE	p	β	SE	p
Mean						
Main effects						
Wait (per month)	-0.16	0.01	0.00	-0.28	0.02	0.00
Shared decision-making						
No effort	-0.74	0.05	0.00	-1.07	0.08	0.00
Some effort	Ref			Ref		
Every effort	0.39	0.06	0.00	0.68	0.08	0.00
Pain (per %)	0.08	0.00	0.00	0.12	0.01	0.00
Interaction terms						
Wait (per month) * Extreme	-0.30	0.03	0.00	-0.48	0.04	0.00
Shared decision-making						
No effort * Extreme	0.31	0.08	0.00	0.42	0.12	0.00
Some effort * Extreme	Ref			Ref		
Every effort * Extreme	-0.10	0.09	0.28	-0.16	0.13	0.21
Pain (per %) * Extreme	-0.01	0.00	0.12	-0.02	0.01	0.00
Standard Deviation						
Main effects						
Wait (per month)				0.36	0.02	0.00
Shared decision-making						
No effort				1.18	0.06	0.00
Some effort				Ref		
Every effort				0.36	0.09	0.00
Pain (per %)				0.11	0.00	0.00
Interaction terms						
Wait (per month) * Extreme				0.20	0.05	0.00
Shared decision-making						
No effort * Extreme				0.15	0.09	0.08
Some effort * Extreme				Ref		
Every effort * Extreme				0.15	0.09	0.08
Pain (per %) * Extreme				0.00	0.01	0.14
Log-Likelihood	-7263			-6392		
AIC	14,559			12,828		
BIC	14,690			13,110		
Observations	26,208			26,208		

* models include dummy variables to control for order

Table 6.5: Regression coefficients for latent class model with four classes*

	Class 1 "Balanced - prioritize outcomes less with severity"			Class 2 "Prioritize outcomes"			Class 3 "Balanced - prioritize outcomes more with severity"			Class 4 "Prioritize SDM"		
Class Share	32%			25%			28%			15%		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Main effects												
Wait (per month)	-0.36	0.03	0.00	-0.13	0.12	0.30	-0.31	0.05	0.00	-0.01	0.08	0.86
Shared decision-making												
No effort	-0.44	0.10	0.00	0.30	0.33	0.37	-1.47	0.22	0.00	-2.23	0.37	0.00
Some effort	Ref			Ref			Ref			Ref		
Every effort	0.26	0.12	0.03	0.41	0.58	0.48	0.94	0.19	0.00	1.66	0.48	0.00
Pain (per %)	0.05	0.01	0.00	0.21	0.02	0.00	0.11	0.01	0.00	0.02	0.02	0.41
Interaction terms												
Wait (per month) * Extreme	-0.52	0.24	0.03	-0.30	0.13	0.03	-0.88	0.10	0.00	-0.27	0.09	0.00
Shared decision-making												
No effort * Extreme	0.91	0.31	0.00	-0.77	0.55	0.16	0.63	0.32	0.05	-0.11	0.58	0.85
Some effort * Extreme	Ref			Ref			Ref			Ref		
Every effort * Extreme	0.24	0.29	0.39	-0.47	0.91	0.60	-0.14	0.32	0.67	0.61	0.66	0.35
Pain (per %) * Extreme	-0.06	0.03	0.02	-0.03	0.03	0.29	0.09	0.02	0.00	0.09	0.05	0.04
Log-likelihood	-5977											
CAIC	12,510											
BIC	12,443											

* models include dummy variables to control for order

6.4.1.2 Robustness

The base case analysis included all respondents except those under 60 years of age, and those who always chose the same alternative. As a robustness check, the analysis was re-run using different samples, excluding respondents who displayed lexicographic preferences; failed the consistency check; and took less than five or ten seconds to complete the survey, on average, per DCE question. The results suggest that our findings are robust to these alternative criteria, with coefficients exhibiting the same magnitude and direction of effect when analysed using conditional logit (Appendix 6.4, Table A.5) and mixed logit models (Appendix 6.4, Table A.6).

6.4.1.3 Estimating the MRS between SDM and health

The MRS between SDM and health, was estimated using the results from conditional logit model that included interaction terms to account for order effects, and modelled wait and pain linearly (Table 6.4, Model 3). In the moderate pain scenario, the results suggest that respondents were willing to forego a 10% (95% CI: 8%-11%) chance of improvement in pain or discomfort to meet with a specialist who made 'some effort' to engage in SDM compared with 'no effort' and a 5% (95% CI: 4%-7%) chance of improvement in pain or discomfort to meet with a specialist who made 'every effort' to engage in SDM compared to 'some effort.'

In the extreme pain scenario, respondents were willing to forego a 6% (95% CI: 4%-8%) chance of improvement in pain or discomfort to meet with a specialist who made 'some effort' to engage in SDM compared to 'no effort' and a 4% (95% CI: 2%-6%) chance of improvement in pain or discomfort to meet with a specialist who made 'every effort' to engage in SDM compared to 'some effort.'

6.4.2 Step two: Estimating the marginal rate of substitution between health outcomes and life-years

The Canadian value set for the EQ-5D indicates that respondents were willing to accept approximately a 5 percent reduction in length-of-life (9.5 years vs. 10 years) to have no pain or discomfort, compared to moderate pain or discomfort. This corresponds to a health state utility value of 0.045.[249] The Canadian value set also found that respondents were willing to accept approximately a 30 percent reduction in length-of-life (7 years vs 10 years) to have no pain or discomfort, compared to extreme pain or discomfort. This corresponds to a health state utility value of 0.300.[249]

6.4.3 Estimating the societal health state utility value of SDM

For the moderate pain scenario, the societal health state utility value of 'no effort' relative to 'some effort' was estimated to be -0.005 (0.10 x 0.045), and for 'every effort' relative to 'some effort' is estimated to be 0.002 (0.05 x 0.045). In the extreme pain scenario, the societal health state utility value of 'no effort' relative to 'some effort' is estimated to

be -0.018 (0.06 x 0.300), and for 'every effort' relative to 'some effort' is estimated to be 0.012 (0.04 x 0.300).

6.5 Discussion

This study used a two-step chained valuation approach to estimate the health state utility value of SDM. In step one, a DCE completed by an online panel from the Canadian general population estimated the MRS between SDM and the potential improvement in pain or discomfort in the context of treatment decision-making for advanced knee OA. In step two, the Canadian population value set was used to estimate the MRS between the potential improvement in pain or discomfort and life years. Together, these two-steps were used to estimate the societal health state utility value of SDM. The health state utility value of a specialist making 'every effort' to engage in SDM (relative to 'some effort') was estimated to be between 0.002 and 0.012, whereas 'no effort' was equivalent to disutility between 0.005 and 0.018. While this health state utility value is small, this may be cost-effective given that the cost required to support the use of SDM-interventions may be small. For example, if the decision aid cost $100, then a gain of 0.01 QALYs is equivalent to $10,000/QALY gained, which is considered cost-effective.

A key finding was that the order in which respondents saw scenarios (moderate then extreme and vice versa) had a statistically significant impact on coefficient estimates. For instance, when wait time and pain were modelled as categorical, seven of the twelve interaction coefficients for order were statistically significant. The presence of order

effects in stated preference surveys has been noted elsewhere.[257] In an attempt to mitigate this issue, survey instructions included an 'advanced disclosure' that informed respondents that they would be asked to complete five choice sets in one scenario, followed by five in the other. Previous research in environmental economics has demonstrated that advanced disclosure failed to remove precedent-dependent order effects in a DCE exercise with two binary choice tasks.[258] To mitigate the influence of order effects on our results, dummy variables were included in all models.

An additional finding was that preference heterogeneity exists. While not of primary relevance in quantifying the health state utility value of SDM for incorporation within an economic evaluation, there are implications for practice. The latent class analysis found several classes of patients with distinct preferences. Some classes exhibited a strong preference for SDM (notably Class 4) while others did not. This preference was not associated with observable characteristics, thus highlighting the importance of the clinical encounter for determining the preferred level of SDM for each patient.

While no previous studies have valued SDM in a manner that can be incorporated within a CEA, several studies provide results that can be compared to these results. Damman et al. explored key quality attributes for TJA using a DCE.[241] While not exploring value in the context of treatment decision-making, this study does provide information on the trade-off between elements of SDM and outcomes. For instance, the DCE included

an attribute called "Conduct of physicians," which was meant to indicate "how the physicians and the nurse practitioners communicate with patients, for example, their politeness, careful listening, and clear explanations." It also included an attribute describing "Pain control," which was described as "how well pain is controlled, for example, whether all possible actions are performed to help the patient with his or her pain." Both attributes were described using a three-star system, with 2-star indicating average performance, and 1-star and 3-star indicating below- and above-average, respectively. The results indicated that pain control was valued about the same as physicians' conduct. Like the DCE described in this chapter, this valuation study chose pain to represent health outcomes, though the description focused on the effort taken to alleviate pain rather than the effectiveness of the treatment. In addition, unlike this valuation study, Damman et al did not value the elements of SDM in the context of treatment decision-making or estimate a value that could be incorporated within the QALY.

Brennan completed a similar valuation study that incorporated the value of physician communication within the QALY.[259] They investigated whether process utility existed depending upon the quality of consultation with between women and their physicians in the context of pelvic floor medicine. Their valuation approach involved a 'bolt-on' domain to the SF-6D (resulting in an SF-7D) that described the quality of consultation on a five-point scale ranging from 'very good' to 'very poor.' Aspects of this

consultation domain were described using four components, including: 'We had a good talk,' 'I felt reassured,' 'The clinician understood what was on my mind,' and 'I felt I was taken care of.' Like the valuation study described in this chapter, Brennan et al. considered the value of elements of SDM in the context of treatment decision-making and estimated this value in a manner that could be incorporated within the QALY, however this value was estimated using a non-trade-off-based method (the VAS) and was completed in a different context.

The results of this study can also be compared with the broader literature related to process utility. Brennan et al. reviewed systematically studies that have valued aspects of process utility for incorporation within the QALY, and found that utility ranged from 0.001 to 0.27 for drug delivery methods, and 0.0005 to 0.031 for screening and testing procedures.[159] In total, five studies included in the review used a chained approach. Birch et al. used a chained SG, finding that alternative management strategies for mildly abnormal Pap smears had utility values ranging from 0.017 to 0.031.[260] In the same clinical context, Howard et al. found that different management strategies for abnormal Pap smears had utility values ranging from 0.0005 to 0.03.[261] Boye et al. used a chained SG, and found that weekly instead of daily injections corresponded to a utility value of 0.023, and increased dosing flexibility around mealtimes was valued at 0.006.[262] Chancellor et al. used a chained TTO, and found inhaled compared to injected insulin had a utility value between 0.01 and 0.08 depending on the scenario.[263] Cook used a chained TTO,

finding that different treatments for gallstone disease had utility values from 0.001 to 0.045.[264] The values estimated above are come from a wide range of contexts and describe different aspects of the process of care. Despite this, they are of a similar order of magnitude to the health state utility value estimated in this chapter (range: 0.002 to 0.018) thereby providing a useful comparison.

While the value of process-based aspects of care may be small, incorporating them into CEA may have an important impact on the results. Brennan found that incorporating the value of consultation quality increased the probability that the intervention was cost-effective from 35% to 60% at a WTP threshold of £20,000 per QALY.[259] This result highlights that when considering the value of investments to improve the process of care, while the benefit may be small at the individual level, the costs may be as well. For instance, if the health state utility value of SDM estimated in this study was assumed to last for one year, an intervention that improved SDM from 'some effort' to 'every effort' in patients with moderate pain or discomfort at a cost of $100 would be cost-effective at a threshold of $50,000/QALY. An additional consideration is the prevalent population who could benefit from SDM-interventions. With over 60,000 Canadians undergoing TKA annually, a small benefit per individual may translate into large overall benefit, and interventions to support SDM may benefit from the relatively large economy of scale.

6.5.1 Incorporating the health state utility value of shared decision-making within the QALY

The health state utility value estimated in this study could be incorporated within the QALY by multiplying the value of SDM by the time over which this utility (or disutility) is present. For example, the value of SDM estimated here could be incorporated into the trial-based CEA from Chapter Two. However, this would require several assumptions. The trial did not measure the level of SDM experienced by participants, thus one assumption would be related to the baseline level of SDM in the usual care arm, and potential improvement in SDM for those exposed to the intervention. These estimates could come from the literature,[265] or be assumed to mirror decision quality, which was measured in the trial.

In addition, incorporating the value of SDM into the QALY, whether a trial- or model-based analysis, requires an assumption about the length of time for which to apply the utility. For instance, this health state utility value could be applied as a one-time benefit or applied over one or more years. In the current study, the chance of improvement in the pain or discomfort attribute description presented an expectation that any improvement would be maintained for 'at least two years.' As such, participants may have been trading-off health with this expectation, meaning that the value estimated implicitly accounts for the potential duration. If this were the case, the most appropriate way of account for this would be to either apply the value as a one-time benefit, or over

one-year. Alternative scenarios could be explored to see how this assumption impacts the cost-effectiveness results.

If it were assumed that decision quality was indicative of the level of SDM experienced in the CEA from Chapter Two, that all patients were experiencing 'moderate' pain or discomfort, and that those who made a quality decision experienced 'every effort' as measured by the CollaboRATE scale, compared to 'some effort' for those that did not, then this would translate into an incremental QALY gain of 0.0002 per patient (Appendix 6.6).

6.5.2 Limitations

There are several limitations that warrant consideration. First, this study quantified the value of SDM in the context of treatment decision-making for advanced knee OA, and the results are not generalizable to other contexts. Second, the SDM level 'some effort' was used as the reference, meaning that 'no effort' corresponded to a disutility and 'every effort' corresponded to a utility gain. Incorporating the value of SDM within the QALY as a utility gain has the potential to lead to a 'quality-adjustment' that is greater than 1, however this could be addressed by treating 'every effort' as the reference level and assigning a disutility for the other two levels. A key assumption of the chained valuation approach is that the intermediate good, in this case the potential improvement in pain or discomfort, is preferred to the item being valued (i.e. process of

SDM). There may be respondents who have a strong preference for greater SDM, where this assumption does not hold. In the DCE, respondents were told to only consider the potential improvement in pain or discomfort and assume other aspects of health were the same between specialists. It is possible that individuals were considering more than just the potential improvement in pain or discomfort, thereby leading to a conservative valuation. Moreover, this chained valuation approach relied on the integration of two separate valuation methods, each of which has different methodological underpinnings.[93]

6.5.3 Conclusion

This chapter aimed to quantify the value of SDM in the context of treatment decision-making in advanced knee OA in a manner that can be incorporated within the QALY. Following Canadian economic evaluation guidelines, it was demonstrated that respondents were willing to forego potential health improvements for greater SDM and estimated the health state utility value under different scenarios.

7 Discussion

This dissertation aimed to quantify the economic value of interventions to support SDM in health care. Toward this aim, this dissertation employed: a trial-based CEA; a resource utilization and cost analysis using linked administrative data; a systematic review; and a DCE. This chapter discusses the key findings of the dissertation and places them in the context of the wider literature, and strengths and limitations of this program of research. This chapter also presents the implications for practice and highlight options for future research that were beyond the scope of this dissertation.

7.1 Key findings

The first empirical study presented in this dissertation was a trial-based CEA of patient decision aids plus a surgeon preference report for patients considering TJA, compared to usual care, using QALYs as the measure of benefit (see Chapter Two). The trial included 344 participants, and the results suggested that, despite fewer patients undergoing TJA, the patient decision aids were dominant, resulting in lower costs and more QALYs than usual care. This analysis represents a novel contribution to the literature, as the two-year time horizon matches the longest follow-up period of any previous CEA of patient decision aids,[96] and it is the only economic evaluation of patient decision aids in the context of treatment decision-making for advanced OA. The time horizon for this analysis allowed some preliminary conclusions to be drawn about the relative cost-effectiveness of patient decision aids, however the decision about whether

to undergo TJA will have impact on health outcomes and costs beyond the two-year time horizon of an RCT.

Chapter Three aimed to address the time horizon issue, by assessing resource use and costs associated with patient decision aids plus a surgeon preference report, at seven-years follow-up. This study linked the trial and administrative data of 324 of the 344 RCT participants and found results that were consistent with the two-year time horizon of the trial: a non-statistically significant reduction in the risk of TJA and costs. This is the first study to provide evidence on the long-term impact of patient decision aids in this context. However, there are two limitations that suggest the long-term implications of patient decision aids in this context are still unclear. First, the original trial was not powered on risk of TJA and using administrative data did not provide a measure of health status. It was therefore unclear whether patients exposed to the patient decision aid and preference report continued to have improved health after the trial follow-up. However, this analysis provided preliminary evidence which may be used in the design of future research studies to be powered to measure long-term health status.

The subsequent chapters of this dissertation addressed the following question: *How do interventions to support SDM provide value?* Chapter 4 explored this issue in depth and argued that conventional CEA using QALYs may underestimate the value of SDM-interventions and discussed how the process of SDM could be valued and incorporated into an economic evaluation. Chapter Four described the concepts, rather than suggest

definitive answers on the best way to measure, value, and integrate the value of SDM-interventions into economic evaluations. This chapter explained that, to be consistent with economic evaluation guidelines, the value of the process of SDM should be quantified through the trade-off with health outcomes using societal preferences.

Chapter Five presented a systematic review of studies that have valued the process of SDM using a DCE. The review identified 25 DCEs and found that across studies, definitions of SDM vary widely, and that most had elicited patient preferences. There was evidence that respondents were willing to wait longer for SDM, pay for SDM, and in some cases, forego health for SDM. However, most studies did not include a health outcome attribute. As a result, it was unclear whether respondents value the process of SDM, or its potential to improve health outcomes. Notably, no studies have valued SDM in the context of treatment decision-making for advanced OA, few studies have valued SDM in a manner consistent with Canadian economic evaluation guidelines, and none have done so in a way that can be incorporated within the QALY.

Chapter Six reported an empirical study that aimed to value the process of SDM in the context of treatment decision-making for TKA in a manner that can be incorporated within the QALY. This study used a chained valuation approach that valued the process of SDM relative to health outcomes using a DCE and used population weights to ascertain the value of health outcomes relative to life years. The results provided evidence that suggested the public may be willing to forego health outcomes for greater SDM. In

addition, the results suggest that there is large heterogeneity in preferences. The health state utility value of SDM was quantified so that it could be incorporated within the QALY framework of evaluation. To incorporate the value of SDM into the QALY required assumptions about whether to treat the value as a utility gain or disutility, and the length of time over which to apply it.

7.2 Economic evaluations of shared decision-making interventions

This dissertation contributes to a small, but growing literature evaluating the economic implications of SDM-interventions. There were no prior CEA of patient decision aids for advanced OA before this dissertation, but there were some examples from other areas of elective surgery. The findings from the studies reported in Chapters Two and Three produced similar results to published examples in other areas; namely a reduction in the uptake of elective surgery resulting in lower incremental costs. A published trial-based CEA found that patient decision aids reduced the uptake of hysterectomy for menorrhagia in the UK (OR=0.60, 95% CI: 0.38-0.96) and lowered costs (mean difference, $1,184, 95% CI: $684-$2,110) but had no effect on health outcomes over the two-year follow-up.[97] Hollinghurst et al. found that patient decision aids resulted in a non-significant reduction in the uptake of elective caesarean section, that translated into lower increment costs compared to usual care at nine-months follow-up.[98] Health outcomes were not impacted but there was a significant reduction in decisional conflict in the decision aid arm.[98]

7.3 Valuing the process of shared decision-making

The systematic review reported in Chapter Four identified two DCEs that have valued SDM relative to health outcomes using societal preferences.[227,232] Both studies valued SDM in the context of health system design and found that respondents were willing to trade health outcomes for greater SDM. This DCE found that for a specific clinical decision, treatment decision-making for advanced knee OA, respondents are willing to make that trade-off. This evidence suggests that it, in this context, may be acceptable to divert resources from activities that improve health outcomes, such as medications or surgery, toward SDM-interventions.

There is a growing literature identifying cases where the health-related QALY may fail to capture aspects of health and health care that matter to patients. This has been raised for public health interventions and programs that straddle health and social care.[154,198] There are many methods available to address these concerns, ranging from small adjustments to radical departures from current methods. The approach taken in this dissertation to capture the value of SDM is best categorized as a small adjustment. The approach remained within a health-related QALY paradigm, choosing to maintain a focus on health outcomes, but also incorporate aspects of the process of care that are valued. This contrasts with others who have advocated for a broadening the evaluative space to include well-being rather than just on health, an example of which is the well-being-adjusted-life-year (WELBY).[194] In valuing the process of SDM, guidelines from

CADTH were followed which suggested valuing non-health consequences relative to health using societal preferences, and reflects the potential health outcomes foregone when investing in process or non-health benefits.[144] This study also chose to value the process of SDM in a manner that can be integrated within the QALY. While this is not the first study to do so, this is not common. For example, a 2013 systematic review by Brennan and Dixon found just 15 studies published between 1996 and 2012 that had valued aspects of health care processes in a manner than can be incorporated within the QALY.

The most comparable study to this dissertation, Brennan (2016), valued improved communication between patients and providers, and incorporated this value into the QALY.[259] Our approach differs in several important ways. The study reported in Chapter Six used a trade-off-based valuation technique (DCE) whereas Brennan did not (visual analogue scale). The motivation for using a DCE was to be consistent with CADTH guidelines, which recommend using trade-off-based techniques. Brennan valued the process of communication directly against six dimensions of the SF-6D. In contrast, in this dissertation SDM was valued indirectly, used a chained valuation approach, against one of the five dimensions of the EQ-5D. The choice to use a chained valuation approach, which has been used to value processes and temporary health states,[159,185] was due to concerns that placing processes and outlines alongside each other may result in processes being over-valued.

In current practice, economic evaluation is just one input (albeit a central one) into the health technology assessment (HTA) process. For instance, the CADTH Health Technology Expert Review Panel (HTERP) Deliberative Framework considers need, benefits, harms, patient preferences, implementation, legal and ethical implications, and environmental impact, alongside the economic impact which includes CEA and budget impact.[266] The Institute for Clinical and Economic Review from the United States considers long-term value for money, which includes comparative clinical effectiveness, estimated incremental cost-effectiveness, contextual considerations, and other benefits or disadvantages, and short-term affordability through the potential budget impact.[267] Within these, and other value frameworks and HTA processes, evidence of the added value of SDM-interventions could be captured and incorporated into resource allocation decisions alongside evidence from an economic evaluation. For example, the 'benefits' domain of CADTH's HTERP framework includes non-health benefits, such as patient autonomy and dignity.[266]

While mechanisms are in place to incorporate the value of SDM-interventions into investment decisions, the reality is that considering disaggregated consequences is complex, and it is unclear whether it leads to quality decisions. Baltussen and Niessen note that in many cases this leads to ad hoc decision-making, with policy-makers using intuitive or heuristic approaches to decision-making or acting out of political self-interest.[268] As a result, they conclude that "policy makers may not always be well placed

to make informed well-thought choices involving trade-offs of societal values" and advocate for a more rational and transparent approach.[268] Multi-criteria decision analysis (MCDA) is advocated as one such approach, and it is has been used to capture benefits beyond QALYs.[269] However there are several challenges to using MCDA. For instance, MCDA is prone to 'double-counting,' with one example being that 'cost' is one criterion often considered separately from 'cost-effectiveness.'[270] In addition, while MCDA does take a structured approach to eliciting weights for different pieces of evidence (e.g., cost-effectiveness, process or non-health benefits, equity), these weights often come from members of a HTA committee and may not reflect societal preferences.[270] Lastly, MCDA may not reflect economic value, in that assigning a weight to one consequence, such as process or non-health consequences, it is not explicit that this requires foregoing another, such as health outcomes.[270]

7.4 Strengths

This dissertation used conventional economic evaluation methods and a novel chained valuation approach to quantify the economic value of interventions that support SDM in the context of treatment decision-making for advanced OA. As detailed throughout this dissertation, many health systems are exploring the implementation SDM-interventions in this context, with the aim of containing rising costs and improving the quality of care. While previous evidence has suggested that SDM-interventions in this context may reduce uptake of TJA and costs, and may improve care, this is the first

analysis which have evaluated the costs and consequences simultaneously to assess whether they provide value.

While the two-year time horizon of the trial (reported in Chapter Two) equals the longest follow-up for an economic evaluation of a decision aid, linking the trial to administrative data (reported in Chapter Three) allowed exploration of a key issue identified in past studies: to determine whether those that choose not to have surgery over the short-term simply delay surgery or forego it altogether. While this analysis is not definitive, it used administrative date to provide the first evidence of the long-term impact of SDM-interventions in this context and suggests that they may have an impact on health resource use and costs for years.

The analysis in Chapter Six undertook rigorous methodological work to quantify the value of SDM-interventions that may not be captured in conventional CEA. Following Canadian economic evaluation guidelines, the process of SDM was valued using a chained approach in a sample of nearly 1,500 Canadians. Furthermore, this value was captured in a manner than can be incorporated within the QALY, which has not been done in the context of SDM-interventions.

7.5 Limitations

In considering the results program of research, several overarching limitations need to be considered.

There are undoubtedly economic implications beyond the seven-year time horizon of Chapter Three. One option was to develop an economic model with a lifetime time horizon, however this option was not pursued for two reasons. First, there was no evidence in Chapter Three that suggested that patient costs and outcomes had changed from the trial-based analysis, and second, there is a lack of data on the comparative costs and outcomes of individuals exposed to the intervention and comparator that would have made modelling challenging. Given the substantial time and effort required to complete this analysis, and the small likelihood that it would change the findings, it was not pursued.

The Informed Medical Decisions Foundation developed the patient decision aids evaluated in this trial, which are now owned and distributed by Health Dialoge Services Organization (www.healthdialoge.com). These patient decision aids are available, however the cost of providing them is unclear, as is their availability outside of the United States. There are other patient decision aids available for advanced OA, however their evidence base is not as robust. For example, the OPTION Grid for knee OA has been evaluated in a single step-wedge trial, which found a significant increase in knowledge and SDM, with no impact on the length of consultation, but the impact of the OPTION Grid on treatment choice, health outcomes, and cost is unclear.[271] In Australia ,there is a trial currently evaluating the impact of a DCE-based decision aid in this context.[272] In Canada, a pilot trial in Alberta is evaluating the impact of a patient-reported outcome

measure (PROM) based decision aid, and surgeon preference report on decision quality, SDM, and health outcomes.[273] These new patient decision aids are substantially different from the one studied in this dissertation, and as such, it cannot be assumed that they will have the same effectiveness and cost-effectiveness profile as those evaluated in the present trial.

One central aim of this dissertation was to quantify the value of the process of SDM in a manner that can be incorporated within the QALY. In Chapter Six, ways of incorporating the value of the process of SDM within the CEA from Chapter Two were discussed. There are several limitations that make this challenging. First, the trial did not measure the impact of patient decision aids on the process of SDM, but on the outcomes of the decision. Thus, while the results found that patient decision aids increased patient knowledge and resulted in a higher proportion of patients making a quality decision, this does not necessarily mean that more SDM occurred. Furthermore, incorporating the value of SDM within the CEA from Chapter Two would require assumptions about the level of SDM experienced by participants in both arms. One of the most important limitations in incorporating the value of SDM estimated in Chapter Six into the analysis from Chapter Two is that the trial reported in Chapter Two included individuals with both hip and knee OA, while the valuation study in Chapter Six was completed in the context of treatment decision-making for those with knee OA. The choice to focus the valuation study on the context of knee OA was made because individuals with knee OA

made up approximately 70% of individuals enrolled in the trial. However, the value of SDM may be different in the context of hip and knee OA.

Lastly, the conflict between the economic evidence presented in this thesis and the results of the clinical trial need to be considered. In real-world decision-making, policy-makers consider both the clinical and economic evidence when making resource allocation decisions. In this current context, there is a conflict between the clinical evidence, which finds no conclusive evidence that patient decisions aids result in a reduction in the uptake of TJA, and the economic evidence, which suggests a high degree of confidence that decision aids are cost-effective over the short-term. This conflict may lead policy makers to choose to wait for additional information, such as a clinical trial that is powered on the outcome of uptake of TJA. However, based on the evidence presented in Chapter Two, an economist would argue that, having considered the uncertainty, there is sufficient evidence to make a policy decision.[134] Chapter Three provided insight into the long-term implications of decision aids in this context and provides information that could be used to inform the design of a future trial, including the required sample size. Despite this, it is unclear whether current evidence is sufficient for policy-makers to decide, or whether additional information is required.

7.6 Implications for practice, health policy, and economic evaluations

This dissertation has several implications for practice. The evidence reported in Chapters Two and Three suggested that decision aids for patients considering TJA are a

cost-effective use of resources. Chapter One, Section 1.2.2.4 described how policies to encourage SDM often have two aims: cost containment and improving care.[15,54] The results in this dissertation suggest that, in this context, patient decision aids deliver on both aims. For example, the within trial CEA (see Chapter Two) suggests patient decision aids are less costly and provide more benefit. Furthermore, there was a high degree of confidence in these results. The argument that SDM-interventions in this context provide value is even more compelling when considering evidence generated in Chapter Six demonstrating that the process of SDM is valued independent of its impact on health outcomes. The analysis from Chapter Six suggests this value is modest, and if incorporated in the within trial economic evaluation from Chapter Two, would result in a dominant intervention achieving even greater benefit.

The trial-based CEA included a population with moderate-to-severe OA. It is possible that decision aids may be even more cost-effective in less severe patients, which may be more indicative of individuals seen in routine care, especially those with knee OA. As described in the Introduction, patients with milder disease are less likely to benefit from surgery and have a higher risk of complications. As a result, TKA is considerably less cost-effective in patients with less severe OA.[274] If SDM-interventions result in a greater proportion of patients with milder disease delaying TJA, it is likely that SDM-interventions would be more cost-effective in this context. Together, this body of evidence suggests that the policy focus on patient decision aids in this context is justified,

as they may be able to deliver on the promise of reducing costs and improving outcomes,[15] both with respect to the health of individuals and the process of care.

A second finding from this dissertation relates to the long-term impact of patient decision aids on uptake of TJA. Given the progressive nature of OA, there are questions about whether individuals exposed to patient decision aids forego TJA, or delay, and if so, for how long. The analysis in Chapter Three suggests that, at seven-years follow-up, the proportion of individuals undergoing TJA may remain lower in those exposed to patient decision aids. While both the original trial and seven-year time horizon analysis in Chapter Three were not powered to detect this outcome, this dissertation provides the first evidence exploring this research question and provides insight that can be used to inform a trial that aims to determine the long-term impact of SDM-interventions on uptake of TJA. For example, the analysis in Chapter Three found that very few individuals underwent TJA after five years follow-up, suggesting that this may be an appropriate time horizon for a future trial. Furthermore, the trial reported in Chapter Two provided some preliminary estimates of the uptake of TJA in both groups which suggests that a future trial would need to include approximately 1,300 participants to ensure 80% power.

A third implication for clinical practice relates to the economic value of SDM-interventions that may not be captured in conventional CEA with health-related QALYs. The study reported in Chapter Six identified that, in the context of treatment decision-

making for advanced knee OA, the process of SDM is valued independent of its potential influence of health outcomes. This suggests that it may be justified to invest in patient decision aids or other SDM-interventions in this context, at the expense of other interventions that provide health benefits, even if SDM-interventions do not improve health outcomes. However, this evidence does not suggest that encouraging SDM should be pursued at any cost and may not be generalized to all SDM-interventions. An appropriate decision should consider the required investment in SDM-interventions, their effectiveness in encouraging SDM, and the value of SDM for the specific decision. There will inevitability be contexts where SDM is valued more, and others where it is valued less, or not at all.

Incorporating the value of the process of SDM into the economic evaluation of SDM-interventions may impact adoption decisions. In the current context, incorporating the value of SDM-interventions into the QALY does not change the adoption decision, but simply renders an already dominant intervention even more dominant. However, in other contexts, this may not be the case. SDM-interventions may prove to be less costly and provide less QALYs than usual care. In such cases, so long as patients are making an informed decision to choose a treatment that provides fewer QALYs, that would likely be acceptable to a decision-maker. However, there are two cases where incorporating the value of the process of SDM may change the adoption decision. The first is where the SDM-intervention is costlier, and provides more QALYs, but falls above the cost-

effectiveness threshold. In this case, considering the value of the process of SDM may result in the cost-effectiveness estimate moving below the threshold, thereby changing the adoption decision. The second case is where SDM-interventions are costlier and provide fewer QALYs. Incorporating the value of the process of SDM within the QALY may result in the CEA estimating that the SDM-intervention provides more QALYs, at which point the question is whether the ICER falls above or below the cost-effectiveness threshold.

Given the resources involved in performing valuation studies, it is worth asking: *What is the likelihood that considering the added value of the process of SDM will impact adoption decisions?* Chapter Six estimated a health state utility value of an improvement from 'no effort' to 'every effort' to engage in SDM to be between 0.007 to 0.033 for the moderate and extreme pain or discomfort scenarios, respectively. While this value will vary depending upon the context, the incremental benefit from a SDM-intervention will always represent a fraction of the overall value of SDM for a given context. The incremental value depends on the baseline level of SDM in usual care and the effectiveness of the SDM-intervention at encouraging SDM behaviours. Evidence suggests that the baseline level of SDM in usual care, as reported by patients is high. For example, 61-86% of patients surveyed in primary care reported the top score on all three components of the CollaboRATE scale, depending upon practice and mode of administration.[265] The baseline level of SDM may be lower using other measures that that

do not exhibit the same ceiling effects. In terms of the effectiveness of SDM-interventions in encouraging SDM, surprisingly, the evidence to support this is limited. The most recent Cochrane review of patient decision aids found only 10 of 105 had evaluated this impact.[13,275] While there is a dearth of evidence, that which is available is encouraging. Of the five trials that directly observed the impact on SDM, all found they resulted in a statistically significant increase in SDM.[13] All told, while there is limited evidence on how much patient decision aids result in increased SDM, the existing literature suggests that the incremental benefit of the process of SDM associated with the implementation of SDM-interventions may be small. Consequently, the largest driver of the value of SDM-interventions may be their potential impact on treatment choice and health outcomes, and the value associated with the process of SDM may only impact the adoption decision in cases where differences in health-related QALYs or costs are small. This was the case for Brennan, where both incremental costs and health-related QALYs, as measured by the SF-6D were small, and incorporating the value of improved communication moved the point estimate from the northwest quadrant of the cost-effectiveness plane (more costly, less benefit) to the northeast (more costly, more benefit).[259]

Moving beyond an economic perspective, which focuses on the mean effects, there are implications for individual patients. For example, the valuation survey described in Chapter Six found substantial heterogeneity in preferences for SDM. For example, the latent class analysis (Table 6.5) found four classes of preferences. Class 2 had a strong

preference for improved outcomes, the main effects coefficients for SDM were not significant. By comparison, members of Class 4 had a strong preference for SDM, as these were the only statistically significant main effects coefficients. While these classes of preferences were present, there did not appear to be demographic or clinical characteristics that predicted class membership (Appendix 6.4, Table A.4). This highlights the importance of the clinical encounter in determining the preferred level of information and involvement in treatment decision-making for each patient.

7.7 Areas for future research

There are several opportunities to expand on the studies presented in this dissertation. It was unclear whether the patient decision aids evaluated in the trial reported in Chapter Two are available for use in routine care in Canada, and if so, how much they cost. There are new patient decision aids under development for the Canadian context, including one for patients considering TKA which uses PROMs to individualize outcomes estimates to patients. Future research, such as randomized controlled trials, are needed to establish their effectiveness and cost-effectiveness.

Participants in the trial (reported in Chapters Two and Three) were recruited from two orthopedic screening clinics, both of which had pre-screening to determine whether they were 'minimally appropriate for considering total joint arthroplasty.'[58] As a result, the study population had moderate or severe pain and functional limitations, meaning that the findings may not be generalizable to other settings. Future research should

investigate the effectiveness of patient decision aids in jurisdictions where pre-screening is not undertaken, and/or in patients with less severe pain and functional impairments.

The inclusion criteria for the trial reported in Chapters Two and Three required that participants were able to read and understand English. Future research should consider translating SDM-interventions into different languages and adapting them for different cultural contexts.[276] Previous research has demonstrated that SDM-interventions can be adapted to meet the decisional needs of different cultural groups [277] and improve patient-doctor communication and decision quality.[278] A 2014 systematic review found that, not only do SDM-interventions improve outcomes for disadvantaged patients, but may be more beneficial to them than higher literacy/socioeconomic status patients.[279]

The study reported in Chapter Six found that participants do value the process of SDM in the context of treatment decision-making for advanced knee OA. However, it is logical to expect this value to vary depending on factors such as the number of treatment options available, and the differences in benefits and harms. Future research should aim to identify the clinical decisions where the process of SDM is most valuable.

An additional finding in the empirical valuation study (see Chapter Six) was that there was considerable heterogeneity in the value of SDM. Previous research has demonstrated that some patients want to be engaged in decisions about their health care, such as younger, female, and more well-educated patients, while others prefer to defer

decision-making to their provider.[202] Efforts to encourage SDM should respect the decision-making preferences of patients, however they should also consider equity. Providing SDM-interventions only in specific clinical contexts or subgroups of patients who express a strong preference for SDM may exacerbate existing disparities rather than reduce them.

One important consideration is how the value of processes is incorporated into economic evaluation. This includes considering whether the value of the process should be treated as a utility gain or disutility, and perhaps most importantly, the duration of the effect. In the valuation survey, respondents were asked to trade health outcomes, described as the potential improvement in pain or discomfort, for greater SDM. Implicitly, this trade-off includes a duration for the potential improvement in pain or discomfort. For instance, survey instructions stated that the potential improvement in pain or discomfort would be present one year after their consult and would be expected to last for at least two years. The discussion section of Chapter Six described how this value could be incorporated with a CEA, which required an assumption that the value of SDM lasted for one-year. However, this could be an under- or over-estimate. Future research should explore how long the process and non-health benefits last.

A further consideration is how different approaches to quantifying and presenting valuations of health care process not captured within the QALY influence decision-making. One assumption is that incorporating the value of health care processes within

the QALY can help ensure that this value is considered appropriately given resource constraints and competing priorities. Current economic evaluation methods recommend presenting disaggregated consequences in the form of an impact inventory but recognize that a single figure may be helpful for decision-making. Currently, there are no established and accepted methods for doing so. Future research should identify the potential benefits and limitations of aggregating and disaggregating consequences, both from the perspective of decision-makers and the subsequent impact on resource allocation decisions.

The study reported in Chapter Five identified that the value of SDM varies depending on the context. For economic evaluations of SDM-interventions, this implies that the value of SDM estimated in one context cannot simply be applied to other treatment decisions. For example, the value of SDM may be dramatically different in the context of treatment decision-making for a chronic condition where patients can switch between treatments, compared to here where undergoing TJA cannot be reversed. However, it may not be realistic to perform a valuation study for each economic evaluation of a SDM-intervention given the resources required. One option for future research is to generate health state utility values for a validated instrument that account for different types of decisions (e.g., chronic vs. one-off). Doing so would allow future economic evaluations to consider this added value of SDM-interventions without completing a valuation study.

7.8 Conclusions

The dissertation reported the first economic evaluation of patient decision aids in the context of treatment decision-making for advanced OA; a clear policy priority. The results described in this dissertation suggest that patient decision aids may be a cost-effective use of resources, resulting in lower costs, improved health outcomes, and provide additional value through improvements in the process of care. While these results are not definitive and cannot be generalized to all SDM-interventions in all contexts, this dissertation provides insight into how SDM-interventions can be valued to inform resource allocation decisions.

References

1. National Health Service - England. Involving people in their own care. (2016). Available at: https://www.england.nhs.uk/ourwork/patients/patient-participation/. (Accessed: 15th March 2016)
2. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. *Institute of Medicine* (2001). Available at: http://www.iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx. (Accessed: 2nd June 2014)
3. Coulter, A. Paternalism or partnership? *BMJ* **319,** 719–720 (1999).
4. Barry, M. J. & Edgman-Levitan, S. Shared decision making--pinnacle of patient-centered care. *N. Engl. J. Med.* **366,** 780–781 (2012).
5. National Health Service - England. Shared Decision Making (SDM) - NHS. (2016). Available at: http://sdm.rightcare.nhs.uk/. (Accessed: 15th March 2016)
6. Makoul, G. & Clayman, M. L. An integrative model of shared decision making in medical encounters. *Patient Educ. Couns.* **60,** 301–312 (2006).
7. Légaré, F., Ratté, S., Gravel, K. & Graham, I. D. Barriers and facilitators to implementing shared decision-making in clinical practice: Update of a systematic review of health professionals' perceptions. *Patient Educ. Couns.* **73,** 526–535 (2008).
8. Coulter, A. Shared decision making: everyone wants it, so why isn't it happening? *World Psychiatry* **16,** 117–118 (2017).
9. Godolphin, W. Shared decision-making. *Heal. Q* **12,** e186–e190 (2009).
10. Légaré, F. *et al.* Interventions for improving the adoption of shared decision making by healthcare professionals. *Cochrane Database Syst. Rev.* CD006732 (2010).
11. Légaré, F. *et al.* Training family physicians in shared decision-making to reduce the overuse of antibiotics in acute respiratory infections: a cluster randomized trial. *CMAJ* **184,** E726-734 (2012).
12. Légaré, F. *et al.* Training health professionals in shared decision-making: An international environmental scan. *Patient Educ. Couns.* **88,** 159–169 (2012).
13. Stacey, D. *et al.* Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst. Rev.* CD001431 (2017).
14. Stacey, D. *et al.* Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst. Rev.* CD001431 (2011). doi:10.1002/14651858.CD001431.pub3
15. Oshima Lee, E. & Emanuel, E. J. Shared decision making to improve care and reduce costs. *N. Engl. J. Med.* **368,** 6–8 (2013).
16. Trenaman, L., Bryan, S. & Bansback, N. The cost-effectiveness of patient decision aids: A systematic review. *Healthc* **2,** 251–257 (2014).
17. Walsh, T. *et al.* Undetermined impact of patient decision support interventions on healthcare costs and savings: systematic review. *BMJ* **348,** g188 (2014).

18. Edwards, A. Patient-based outcome results from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. *Fam. Pract.* **21,** 347–354 (2004).
19. Bieber, C. *et al.* A shared decision-making communication training program for physicians treating fibromyalgia patients: effects of a randomized controlled trial. *J. Psychosom. Res.* **64,** 13–20 (2008).
20. Murray, C. & Frenk, J. A WHO Framework for Health System Performance Assessment. (2000).
21. National Health Service - England. NHS England commitment to Shared Decision Making. (2016). Available at: https://www.england.nhs.uk/ourwork/pe/sdm/commitment/. (Accessed: 15th March 2016)
22. National Health Service - England. Quality standard for patient experience in adult NHS services. (2012). Available at: https://www.nice.org.uk/guidance/qs15. (Accessed: 15th March 2016)
23. British Columbia Ministry of Health. Setting priorities for the B.C. Health System. (2014). Available at: http://www.health.gov.bc.ca/library/publications/year/2014/Setting-priorities-BC-Health-Feb14.pdf. (Accessed: 27th July 2018)
24. Gravel, K., Légaré, F. & Graham, I. D. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. *Implement Sci* **1,** 16 (2006).
25. Spatz, E. S., Elwyn, G., Moulton, B. W., Volk, R. J. & Frosch, D. L. Shared decision making as part of value based care: new U.S. policies challenge our readiness. *Z. Für Evidenz Fortbild. Qual. Im Gesundheitswesen* **123–124,** 104–108 (2017).
26. Altman, R. *et al.* Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum.* **29,** 1039–1049 (1986).
27. Zhang, Y. & Jordan, J. M. Epidemiology of Osteoarthritis. *Clin. Geriatr. Med.* **26,** 355–369 (2010).
28. Sayre, E. C. *et al.* The effect of disease site (knee, hip, hand, foot, lower back or neck) on employment reduction due to osteoarthritis. *PloS One* **5,** e10470 (2010).
29. Lawrence, R. C. *et al.* Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheum.* **58,** 26–35 (2008).
30. Kopec, J. A. *et al.* Descriptive epidemiology of osteoarthritis in British Columbia, Canada. *J. Rheumatol.* **34,** 386–393 (2007).
31. Birtwhistle, R. *et al.* Prevalence and management of osteoarthritis in primary care: an epidemiologic cohort study from the Canadian Primary Care Sentinel Surveillance Network. *CMAJ Open* **3,** E270–E275 (2015).

32. MacDonald, K. V., Sanmartin, C., Langlois, K. & Marshall, D. A. Symptom onset, diagnosis and management of osteoarthritis. *Health Rep.* **25,** 10–17 (2014).
33. McAlindon, T. E. *et al.* OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthr Cartil* **22,** 363–388 (2014).
34. Hochberg, M. C. *et al.* American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res.* **64,** 465–474 (2012).
35. Murray, D. G. History of Total Knee Replacement. in *Total Knee Replacement* 3–15 (Springer, 1991).
36. Katti, K. S. Biomaterials in total joint replacement. *Colloids Surf. B Biointerfaces* **39,** 133–142 (2004).
37. Harris, W. H. & Sledge, C. B. Total hip and total knee replacement. *N. Engl. J. Med.* **323,** 725–731 (1990).
38. Ethgen, O., Bruyère, O., Richy, F., Dardennes, C. & Reginster, J.-Y. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J. Bone Joint Surg. Am.* **86-A,** 963–974 (2004).
39. Chang, R. W., Pellisier, J. M. & Hazen, G. B. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. *JAMA* **275,** 858–865 (1996).
40. Losina, E. *et al.* Cost-effectiveness of total knee arthroplasty in the United States. *Arch. Intern. Med.* **169,** 1113–1122 (2009).
41. Daigle, M. E., Weinstein, A. M., Katz, J. N. & Losina, E. The cost-effectiveness of total joint arthroplasty: A systematic review of published literature. *Best Pract. Res. Clin. Rheumatol.* **26,** 649–658 (2012).
42. Canadian Institute for Health Information (CIHI). Hip and Knee Replacements in Canada: 2015/16 Quick Stats. (2017). Available at: https://www.cihi.ca/en/joint-replacements. (Accessed: 21st December 2016)
43. Hawker, G. *et al.* Perspectives of Canadian stakeholders on criteria for appropriateness for total joint arthroplasty in patients with hip and knee osteoarthritis. *Arthritis Rheumatol.* **67,** 1806–1815 (2015).
44. Jones, C. A., Beaupre, L. A., Johnston, D. W. C. & Suarez-Almazor, M. E. Total joint arthroplasties: current concepts of patient outcomes after surgery. *Rheum. Dis. Clin. North Am.* **33,** 71–86 (2007).
45. Bourne, R. B., Chesworth, B. M., Davis, A. M., Mahomed, N. N. & Charron, K. D. J. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? *Clin. Orthop.* **468,** 57–63 (2010).
46. Hamilton, D. F. *et al.* What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. *BMJ Open* **3,** e002525 (2013).
47. Hawker, G. A. *et al.* Which patients are most likely to benefit from total joint arthroplasty? *Arthritis Rheum.* **65,** 1243–1252 (2013).

48. Talmo, C. T., Aghazadeh, M. & Bono, J. V. Perioperative complications following total joint replacement. *Clin. Geriatr. Med.* **28,** 471–487 (2012).
49. Labek, G., Thaler, M., Janda, W., Agreiter, M. & Stöckl, B. Revision rates after total joint replacement: cumulative results from worldwide joint register datasets. *J Bone Jt. Surg Br* **93-B,** 293–297 (2011).
50. Barnett, A. J. & Toms, A. D. Revision total hip and knee replacement. *Clin. Geriatr. Med.* **28,** 431–446 (2012).
51. Hoffmann, T. C. & Del Mar, C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA Intern. Med.* **175,** 274–286 (2015).
52. Skou, S. T. *et al.* A randomized, controlled trial of total knee replacement. *N. Engl. J. Med.* **373,** 1597–1606 (2015).
53. CMS cancels Shared Decision Making ACO model. *Health Exec* (2018). Available at: http://www.healthexec.com/topics/care-delivery/cms-cancels-shared-decision-making-aco-model. (Accessed: 6th April 2018)
54. Ibrahim, S. A. Decision aids and elective joint replacement - how knowledge affects utilization. *N. Engl. J. Med.* **376,** 2509–2511 (2017).
55. Arterburn, D. *et al.* Introducing decision aids at Group Health was linked to sharply lower hip and knee surgery rates and costs. *Health Aff. (Millwood)* **31,** 2094–2104 (2012).
56. Borkhoff, C. M. *et al.* The effect of patients' sex on physicians' recommendations for total knee arthroplasty. *CMAJ* **178,** 681–687 (2008).
57. Sepucha, K. *et al.* Patient decision aids improve decision quality and patient experience and reduce surgical rates in routine orthopaedic care: a prospective cohort study. *J. Bone Joint Surg. Am.* **99,** 1253–1260 (2017).
58. Stacey, D. *et al.* Impact of patient decision aids on appropriate and timely access to hip or knee arthroplasty for osteoarthritis: a randomized controlled trial. *Osteoarthr Cartil* **24,** 99–107 (2015).
59. Bozic, K. J. *et al.* Shared decision making in patients with osteoarthritis of the hip and knee: results of a randomized controlled trial. *J. Bone Joint Surg. Am.* **95,** 1633–1639 (2013).
60. Ibrahim, S. A. *et al.* Effect of a decision aid on access to total knee replacement for black patients with osteoarthritis of the knee: a randomized clinical trial. *JAMA Surg.* **152,** e164225 (2017).
61. Wilson, M. G., May, D. S. & Kelly, J. J. Racial differences in the use of total knee arthroplasty for osteoarthritis among older Americans. *Ethn. Dis.* **4,** 57–67 (1994).
62. Skinner, J., Weinstein, J. N., Sporer, S. M. & Wennberg, J. E. Racial, ethnic, and geographic disparities in rates of knee arthroplasty among Medicare patients. *N. Engl. J. Med.* **349,** 1350–1359 (2003).

63. Stacey, D. *et al.* Decision aid for patients considering total knee arthroplasty with preference report for surgeons: a pilot randomized controlled trial. *BMC Musculoskelet. Disord.* **15,** 54 (2014).
64. Oxford Dictionaries | The World's Most Trusted Dictionary Provider. *Oxford Dictionaries* (2018). Available at: https://www.oxforddictionaries.com/. (Accessed: 15th May 2018)
65. Committee on Assessing and Valuing the Services of Aquatic and Related Terrestrial Ecosystems. *Valuing Ecosystem Services: Toward Better Environmental Decision-Making*. (National Academies Press, 2005).
66. Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L. & Torrance, G. W. *Methods for the Economic Evaluation of Health Care Programmes*. (Oxford University Press, 2015).
67. Williams, A. The economic role of 'health indicators'. in *Measuring the social benefits of medicine* (Office of Health Economics, 1983).
68. Culyer, A. J. The normative economics of health care finance and provision. *Oxf. Rev. Econ. Policy* **5,** 34–58 (1989).
69. Brouwer, W. B. F., Culyer, A. J., van Exel, N. J. A. & Rutten, F. F. H. Welfarism vs. extra-welfarism. *J. Health Econ.* **27,** 325–338 (2008).
70. Briggs, A. H. & O'Brien, B. J. The death of cost-minimization analysis? *Health Econ.* **10,** 179–184 (2001).
71. Dakin, H. & Wordsworth, S. Cost-minimisation analysis versus cost-effectiveness analysis, revisited. *Health Econ.* **22,** 22–34 (2013).
72. McIntosh, E., Donaldson, C. & Ryan, M. Recent advances in the methods of cost-benefit analysis in healthcare. Matching the art to the science. *PharmacoEconomics* **15,** 357–367 (1999).
73. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. (2013). Available at: https://www.nice.org.uk/article/pmg9/chapter/Foreword. (Accessed: 27th July 2018)
74. Coast, J. Maximisation in extra-welfarism: A critique of the current position in health economics. *Soc. Sci. Med. 1982* **69,** 786–792 (2009).
75. Weinstein, M. C., Torrance, G. & McGuire, A. QALYs: the basics. *Value Health* **12,** S5–S9 (2009).
76. Torrance, G. W. Utility approach to measuring health-related quality of life. *J. Chronic Dis.* **40,** 593–603 (1987).
77. Macran, S. & Kind, P. 'Death' and the valuation of health-related quality of life. *Med. Care* **39,** 217–227 (2001).
78. Carson, R. T. & Louviere, J. J. A common nomenclature for stated preference elicitation approaches. *Environ. Resour. Econ.* **49,** 539–559 (2011).

79. Ryan, M., Gerard, K. & Amaya-Amaya, M. *Using discrete choice experiments to value health and health care*. (Springer, 2008).
80. Hanley, N., Shogren, J. & White, B. *Introduction to Environmental Economics*. (Oxford University Press, 2013).
81. Perman, R., Yu, M., McGilvray, J. & Common, M. *Natural Resource and Environmental Economics*. (Pearson Education, 2003).
82. Rohlfs, C., Sullivan, R. & Kniesner, T. New estimates of the value of a statistical life using air bag regulations as a quasi-experiment. *Am. Econ. J. Econ. Policy* **7,** 331–359 (2015).
83. Ali, S. & Ronaldson, S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br. Med. Bull.* **103,** 21–44 (2012).
84. Clark, M. D., Determann, D., Petrou, S., Moro, D. & de Bekker-Grob, E. W. Discrete choice experiments in health economics: a review of the literature. *PharmacoEconomics* **32,** 883–902 (2014).
85. Flynn, T. N. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev. Pharmacoecon. Outcomes Res.* **10,** 259–267 (2010).
86. Rabin, R. & de Charro, F. EQ-5D: a measure of health status from the EuroQol Group. *Ann. Med.* **33,** 337–343 (2001).
87. Donaldson, C. Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'? *Soc. Sci. Med.* **49,** 551–563 (1999).
88. McNamee, P. *et al.* Chained time trade-off and standard gamble methods. Applications in oesophageal cancer. *Eur J Health Econ* **5,** 81–86 (2004).
89. von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior*. (Princeton University Press, 1944).
90. Flynn, T. N. Using conjoint analysis and choice experiments to estimate QALY values. *PharmacoEconomics* **28,** 711–722 (2010).
91. Fraenkel, L., Bodardus, S. & Wittink, D. R. Understanding patient preferences for the treatment of lupus nephritis with adaptive conjoint analysis. *Med. Care* **39,** 1203–1216 (2001).
92. Louviere, J. J., Flynn, T. N. & Carson, R. T. Discrete choice experiments are not conjoint analysis. *J. Choice Model.* **3,** 57–72 (2010).
93. Bansback, N., Brazier, J., Tsuchiya, A. & Anis, A. Using a discrete choice experiment to estimate health state utility values. *J. Health Econ.* **31,** 306–318 (2012).
94. Murray, E. *et al.* Randomised controlled trial of an interactive multimedia decision aid on hormone replacement therapy in primary care. *BMJ* **323,** 490–493 (2001).
95. Murray, E. *et al.* Randomised controlled trial of an interactive multimedia decision aid on benign prostatic hypertrophy in primary care. *BMJ* **323,** 493–496 (2001).

96. Kennedy, A. D. M. *et al.* A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women's preferences in the management of menorrhagia. *Health Technol Assess* **7,** 1–76 (2003).
97. Kennedy, A. D. M. *et al.* Effects of decision aids for menorrhagia on treatment choices, health outcomes, and costs: a randomized controlled trial. *JAMA* **288,** 2701–2708 (2002).
98. Hollinghurst, S. *et al.* Economic evaluation of the DiAMOND randomized trial: cost and outcomes of 2 decision aids for mode of delivery among women with a previous cesarean section. *Med. Decis. Making* **30,** 453–463 (2010).
99. Tubeuf, S. *et al.* Cost effectiveness of a web-based decision aid for parents deciding about MMR vaccination: a three-arm cluster randomised controlled trial in primary care. *Br. J. Gen. Pract. J. R. Coll. Gen. Pract.* **64,** e493-499 (2014).
100. Patel, S. *et al.* Primum non nocere: shared informed decision making in low back pain--a pilot cluster randomised trial. *BMC Musculoskelet. Disord.* **15,** 282 (2014).
101. Parkinson, B. *et al.* Cost-effectiveness of the BRECONDA decision aid for women with breast cancer: results from a randomized controlled trial. *Psychooncology.* Epub ahead of print (2018).
102. Trenaman, L. *et al.* Exploring the potential cost-effectiveness of patient decision aids for use in adults with obstructive sleep apnea: a case study. *Med Decis Mak.* **35,** 271–82 (2015).
103. Cantor, S. B., Rajan, T., Linder, S. K. & Volk, R. J. A framework for evaluating the cost-effectiveness of patient decision aids: A case study using colorectal cancer screening. *Prev. Med.* **77,** 168–173 (2015).
104. Penton, H. *et al.* Potential cost-effectiveness for using patient decision aids to guide osteoporosis treatment. *Osteoporos Int* **27,** 2697–2707 (2016).
105. Vuorma, S. *et al.* A randomized trial among women with heavy menstruation–impact of a decision aid on treatment outcomes and costs. *Health Expect.* **7,** 327–337 (2004).
106. Canadian Foundation for Healthcare Improvement. Healthcare priorities in Canada: A backgrounder. (2014). Available at: http://www.cfhi-fcass.ca/sf-docs/default-source/documents/harkness-healthcare-priorities-canada-backgrounder-e.pdf. (Accessed: 12th August 2016)
107. Canadian Medical Association. Health Care Transformation in Canada. (2009). Available at: https://www.cma.ca/Assets/assets-library/document/en/advocacy/policy-research/CMA_Policy_Health_Care_Transformation_in_Canada_Change_that_Works_Care_that_Lasts_PD10-05-e.pdf. (Accessed: 12th August 2016)
108. Weston, W. W. Informed and shared decision-making: the crux of patient-centred care. *CMAJ* **165,** 438–439 (2001).

109. Nwachukwu, B. U. *et al.* Current status of cost utility analyses in total joint arthroplasty: a systematic review. *Clin. Orthop.* **473,** 1815–1827 (2015).
110. Trenaman, L., Bryan, S. & Bansback, N. The cost-effectiveness of patient decision aids: A systematic review. *Healthc. Amst. Neth.* **2,** 251–257 (2014).
111. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada, 3rd Edition. (2006). Available at: https://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf. (Accessed: 27th July 2018)
112. Ontario Ministry of Health and Long-term Care. Schedule of Benefits. Available at: http://www.health.gov.on.ca/english/providers/program/ohip/sob/physserv/sob_master11062015.pdf. (Accessed: 11th September 2015)
113. Canadian Institute for Health Information. CIHI Patient Cost Estimator. (2015). Available at: https://www.cihi.ca/en/spending-and-health-workforce/spending/patient-cost-estimator. (Accessed: 11th September 2015)
114. Hospital Central Agreement. Collective Agreement Between Ontario Nurses' Association and Hospitals. Available at: https://www.ona.org/documents/File/pdf/cas/hospitals/HospitalCentralAgreement-English_March312016.pdf. (Accessed: 11th September 2015)
115. Wailoo, A., Hernandez Alava, M. & Escobar Martinez, A. Modelling the relationship between the WOMAC Osteoarthritis Index and EQ-5D. *Health Qual. Life Outcomes* **12,** 37 (2014).
116. Xie, F. *et al.* Use of a disease-specific instrument in economic evaluations: mapping WOMAC onto the EQ-5D utility index. *Value Health* **13,** 873–878 (2010).
117. Fitzgerald, P., Goodacre, S. W., Cross, E. & Dixon, S. Cost-effectiveness of point-of-care biomarker assessment for suspected myocardial infarction: the randomized assessment of treatment using panel Assay of cardiac markers (RATPAC) trial. *Acad Emerg Med* **18,** 488–495 (2011).
118. Manca, A., Hawkins, N. & Sculpher, M. J. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ.* **14,** 487–496 (2005).
119. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple Imputation by Chained Equations: What is it and how does it work? *Int. J. Methods Psychiatr. Res.* **20,** 40–49 (2011).
120. Faria, R., Gomes, M., Epstein, D. & White, I. R. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics* **32,** 1157–1170 (2014).
121. Little, R. J. A. & Rubin, D. B. *Statistical Analysis With Missing Data*. (Wiley, 1987).
122. Briggs, A. H., Wonderling, D. E. & Mooney, C. Z. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Econ.* **6,** 327–340 (1997).

123. Black, W. C. The CE plane: a graphic representation of cost-effectiveness. *Med. Decis. Making* **10,** 212–214 (1990).
124. Fenwick, E., Claxton, K. & Sculpher, M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ.* **10,** 779–787 (2001).
125. Barton, G. R. *et al.* Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health Qual. Life Outcomes* **6,** 51 (2008).
126. Grootendorst, P. *et al.* A model to estimate health utilities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. *J. Rheumatol.* **34,** 534–542 (2007).
127. Stacey, D. *et al.* Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst. Rev.* CD001431 (2014).
128. Hudak, P. L. *et al.* "You're perfect for the procedure! Why don't you want it?" Elderly arthritis patients' unwillingness to consider total joint arthroplasty surgery: a qualitative study. *Med. Decis. Making* **22,** 272–278 (2002).
129. Quintana, J. M. *et al.* Health-related quality of life and appropriateness of knee or hip joint replacement. *Arch. Intern. Med.* **166,** 220–226 (2006).
130. Gonzalez Sáenz de Tejada, M. *et al.* Patient expectations and health-related quality of life outcomes following total joint replacement. *Value Health* **13,** 447–454 (2010).
131. Klett, M.-J., Frankovich, R., Dervin, G. F. & Stacey, D. Impact of a surgical screening clinic for patients with knee osteoarthritis: a descriptive study. *Clin J Sport Med* **22,** 274–277 (2012).
132. Losina, E., Thornhill, T. S., Rome, B. N., Wright, J. & Katz, J. N. The dramatic increase in total knee replacement utilization rates in the United States cannot be fully explained by growth in population size and the obesity epidemic. *J. Bone Joint Surg. Am.* **94,** 201–207 (2012).
133. Longworth, L. & Rowen, D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value Health* **16,** 202–210 (2013).
134. Claxton, K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J. Health Econ.* **18,** 341–364 (1999).
135. Whitehurst, D. G. T. & Bryan, S. Trial-based clinical and economic analyses: the unhelpful quest for conformity. *Trials* **14,** 421 (2013).
136. Petrou, S. & Gray, A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ* **342,** d1548 (2011).
137. Ramsey, S. *et al.* Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* **8,** 521–533 (2005).
138. Sculpher, M. J., Claxton, K., Drummond, M. & McCabe, C. Whither trial-based economic evaluation for health care decision making? *Health Econ.* **15,** 677–687 (2006).

139. van Walraven, C. & Hawken, S. Competing risk bias in Kaplan-Meier risk estimates can be corrected. *J. Clin. Epidemiol.* **70,** 101–105 (2016).
140. Lacny, S. *et al.* Kaplan-Meier survival analysis overestimates cumulative incidence of health-related events in the presence of competing risks: a meta-analysis. *J. Clin. Epidemiol.* **93,** 25–35 (2017).
141. Austin, P. C. & Fine, J. P. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat. Med.* **36,** 4391–4400 (2017).
142. Wodchis, W., Bushmeneva, K., Nikitovic, M. & McKillip, I. Guidelines on person-level costing using administrative databases in Ontario. (2013). Available at: http://www.hsprn.ca/uploads/files/Guidelines_on_PersonLevel_Costing_May_2013.pdf. (Accessed: 9th December 2014)
143. Canadian Institute for Health Information. Cost of a Standard Hospital Stay. Available at: http://indicatorlibrary.cihi.ca/display/HSPIL/Cost+of+a+Standard+Hospital+Stay. (Accessed: 23rd November 2017)
144. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. 4th Edition. (2017). Available at: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf. (Accessed: 27th July 2018)
145. Lumley, T., Diehr, P., Emerson, S. & Chen, L. The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* **23,** 151–169 (2002).
146. Bachmeier, C. J. M. *et al.* A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthr Cartil* **9,** 137–146 (2001).
147. Beswick, A. D., Wylde, V., Gooberman-Hill, R., Blom, A. & Dieppe, P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* **2,** e000435 (2012).
148. Blom, A. W. *et al. Improving patients' experience and outcome of total joint replacement: the RESTORE programme*. (NIHR Journals Library, 2016).
149. Kurtz, S. M. *et al.* Infection Burden for Hip and Knee Arthroplasty in the United States. *J. Arthroplasty* **23,** 984–991 (2008).
150. Nilsdotter, A.-K. & Lohmander, L. S. Age and waiting time as predictors of outcome after total hip replacement for osteoarthritis. *Rheumatology* **41,** 1261–1267 (2002).
151. Sepucha, K. R. *et al.* Establishing the effectiveness of patient decision aids: key constructs and measurement instruments. *BMC Med. Inform. Decis. Mak.* **13,** S12 (2013).

152. Brazier, J. E., Dixon, S. & Ratcliffe, J. The role of patient preferences in cost-effectiveness analysis: a conflict of values? *PharmacoEconomics* **27,** 705–712 (2009).
153. Payne, K., McAllister, M. & Davies, L. M. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Econ.* **22,** 258–271 (2013).
154. Greco, G., Lorgelly, P. & Yamabhai, I. Outcomes in economic evaluations of public health interventions in low- and middle-income countries: health, capabilities and subjective wellbeing. *Health Econ.* **25 Suppl 1,** 83–94 (2016).
155. Donaldson, C. & Shackley, P. Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc. Sci. Med.* **44,** 699–707 (1997).
156. Donabedian, A. The quality of care: How can it be assessed? *JAMA* **260,** 1743–1748 (1988).
157. Donabedian, A. *An Introduction to Quality Assurance in Health Care*. (Oxford University Press, USA, 2002).
158. Donabedian, A. Evaluating the quality of medical care. *Milbank Q.* **83,** 691–729 (2005).
159. Brennan, V. K. & Dixon, S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *PharmacoEconomics* **31,** 677–691 (2013).
160. Mooney, G. Beyond health outcomes: The benefits of health care. *Health Care Anal.* **6,** 99–105 (1998).
161. van Mastrigt, G. A., Paulus, A. T., Aarts, M.-J., Evers, S. M. & Alayli-Goebbels, A. F. A qualitative study on the views of experts regarding the incorporation of non-health outcomes into the economic evaluations of public health interventions. *BMC Public Health* **15,** 954 (2015).
162. Ryan, M. & Shackley, P. Assessing the benefits of health care: how far should we go? *Qual. Health Care* **4,** 207–213 (1995).
163. Weisbrod, B. A. Collective-consumption services of individual-consumption goods. *Q. J. Econ.* **78,** 471–477 (1964).
164. McGuire, A., Henderson, J. & Mooney, G. H. *The Economics of Health Care: An Introductory Text*. (Routledge & Kegan Paul, 1988).
165. Gjerberg, E., Lillemoen, L., Førde, R. & Pedersen, R. End-of-life care communications and shared decision-making in Norwegian nursing homes - experiences and perspectives of patients and relatives. *BMC Geriatr.* **15,** 103 (2015).
166. Brouwer, W. B. F., Exel, N. J. A. van, Berg, B. van den, Bos, G. A. M. van den & Koopmanschap, M. A. Process utility from providing informal care: the benefit of caring. *Health Policy* **74,** 85–99 (2005).
167. Wittenberg, E. & Prosser, L. A. Disutility of illness for caregivers and families: a systematic review of the literature. *PharmacoEconomics* **31,** 489–500 (2013).

168. Wilson, S. R. *et al.* Shared treatment decision making improves adherence and outcomes in poorly controlled asthma. *Am. J. Respir. Crit. Care Med.* **181,** 566–577 (2010).
169. Härter, M. *et al.* Shared decision making and the use of decision aids. *Dtsch. Ärztebl. Int.* **112,** 672–679 (2015).
170. Shabason, J. E., Mao, J. J., Frankel, E. S. & Vapiwala, N. Shared decision-making and patient control in radiation oncology: implications for patient satisfaction. *Cancer* **120,** 1863–1870 (2014).
171. Al-Janabi, H., van Exel, J., Brouwer, W. & Coast, J. A framework for including family health spillovers in economic evaluation. *Med Decis Mak.* **36,** 176–186 (2016).
172. Bozic, K. J. *et al.* Patient, surgeon, and healthcare purchaser views on the use of decision and communication aids in orthopaedic surgery: a mixed methods study. *BMC Health Serv. Res.* **14,** 366 (2014).
173. Mather, R. C. *et al.* Economic evaluation of access to musculoskeletal care: the case of waiting for total knee arthroplasty. *BMC Musculoskelet. Disord.* **15,** 22 (2014).
174. Yorgason, J. B., Roper, S. O., Sandberg, J. G. & Berg, C. A. Stress spillover of health symptoms from healthy spouses to patient spouses in older married couples managing both diabetes and osteoarthritis. *Fam Syst Health* **30,** 330–343 (2012).
175. Sepucha, K. R. *et al.* Decision quality instrument for treatment of hip and knee osteoarthritis: a psychometric evaluation. *BMC Musculoskelet. Disord.* **12,** 149 (2011).
176. Sepucha, K. R. & Scholl, I. Measuring shared decision making: a review of constructs, measures, and opportunities for cardiovascular care. *Circ. Cardiovasc. Qual. Outcomes* **7,** 620–626 (2014).
177. Scholl, I. *et al.* Measurement of shared decision making – a review of instruments. *Z. Für Evidenz Fortbild. Qual. Im Gesundheitswesen* **105,** 313–324 (2011).
178. Brehaut, J. C. *et al.* Validation of a decision regret scale. *Med Decis Mak.* **23,** 281–292 (2003).
179. Elwyn, G. *et al.* Developing CollaboRATE: A fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient Educ. Couns.* **93,** 102–107 (2013).
180. Dolan, J. G. A method for evaluating health care providers' decision making: the Provider Decision Process Assessment Instrument. *Med Decis Mak.* **19,** 38–41 (1999).
181. Barr, P. J. *et al.* The psychometric properties of Observer OPTION(5), an observer measure of shared decision making. *Patient Educ. Couns.* **98,** 970–976 (2015).
182. Culyer, A. J. Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use. *Health Econ Policy Law* **11,** 15–32 (2016).
183. Ryan, M., Kinghorn, P., Entwistle, V. A. & Francis, J. J. Valuing patients' experiences of healthcare processes: Towards broader applications of existing methods. *Soc. Sci. Med.* **106,** 194–203 (2014).

184. Higgins, A., Barnett, J., Meads, C., Singh, J. & Longworth, L. Does convenience matter in health care delivery? A systematic review of convenience-based aspects of process utility. *Value Health* **17,** 877–887 (2014).
185. Wright, D. R., Wittenberg, E., Swan, J. S., Miksad, R. A. & Prosser, L. A. Methods for measuring temporary health States for cost-utility analyses. *PharmacoEconomics* **27,** 713–723 (2009).
186. Versteegh, M. M. & Brouwer, W. B. F. Patient and general public preferences for health states: A call to reconsider current guidelines. *Soc. Sci. Med.* **165,** 66–74 (2016).
187. Krahn, M. *et al.* Patient and community preferences for outcomes in prostate cancer: implications for clinical policy. *Med. Care* **41,** 153–164 (2003).
188. Noel, C. W. *et al.* Comparison of health state utility measures in patients with head and neck cancer. *JAMA Otolaryngol Head Neck Surg* **141,** 696–703 (2015).
189. Peeters, Y. & Stiggelbout, A. M. Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value Health* **13,** 306–309 (2010).
190. Stiggelbout, A. M. & de Vogel-Voogt, E. Health state utilities: a framework for studying the gap between the imagined and the real. *Value Health* **11,** 76–87 (2008).
191. Zethraeus, N. & Johannesson, M. A comparison of patient and social tariff values derived from the time trade-off method. *Health Econ.* **8,** 541–545 (1999).
192. Harrison, M., Milbers, K., Hudson, M. & Bansback, N. Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. *BMJ Open* **7,** e014719 (2017).
193. Fontana, M. *et al.* Patient-accessible tool for shared decision making in cardiovascular primary prevention balancing longevity benefits against medication disutility. *Circulation* **129,** 2539–2546 (2014).
194. Brazier, J. & Tsuchiya, A. Improving cross-sector comparisons: going beyond the health-related QALY. *Appl. Health Econ. Health Policy.* **13,** 557–565 (2015).
195. Tinelli, M., Ryan, M. & Bond, C. What, who and when? Incorporating a discrete choice experiment into an economic evaluation. *Health Econ. Rev.* **6,** 31 (2016).
196. Sanders, G. D. *et al.* Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA* **316,** 1093–1103 (2016).
197. Al-Janabi, H. *et al.* Measuring health spillovers for economic evaluation: a case study in meningitis. *Health Econ.* **25,** 1529–1544 (2016).
198. Wildman, J., McMeekin, P., Grieve, E. & Briggs, A. Economic evaluation of integrated new technologies for health and social care: Suggestions for policy makers, users and evaluators. *Soc. Sci. Med.* **169,** 141–148 (2016).

199. Edwards, A. & Elwyn, G. *Shared Decision-making in Health Care: Achieving Evidence-based Patient Choice*. (Oxford University Press, 2009).
200. Drake, R. E. & Deegan, P. E. Shared decision making is an ethical imperative. *Psychiatr. Serv.* **60,** 1007–1007 (2009).
201. Whitney, S. N. A new model of medical decisions: exploring the limits of shared decision making. *Med Decis Mak.* **23,** 275–280 (2003).
202. Say, R., Murtagh, M. & Thomson, R. Patients' preference for involvement in medical decision making: A narrative review. *Patient Educ. Couns.* **60,** 102–114 (2006).
203. Goranitis, I. *et al.* Maximizing health or sufficient capability in economic evaluation? A methodological experiment of treatment for drug addiction. *Med Decis Mak.* **37,** 498–511 (2017).
204. Ryan, M. Discrete choice experiments in health care. *BMJ* **328,** 360–361 (2004).
205. University of York, Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews. (2001). Available at: http://www.york.ac.uk/inst/crd/crdreports.htm. (Accessed: 21st July 2017)
206. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* **6,** e1000100 (2009).
207. Vass, C., Gray, E. & Payne, K. Discrete choice experiments of pharmacy services: a systematic review. *Int. J. Clin. Pharm.* **38,** 620–630 (2016).
208. de Bekker-Grob, E. W., Ryan, M. & Gerard, K. Discrete choice experiments in health economics: a review of the literature. *Health Econ.* **21,** 145–172 (2012).
209. Joy, S. M., Little, E., Maruthur, N. M., Purnell, T. S. & Bridges, J. F. P. Patient preferences for the treatment of type 2 diabetes: a scoping review. *PharmacoEconomics* **31,** 877–892 (2013).
210. Akkazieva, B., Gulacsi, L., Brandtmuller, A., Péntek, M. & Bridges, J. F. P. Patients' preferences for healthcare system reforms in Hungary: a conjoint analysis. *Appl. Health Econ. Health Policy.* **5,** 189–198 (2006).
211. Mekonnen, A. B. & Enqusselasie, F. Patients' preferences for attributes related to health care services at hospitals in Amhara Region, northern Ethiopia: a discrete choice experiment. *Patient Pref Adherence* **10,** 1293–301 (2015).
212. Cheraghi-Sohi, S. *et al.* What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. *Ann. Fam. Med.* **6,** 107–115 (2008).
213. Davison, S. N., Kromm, S. K. & Currie, G. R. Patient and health professional preferences for organ allocation and procurement, end-of-life care and organization of care for patients with chronic kidney disease using a discrete choice experiment. *Nephrol. Dial. Transplant.* **25,** 2334–2341 (2010).

214. Gidman, W., Elliott, R., Payne, K., Meakin, G. H. & Moore, J. A comparison of parents and pediatric anesthesiologists' preferences for attributes of child daycase surgery: a discrete choice experiment: comparison of parents and anesthesiologists' preferences. *Pediatr. Anesth.* **17,** 1043–1052 (2007).
215. Groenewoud, S. *et al.* What influences patients' decisions when choosing a health care provider? Measuring preferences of patients with knee arthrosis, chronic depression, or alzheimer's disease, using discrete choice experiments. *Health Serv Res* **50,** 1941–1972 (2015).
216. Hendrix, M., Pavlova, M., Nieuwenhuijze, M. J., Severens, J. L. & Nijhuis, J. G. Differences in preferences for obstetric care between nulliparae and their partners in the Netherlands: a discrete-choice experiment. *J. Psychosom. Obstet. Gynecol.* **31,** 243–251 (2010).
217. Hundley, V., Ryan, M. & Graham, W. Assessing women's preferences for intrapartum care. *Birth* **28,** 254–263 (2001).
218. Kessels, R., Van Herck, P., Dancet, E., Annemans, L. & Sermeus, W. How to reform western care payment systems according to physicians, policy makers, healthcare executives and researchers: a discrete choice experiment. *BMC Health Serv Res* **15,** 191 (2015).
219. Krucien, N., Le Vaillant, M. & Pelletier-Fleury, N. What are the patients' preferences for the Chronic Care Model? An application to the obstructive sleep apnoea syndrome. *Health Expect. Int. J. Public Particip. Health Care Health Policy* **18,** 2536–2548 (2015).
220. Longo, M. F. *et al.* Involving patients in primary care consultations: assessing preferences using discrete choice experiments. *Br J Gen Pr.* **56,** 35–42 (2006).
221. Longworth, L., Ratcliffe, J. & Boulton, M. Investigating women's preferences for intrapartum care: home versus hospital births. *Health Soc. Care Community* **9,** 404–413 (2001).
222. Pavlova, M., Hendrix, M., Nouwens, E., Nijhuis, J. & van Merode, G. The choice of obstetric care by low-risk pregnant women in the Netherlands: Implications for policy and management. *Health Policy* **93,** 27–34 (2009).
223. Peacock, S. *et al.* A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services. *Br. J. Cancer* **95,** 1448–1453 (2006).
224. Rischatsch, M. & Zweifel, P. What do physicians dislike about managed care? Evidence from a choice experiment. *Eur. J. Health Econ.* **14,** 601–613 (2013).
225. Salkeld, G., Solomon, M., Butow, P. & Short, L. Discrete-choice experiment to measure patient preferences for the surgical management of colorectal cancer. *Br. J. Surg.* **92,** 742–747 (2005).
226. Schellings, R. *et al.* The development of quality indicators in mental healthcare: a discrete choice experiment. *BMC Psychiatry* **12,** 1 (2012).

227. Scuffham, P. A., Whitty, J. A., Taylor, M. & Saxby, R. C. Health system choice. *Appl. Health Econ. Health Policy.* **8,** 89–97 (2010).
228. Tinelli, M. *et al.* Decision-making criteria among European patients: exploring patient preferences for primary care services. *Eur. J. Public Health* **25,** 3–9 (2015).
229. Haaren-ten Haken, T. *et al.* Eliciting preferences for key attributes of intrapartum care in the Netherlands. *Birth* **41,** 185–194 (2014).
230. van Helvoort-Postulart, D., Dellaert, B. G. C., van der Weijden, T., von Meyenfeldt, M. F. & Dirksen, C. D. Discrete choice experiments for complex health-care decisions: does hierarchical information integration offer a solution? *Health Econ.* **18,** 903–920 (2009).
231. Vick, S. & Scott, A. Agency in health care. Examining patients' preferences for attributes of the doctor–patient relationship. *J. Health Econ.* **17,** 587–605 (1998).
232. Watson, V., Carnon, A., Ryan, M. & Cox, D. Involving the public in priority setting: a case study using discrete choice experiments. *J. Public Health* **34,** 253–260 (2012).
233. Huppelschoten, A. G. *et al.* The monetary value of patient-centred care: results from a discrete choice experiment in Dutch fertility care. *Hum. Reprod. Oxf. Engl.* **29,** 1712–1720 (2014).
234. Mühlbacher, A. C., Bethge, S., Reed, S. D. & Schulman, K. A. Patient preferences for features of health care delivery systems: a discrete choice experiment. *Health Serv Res* **51,** 704–27 (2016).
235. Rechel, B. *et al.* Public reporting on quality, waiting times and patient experience in 11 high-income countries. *Health Policy* **120,** 377–383 (2016).
236. Trenaman, L. *et al.* Decision aids for patients considering total joint replacement: a cost-effectiveness analysis alongside a randomised controlled trial. *Osteoarthr Cartil* **25,** 1615–22 (2017).
237. National Health Service - England. NHS Choices: Find knee replacement services. (2016). Available at: http://www.nhs.uk/Service-Search/Knee-replacement/LocationSearch/1378. (Accessed: 27th September 2016)
238. Lancaster, K. J. A New Approach to Consumer Theory. *J. Polit. Econ.* **74,** 132–157 (1966).
239. Healthwise. Arthritis: Should I Have Knee Replacement Surgery? Available at: http://www.healthwise.net/cochranedecisionaid/Content/StdDocument.aspx?DOCHWID=uh1514. (Accessed: 11th April 2018)
240. Helter, T. M. & Boehler, C. E. H. Developing attributes for discrete choice experiments in health: a systematic literature review and case study of alcohol misuse interventions. *J. Subst. Use* **21,** 662–668 (2016).
241. Damman, O. C., Spreeuwenberg, P., Rademakers, J. & Hendriks, M. Creating compact comparative health care information: what are the key quality attributes to present for cataract and total hip or knee replacement surgery? *Med. Decis. Making* **32,** 287–300 (2012).

242. Victoor, A., Delnoij, D. M. J., Friele, R. D. & Rademakers, J. J. D. J. M. Determinants of patient choice of healthcare providers: a scoping review. *BMC Health Serv. Res.* **12,** 272 (2012).
243. Yahanda, A. T., Lafaro, K. J., Spolverato, G. & Pawlik, T. M. A systematic review of the factors that patients use to choose their surgeon. *World J. Surg.* **40,** 45–55 (2016).
244. Barr, P. J. *et al.* The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. *J. Med. Internet Res.* **16,** e2 (2014).
245. Lange, T. *et al.* Outcome assessment in total knee arthroplasty: a systematic review and critical appraisal. *J. Arthroplasty* **32,** 653-665.e1 (2017).
246. Marshall, D. A. *et al.* How do patients trade-off surgeon choice and waiting times for total joint replacement: a discrete choice experiment. *Osteoarthr Cartil* **26,** 522–530 (2018).
247. Desmeules, F., Dionne, C. E., Belzile, E., Bourbonnais, R. & Frémont, P. The burden of wait for knee replacement surgery: effects on pain, function and health-related quality of life at the time of surgery. *Rheumatol. Oxf. Engl.* **49,** 945–954 (2010).
248. Reed Johnson, F. *et al.* Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health* **16,** 3–13 (2013).
249. Bansback, N., Tsuchiya, A., Brazier, J. & Anis, A. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PloS One* **7,** e31115 (2012).
250. Degner, L. F., Sloan, J. A. & Venkatesh, P. The Control Preferences Scale. *Can. J. Nurs. Res. Rev. Can. Rech. En Sci. Infirm.* **29,** 21–43 (1997).
251. Coast, J. & Horrocks, S. Developing attributes and levels for discrete choice experiments using qualitative methods. *J. Health Serv. Res. Policy* **12,** 25–30 (2007).
252. Coast, J. *et al.* Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ.* **21,** 730–741 (2012).
253. Vass, C. M., Wright, S., Burton, M. & Payne, K. Scale heterogeneity in healthcare discrete choice experiments: a primer. *The Patient* **11,** 167–173 (2018).
254. Nylund, K. L., Asparouhov, T. & Muthén, B. O. Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Struct. Equ. Model. Multidiscip. J.* **14,** 535–569 (2007).
255. Groothuis-Oudshoorn, C. G. M., Flynn, T. N., Yoo, H. I., Magidson, J. & Oppe, M. Key issues and potential solutions for understanding healthcare preference heterogeneity free from patient-level scale confounds. *The Patient* Epub ahead of print (2018).
256. Hole, A. R. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ.* **16,** 827–840 (2007).

257. Day, B. *et al.* Ordering effects and choice set awareness in repeat-response stated preference studies. *J. Environ. Econ. Manag.* **63,** 73–91 (2012).
258. Bateman, I. J., Day, B. H., Dupont, D. P. & Georgiou, S. Procedural Invariance Testing of the One-and-One-Half-Bound Dichotomous Choice Elicitation Method. *Rev. Econ. Stat.* **91,** 806–820 (2009).
259. Brennan, V. K. A study of process utility. (The University of Sheffield, 2016).
260. Birch, S., Melnikow, J. & Kuppermann, M. Conservative versus aggressive follow up of mildly abnormal Pap smears: testing for process utility. *Health Econ.* **12,** 879–884 (2003).
261. Howard, K., Salkeld, G., McCaffery, K. & Irwig, L. HPV triage testing or repeat Pap smear for the management of atypical squamous cells (ASCUS) on Pap smear: is there evidence of process utility? *Health Econ.* **17,** 593–605 (2008).
262. Boye, K. S. *et al.* Utilities and disutilities for attributes of injectable treatments for type 2 diabetes. *Eur. J. Health Econ. HEPAC Health Econ. Prev. Care* **12,** 219–230 (2011).
263. Chancellor, J. *et al.* Preferences of patients with diabetes mellitus for inhaled versus injectable insulin regimens. *PharmacoEconomics* **26,** 217–234 (2008).
264. Cook, J., Richardson, J. & Street, A. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. *Health Econ.* **3,** 157–168 (1994).
265. Forcino, R. C. *et al.* Using CollaboRATE, a brief patient-reported measure of shared decision making: Results from three clinical settings in the United States. *Health Expect. Int. J. Public Particip. Health Care Health Policy* (2017).
266. Health Technology Expert Review Panel: Developing Recommendations on the Optimal Use of Medical Devices, Diagnostic Tests, and Procedures | CADTH.ca. Available at: https://www.cadth.ca/health-technology-expert-review-panel-developing-recommendations-optimal-use-medical-devices. (Accessed: 12th May 2017)
267. Institute for Clinical and Economic Review. Final Value Assessment Framework: Updates for 2017-2019. Available at: https://icer-review.org/material/final-vaf-2017-2019/. (Accessed: 16th February 2018)
268. Baltussen, R. & Niessen, L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff. Resour. Alloc.* **4,** 14 (2006).
269. Thokala, P. & Duenas, A. Multiple criteria decision analysis for health technology assessment. *Value Health* **15,** 1172–1181 (2012).
270. Campillo-Artero, C., Puig-Junoy, J. & Culyer, A. J. Does MCDA trump CEA? *Appl. Health Econ. Health Policy.* **16,** 147–151 (2018).
271. Elwyn, G. *et al.* Supporting shared decision making using an Option Grid for osteoarthritis of the knee in an interface musculoskeletal clinic: A stepped wedge trial. *Patient Educ. Couns.* **99,** 571–577 (2016).

272. Dowsey, M. M. *et al.* Using discrete choice experiments as a decision aid in total knee arthroplasty: study protocol for a randomised controlled trial. *Trials* **17,** 416 (2016).
273. Bansback, N., Trenaman, L., Bryan, S. & Johnson, J. Using routine Patient Reported Outcome Measures to enhance patient decision making: a proof of concept study. *Qual Life Res* **24,** 109 (2015).
274. Ferket, B. S. *et al.* Impact of total knee replacement practice: cost effectiveness analysis of data from the Osteoarthritis Initiative. *BMJ* **356,** j1131 (2017).
275. Montori, V. M., Kunneman, M. & Brito, J. P. Shared decision making and improving health care: the answer is not in. *JAMA* **318,** 617–618 (2017).
276. Alden, D. L., Friend, J., Schapira, M. & Stiggelbout, A. Cultural targeting and tailoring of shared decision making technology: a theoretical framework for improving the effectiveness of patient decision aids in culturally diverse groups. *Soc. Sci. Med. 1982* **105,** 1–8 (2014).
277. Jull, J., Giles, A., Minwaashin Lodge, The Aboriginal Women's Support Centre, Boyer, Y. & Stacey, D. Cultural adaptation of a shared decision making tool with Aboriginal women: a qualitative study. *BMC Med. Inform. Decis. Mak.* **15,** 1 (2015).
278. Nathan, A. G., Marshall, I. M., Cooper, J. M. & Huang, E. S. Use of decision aids with minority patients: a systematic review. *J. Gen. Intern. Med.* **31,** 663–676 (2016).
279. Durand, M.-A. *et al.* Do interventions designed to support shared decision-making reduce health inequalities? A systematic review and meta-analysis. *PloS One* **9,** e94670 (2014).

Appendix 2.1 Study Protocol

Economic evaluation of a decision aid compared to usual care for patients considering joint replacement surgery: follow-up of a multicentre randomised trial

STUDY PROTOCOL

Protocol Summary

Aim	To determine the cost-effectiveness of implementing a patient decision aid, compared to usual care, at orthopedic screening clinics for patients considered candidates for joint replacement surgery
Design	1. 2-year within-trial economic analysis 2. 5-year economic analysis
Intervention	Informed Medical Decisions Foundation decision aid (50-minute video and accompanying booklet) plus surgeon preference report
Comparator	Standard information pamphlet prepared by the hospital which outlines preparation for surgery, recovery after surgery, and discharge plans
Perspective	Health system
Guidelines for Analysis	Economic Evaluation in Clinical Trials – Glick et al. (book) International Society for Pharmacoeconomics and Outcomes Research (ISPOR) – Society for Medical Decision Making (SMDM) Modeling Good Research Practices Person-level costing using administrative data in Ontario - Health System Performance Research Network (electronic)
Guidelines for Reporting	Consolidated Health Economic Evaluation Reporting Standards (CHEERS)
Primary Study Outcomes and Data Sources	**Health care utilization and cost** 1. 2-year within trial economic analysis: patient diary and Ontario average costs for surgery 2. 5-year economic analysis: ICES administrative databases **Quality-of-life** 1. 2-year within trial economic analysis: WOMAC (mapped onto EQ-5D) 2. 5-year economic analysis: WOMAC (mapped onto EQ-5D), modelled beyond trial for 5-years follow-up
Primary Economic Outcomes	**Incremental cost** ($\triangle C$) of patient decision aid (PtDA) strategy compared to usual care (UC); $\triangle C = C_{PtDA} - C_{UC}$ **Incremental effect** ($\triangle E$) in quality-adjusted life-years (QALYs) of patient decision aid strategy compared to usual care; $\triangle E = E_{PtDA} - E_{UC}$ **Incremental cost-effectiveness ratio** (ICER); ICER = $\triangle C / \triangle E$
Discount Rate	5% per CADTH guidelines
Sensitivity Analysis	**Costs** • BC unit costs (as opposed to Ontario) for hip and knee surgeries • Perspective: Health system plus all out of pocket costs incurred by patients **HRQoL** • Using individual preferences (rather than societal) for valuation based on the TTO **Discount rate** • Costs and outcomes undiscounted (0%) • Costs and outcomes discounted at 3% • Costs discounted at 3% and 5% with outcomes undiscounted (0%) **Complete case analysis**

Table of Contents

Tables and Figures

Acronyms

CADTH: Canadian Agency for Drugs and Technology in Health
CACS: Comprehensive Ambulatory Classification System
CCRS: Continuing Care Reporting System
CIHI: Canadian Institutes for Health Information
CMG: Case Mix Group
CMI: Case Mix Index
CPRWPD: Cost per RUG Weighted Patient Day
CPWC: Cost per Weighted Case
DAD: Discharge Abstract Database
Eq-5D: Euroquol 5-Dimension
HCD: Home Care Database
HRPRN: Health System Performance Research Network
HRQoL: Health-Related Quality-Of-Life
ICER: Incremental Cost-Effectiveness Ratio
ICES: Institute of Clinical and Evaluative Sciences
LOS: Length of Stay
MAR: Missing at Random
MCAR: Missing Completely at Random
MNAR: Missing Not at Random
MICE: Multiple Imputation Using Chain Equations
NACRS: National Ambulatory Care Reporting System
NRS: National Rehabilitation Reporting System
OA: Osteoarthritis
ODB: Ontario Drug Benefit (ODB)
OHIP: Ontario Health Insurance Program
OLS: Ordinary Least Squares
PSA: Probabilistic Sensitivity Analysis
PSS: Program and Support Services
QALY: Quality-Adjusted Life-Year
QCH: Queensway-Carleton Hospital
RCT: Randomised Controlled Trial
RF: Raw Food
RIW: Resource Intensity Weight
RPDB: Registered Persons Database
RPG: Rehabilitation Patient Group
RUG: Resource Utilization Groups
TOH: The Ottawa Hospital
TTO: Time Trade-Off
WOMAC: Western Ontario McMaster Osteoarthritis Index

Original study

The original study built on a previous pilot study,(1) and had two complementary components:

1. a multicentre randomised controlled trial (RCT) comparing a patient decision aid plus preference report to usual care to assess their effectiveness based on clinical and decision-related outcomes for individuals considering joint replacement surgery.
2. an economic evaluation to compare the cost-effectiveness of the two strategies as a means of identifying the most efficient provision of care.

Eligible patients who consented to participate in the RCT were randomly assigned to receive a decision aid plus preference report or usual care. Clinical history was taken at baseline, with follow-ups at 6, 12, 18, and 24 months.

Approval from this study was obtained by the research ethics board at The Ottawa Hospital (TOH).

Clinical centres

Patients were recruited from two orthopedic screening clinics at TOH, and Queensway-Carleton Hospital (QCH). Both of these centres are located in Ottawa, Ontario.

Study population

Eligible patients were those with moderate to severe hip or knee osteoarthritis (OA). Exclusion criteria included having inflammatory arthritis, a previous joint arthroplasty surgical consultation, or osteotomy. In addition, those with visual or hearing impairments, and those who were unable to understand English or did not have access to a VCR or DVD player were excluded from the trial. Patients were assessed for eligibility during the screening visit. Patients at TOH were assessed during a 15-minute consultation with a sports medicine physician, while patients at QCH were assessed during a 60-minute visit with an advanced practice physiotherapist or nurse practitioner. All patients were asked to self-report their pain, stiffness, and functioning using the Western Ontario McMaster Osteoarthritis (WOMAC) index. Candidacy for surgery was assessed using the 7-item Western Wait List Hip Knee Priority Tool. If patients were deemed poor candidates for surgery, they were sent back to the referring physician. Those consenting to participate in the study were randomised to one of two strategies.

Health technology strategies being compared

Decision aid plus surgeon preference report

Individuals allocated to the *decision aid plus surgeon preference report* strategy received a copy of the decision aid which was developed by the Informed Medical Decisions Foundation. It consists of a 50-minute video and accompanying booklet. There were unique videos for those considering hip and knee surgeries. Approximately 2-weeks after the clinical visit, participants were contacted by a study coordinator. Their knowledge, values, preferred treatment choice, and decisional conflict were assessed. These findings were combined with patients' clinical assessment to create a one-page preference report that was placed in the patient's file for the surgeon.

Usual care

Those in the *usual care* strategy received a standard information pamphlet prepared by the hospital which outlines preparation for surgery, recovery after surgery, and discharge plans. No written information was provided on benefits and harms of surgery or alternative options that could support decision-making. The clinical information collected during the screening visit was compiled into a half-page summary and placed in the patient's file for the surgeon.

Findings

The typical patient was 66 years old, retired, living with someone, and about half had high school education or less. Median time from referral to off the wait list was 118 days (95% CI: 109-140) in the intervention and 144 days (95% CI: 121-164) in the control arm (log-rank p=0.0691). Surgery rates were 73.2% in the intervention strategy, and 80.5% in the control (RR= 0.91, 95% CI 0.81-1.03) with 12 intervention (7.3%) and 8 control participants (4.9%) returning to have surgery within 2 years (p=0.79). Decision quality was defined as "the extent to which patients' decisions were informed and values congruent with their choice" Decision quality was considered reached if a patient scored >66% on the knowledge test and if their predicted probability of surgery based on values corresponded with their actual choice. Based on this definition, decision quality was reached by 56.1% in the intervention group and 44.5% in the control (RR=1.23; 95% CI 1.00-1.53).

Economic evaluation

Participants were recruited between May 2008 and October 2009, with the trial following all individuals for two-years after enrolment. Trial results (outlined above), are under review, and indicate that at two years follow-up fewer people in the decision aid arm underwent surgery, though this difference was not significantly different. Information on the cost-effectiveness of this intervention can help describe whether a patient decision aid provides value for money in patients considering joint replacement surgery. The economic evaluation will include two initial analyses including a two-year within trial economic analysis, and a five-year economic analysis which will use linked ICES data and model the quality of life outcomes beyond the trial period. These two analyses will use the same guidelines for analysis and reporting and principal study outcomes, as outlined below.

Guidelines for analysis and reporting

Analysis will be based on guidance from three primary sources:

1. Economic Evaluation in Clinical Trials, (2)
2. ISPOR-SMDM modeling best practices guidelines, (3–5)
3. Health System Performance Research Network guidelines on person level costing using Ontario administrative data (6)

Reporting of the economic analysis will follow the Consolidated Health Economic Evaluation Reporting Standards (CHEERS).(7) (see Appendix 1)

The principal study outcome measures

The primary outcomes for measuring the differences in effect between the patient decision aid plus preference report and usual care are:

- Health system costs including all those related to the participants affected joint
- Health-related quality-of-life (HRQoL) – measured using the WOMAC, and converted to EQ-5D health utilities

All analyses will be from a health systems perspective, which will take into account costs incurred in the health care system, in addition to out-of-pocket medication and treatment costs incurred by patients.

1. Two-year within trial economic analysis

Methods

Health care resource use

Data on health-care resource use were collected prospectively. During the trial, patients were asked to provide information whether they had undergone joint replacement, the number of doctor visits, pharmaceuticals, physiotherapy and other extended health services, and other expenses, such as parking while attending medical appointments. In all cases it was specified that the costs should be for the "joint problem". Diaries were collected at 6, 12, 18, and 24 months.

Costs

The cost for each individual patient in trial will be calculated by multiplying his or her use of health-care resources by the associated unit costs. Average Ontario unit costs will be used because the primary objective of the study is to evaluate the incremental cost and benefit of the intervention under question, not the differences in hospitals, where provider-specific costs may differ based on fixed costs associated with activities such as teaching and hospital size.

Total costs will include the costs of surgery, GP visits, hospital admissions and medication.

Quality-adjusted life-years

Health outcomes will be expressed in QALYs. During the trial, HRQoL was assessed at baseline, 6, 12, 18, and 24 months using the WOMAC. The WOMAC can be mapped to the EQ-5D using a methodology outlined by Xie et al.(8) The EQ-5D is a standardised and validated generic instrument that measures HRQoL. It has five dimensions, including mobility, self-care, usual activities, pain and discomfort, and anxiety and depression.(9) It provides a measure of HRQoL on a scale from 0 (equivalent to death) to 1 (perfect health).

Mapping was completed by using regression models with OLS estimators to establish the mapping function, where the WOMAC was represented as explanatory variables in four ways, 1) total score; 2) domain scores; 3) domain scores plus pair-wise interaction terms to account for possible nonlinearities; and 4) individual item scores. They found that the OLS estimator and the WOMAC domain scores as explanatory variables had the best fit and was chosen by the preferred mapping model.

QALYs for each patient will be calculated as the area under the curve following the trapezium rule, which assumes linear interpolation between follow-up points.(10) Incremental mean QALYs between the two policies under consideration will be estimated with and without adjustment for baseline utility, using OLS regression.(11)

Discounting

Costs and outcomes in year two will be discounted using a 5% annual discount rate, as per Canadian Agency for Drugs and Technology in Health (CADTH) guidelines.(12) The influence of the discount rate will be explored in a sensitivity analysis, where it will be varied to 0% and 3%.

Missing data and multiple imputation

Missing data is a concern. For the economic evaluation, this will include WOMAC scores (and TTO values), that will be used to determine the 'benefit' provided in QALYs. In addition, there are some missing cost data from out-of-pocket payments, such as extended health care use, physiotherapy visits, and medication.

In order to determine the best method to address missing data, we will first try to identify which type of missing data we have. There are three types, including data that is missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).(13,14)

- MCAR: the missing data are the same, except for random variation, from the data that is complete
- MAR: the data that are missing differ from the complete data in predictable ways, meaning that after adjustment for observable differences, they are the same, except for random variation
- MNAR: the data that are missing differ from the complete data in unpredictable ways, meaning that the two are not the same

While there are no formal ways to distinguish between the different types of missing data, there are some exploratory tests that can provide insight. For example, if the clinical and demographic characteristics of participants with censored data differ from those without missing data, then we can rule out MCAR (because it is not missing at random). In medical trials, missing data is often MAR, in which case multiple imputation is the technique preferred to overcome this limitation.

Multiple imputation presents a number of important advantages over standard methods for addressing missing data. It makes full use of all available data, it incorporates uncertainty associated with missing data, and it ensures unbiased estimates and standard errors. Broadly speaking, multiple imputation follows three steps: (13)

1. regression models are used to predict plausible values for the missing observations from the observed values, the missing data are filled in *m* times to generate *m* complete datasets
2. the *m* complete datasets are analysed independently using complete-case methods
3. the results from the *m* complete datasets are combined to generate mean estimates of costs, QALYs, variance, and confidence intervals using Rubin's rules.

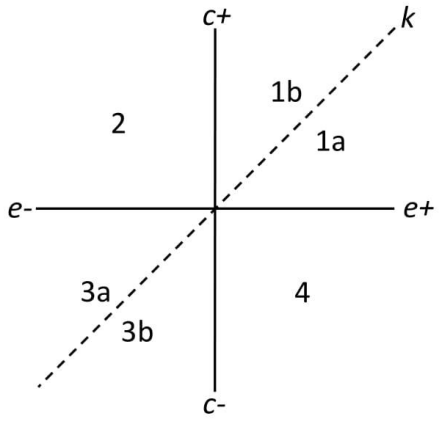
Owing to the fact that in this trial, there are missing data for both costs (from patient questionnaires), and WOMAC scores that are used to derive the measure of benefit (QALYs), multiple imputation using chain equations (MICE) will be employed. This involves individually predicting each variable using its own regression model.(13)

Given the extent of missing data, the multiple imputed data will be presented as the base case, with the complete-case analysis explored in a sensitivity analysis.

Incremental analysis

The cost-effectiveness of the decision aid plus preference report will be evaluated by comparing the costs and QALYs incurred with those in the usual care arm at 2-years of follow-up, using conventional decision rules and estimating ICERs as appropriate. If one intervention is associated with greater mean QALYs and lower mean costs it is deemed cost-effective by dominance. The ICER is calculated if either treatment arm does not dominate. The ICER summarises the additional costs associated with one intervention over the other and relates this to the additional benefits.

Figure 1. Cost-effectiveness plane, adapted from Black et al.(15)



Cost-effectiveness results are often presented on the 'cost-effectiveness plane' which provides a visual representation of the incremental costs and effectiveness of the two strategies being compared (the ICER). The plane is divided into four quadrants (see figure 1), and denote incremental differences in costs and effect between the technology under consideration and the comparator (often usual care). The four quadrants and associated decision rules are demonstrated in table 1.

Table 1. Quadrants of cost-effectiveness plane, decision rules, and adoption decision

	Compared to Usual Care			
Quadrant	Cost	Effect (QALYs)	Decision Rule	Adoption Decision

1a	Increased	Increased	Threshold	Adopt
1b	Increased	Increased	Threshold	Do not adopt
2	Increased	Decreased	Dominance	Do not adopt
3a	Decreased	Decreased	Threshold	Do not adopt
3b	Decreased	Decreased	Threshold	Adopt
4	Decreased	Increased	Dominance	Adopt

Decision-making in quadrants 2 and 4 is quite straightforward. Technologies with an ICER in quadrant 2 would not be adopted on dominance because they cost more and are less effective than current practice. Conversely, technologies with an ICER in quadrant IV would be adopted on dominance because they cost less and are more effective than current practice.

Decision-making in quadrants 1 and 3 is more difficult. Technologies with an ICER in quadrant 1 are more costly and more effective. In this case decision-makers need to decide whether the increased effectiveness justifies the cost. The decision-makers threshold is denoted by *k*, which is the maximum they are willing to pay for the benefit. If the ICER falls below this threshold (1a) then the decision would be to adopt. If the ICER is above the threshold (1b), the technology would not be adopted.

In quadrant 3 the new technology is less effective and costs less. In this case it may be worthwhile to accept the reduced effectiveness if the costs saved could be used to gain greater benefit elsewhere. Thus, ICERs falling under the threshold (3b) would be adopted because the costs saved could be allocated to other technologies and thereby generate greater effect. ICERs above the threshold (3a) would not be adopted because the costs saved could not generate the same benefit if allocated to other technologies.

Bootstrapping to represent uncertainty in the ICER

In order to estimate a confidence interval around the ICER, bootstrapping will be used following guidance from Briggs et al..(16) This process will be carried out as follows:

1. sample with replacement cost/effect pairs from the decision aid plus preference report group and use these to calculate cost (C_{PtDA}) and effect (E_{PtDA}) estimates
2. sample with replacement cost/effect pairs from the usual care group and use these to calculate cost (C_{UC}) and effect (E_{UC}) estimates
3. calculate the bootstrap estimate of the ICER according to the following formula:

$$ICER = \frac{\triangle C}{\triangle E} = \frac{C_{PtDA} - C_{UC}}{E_{PtDA} - E_{UC}}$$

4. Repeat the above process 50-200 times to give a bootstrap estimate of the ICER statistic

These results will be presented on the cost-effectiveness plane as a scatter plot, and as a cost-effectiveness acceptability curve.

Sensitivity analysis

In order to understand the influence of various factors on the outcomes of the analysis, sensitivity analysis will be undertaken, as follows:

- Costs
  - BC unit costs (as opposed to Ontario) for hip and knee surgeries
  - Perspective: Health system plus all out of pocket costs incurred by patients
- HRQoL
  - Using individual preferences (rather than societal) for valuation based on the TTO
- Discount rate
  - Costs and outcomes undiscounted (0%)
  - Costs and outcomes discounted at 3%
  - Costs discounted at 3% and 5% with outcomes undiscounted (0%)
- Complete case analysis: only individuals with complete data will be used for this analysis

2. Five-year economic analysis

In the two-year analysis, cost data was derived from patient diaries, collected at 6, 12, 18, and 24 months follow-up. Administrative data for up to 5-year follow-up is also available to derive costs, and, when combined with the two-year economic analysis, will be able to assess the robustness of the conclusions. Ultimately, these two analyses will help determine whether a lifetime model is required.

Methods

Health care resource use

Trial data will be linked to administrative data on health-care utilization held by the Institute of Clinical and Evaluative Sciences (ICES). Databases queried for the analysis will include the Discharge Abstract Database (CIHI-DAD), National Ambulatory Care Reporting System (NACRS), Continuing Care Reporting System (CCRS), and National Rehabilitation Reporting System (NRS) held by the Canadian Institutes for Health Information (CIHI), in addition to the Ontario Drug Benefit (ODB) Program, Ontario Health Insurance Program (OHIP), Home Care Database (HCD), and the Registered Persons Database (RPDB) (see appendix 1). Person-level utilization and costing will be accomplished following guidelines set out by the Health System Performance Research Network (HSPRN),(6) outlined below.

Utilization intensity and case-mix adjustment

Even among individuals with the same diagnosis, there is variability in disease severity. As a consequence, there is variability in resource utilization (and subsequent cost). In order to account for this variation, which can be the result of a variety of factors, including age, sex, or presence of co-morbid conditions, cases will be costed based on weighting that takes into account the amount of resources used (including administrative, staff, supplies, and technology) relative to other patients. Derivation of resource utilization for each of the databases will be accomplished as follows:

DAD, NACRS: All discharges are classified with as a case-mix group (CMG) in the DAD or Comprehensive Ambulatory Classification System (CACS), both of which can be used to derive a Resource Intensity Weight (RIW).

NRS: Discharges are associated with a Rehabilitation Patient Group (RPG) case mix classification methodology and weighting system, which corresponds with a unique RCW. Unlike RIW from the DAD and NACRS, RCW are not provided by the NRS and will be calculated or assigned.

CCRS: Weighting in the CCRS is through Resource Utilization Groups (RUG or RUG-II), with each RUG associated with a Case Mix Index (CMI) which corresponds to the weight of a daily cost of care relative to the average use among the population.

HCD: Each record in the HCD database represents a single visit or service received, with the exception of select services, such as respite and nursing, which are recorded by hours received.

Costs

DAD: Provincial average Cost per Weighted Case (CPWC) will be multiplied by a person's RIW. Annual CPWC are readily available.

NACRS: Will be calculated by multiplying RIW weights by provincial cost per CACS weighted case (CPCACSC). The CPWC will be calculated by dividing the total provincial costs by the total number of weighted cases.

NRS: Will be calculated by multiplying RCW by CPWC. CPWC are available for years 2002 to 2009.

CCRS: Complex continuing care costs will be calculated as the sum of CMI multiplied by cost per RUG weighted patient day (CPRWPD) and by length of stay (LOS). Long-term care costs will be calculated as the sum of nursing and personal care costs multiplied by CMI, plus program and support services (PSS), raw food (RF), and other accommodation, multiplied by length of stay.

OHIP: Cost of physician and laboratory services will be taken directly from the OHIP database, as they can be directly linked to individuals where fee-for-service billing is used. In settings where other payment models are used, such as capitation, salary, or alternative payment plans, an alternative approach will be used.

HCD: Will be calculated by multiplying the cost per visit by the number of visits, or the cost per hour multiplied by the total number of hours.

ODB: Prescription drug payments paid to pharmacies for dispensing medications are recorded in the ODB.

Table 2. Health service, associated database, healthcare utilization and cost

Type of Service	Utilization Database	Utilization Unit	Weight Name	Unit Cost
Group 1: Short Episodes (<60 days)				
Acute Hospitalization	DAD	Weighted Case	RIW	CPWC
Same Day Surgery & Outpatient	NACRS	Weighted Case	RIW	CPWC
Emergency Department	NACRS	Weighted Case	RIW	CPWC
Inpatient Rehabilitation	NRS	Weighted Case	RCW	CPWC

Group 2: Longer Term Episodes				
Complex Continuing Care	CCRS	Weighted Day	CMI	CPRWPD
Long Term Care Home	CCRS	Weighted Day	CMI	CPRWPD
Group 3: Visits/Claims				
Primary and Specialist Physician	OHIP	Visit	n/a	Fee Paid
Home Care	HCD	Visit	n/a	Cost/Visit
Pharmaceuticals	ODB	Prescription	n/a	Amount Paid

Wodchis et al., 2013

As with the two-year economic analysis, the base-case will only consider osteoarthritis-related costs. These include surgery, complications due to surgery, analgesic medications, joint –related visits to and from the GP and hospital inpatient, outpatient and day visits. In order to determine which healthcare utilization and subsequent costs are related to osteoarthritis, a rheumatologist (blinded) will review all relevant codes in the data (e.g., diagnosis codes in OHIP, drug identification number in ODB). Codes will be assigned to one of two categories: 1) OA-related, and 2) non-OA-related. Only those categorised as OA-related will be used in the base-case analysis, while all costs will be investigated in a sensitivity analysis. Costs will be adjusted at baseline, using 6-months costs prior to entry into the study (for those with this data available). Adjustment will be done using OLS regression.

From the administrative databases, the primary concern is that the OBD only includes individuals age 65 and older, or those on disability. As the mean age of both the control (66.9 years, SD=9.8) and intervention (66.1 years, SD=9.8) are around this cut-off point, a number of patients will be missing data on prescription medication costs which are paid for out-of-pocket, or through extended health insurance.

Quality-adjusted life-years

For the five-year evaluation, it will be required to model quality of life beyond the two-year RCT follow-up period. This will involve creating a number of unique health states that correspond to six-month periods as in the original trial. Based on health care utilization, individuals will be into one of this health states, and their quality of life will be imputed based on the corresponding population experiencing that health state during the two-year follow-up. For example, if an individual undergoes a surgery at 3 years, their quality of life during that six-month period will be based on similar individuals who underwent a surgery during the RCT follow-up period. QALYs will be calculated in the same manner as the two-year analysis.

Discounting

Costs and outcomes in years 2-5 will be discounted using a 5% annual discount rate, as per Canadian Agency for Drugs and Technology in Health (CADTH) guidelines.(12) The influence of the discount rate will be explored in a sensitivity analysis.

Incremental analysis

The incremental analysis will be completed in the same manner as the two-year analysis, including through the use of bootstrapping to represent uncertainty around the ICER.

Sensitivity Analysis

The sensitivity analysis will be completed in the same manner as the two-year analysis.

3. Exploring the need for a long-term model

Based on the results of the 1) two-year economic analysis, and 2) five-year economic analysis, there may be sufficient uncertainty remaining to justify the development of a long-term model to fully elucidate the economic impact of the patient decision aid. This decision will be informed on the incremental differences and trends observed in costs and quality-of-life between the two strategies at follow-up.

If there is sufficient uncertainty in the conclusions to justify modelling over the long-term, modelling will begin by reviewing past economic models of hip and knee replacement.

References

1. Stacey D, Hawker G, Dervin G, Tugwell P, Boland L, Pomey M-P, et al. Decision aid for patients considering total knee arthroplasty with preference report for surgeons: a pilot randomized controlled trial. BMC Musculoskelet Disord. 2014;15(1):54.

2. Glick H, Doshi J, Sonnad S, Polsky D. Economic evaluation in clinical trials, volume 2. Oxford, UK: Oxford University Press; 2014.

3. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2012 Oct;15(6):843–50.

4. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD, et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Mak Int J Soc Med Decis Mak. 2012 Oct;32(5):722–32.

5. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Med Decis Mak Int J Soc Med Decis Mak. 2012 Oct;32(5):678–89.

6. Wodchis W, Bushmeneva K, Nikitovic M, McKillip I. Guidelines on person-level costing using administrative databases in Ontario [Internet]. Health System Performance Research Network; 2013. Available from: http://www.hsprn.ca/uploads/files/Guidelines_on_PersonLevel_Costing_May_2013.pdf

7. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMC Med. 2013 Mar 25;11(1):80.

8. Xie F, Pullenayegum EM, Li S-C, Hopkins R, Thumboo J, Lo N-N. Use of a disease-specific instrument in economic evaluations: mapping WOMAC onto the EQ-5D utility index. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2010 Dec;13(8):873–8.

9. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001 Jul;33(5):337–43.

10. Fitzgerald P, Goodacre SW, Cross E, Dixon S. Cost-effectiveness of point-of-care biomarker assessment for suspected myocardial infarction: the randomized assessment of treatment using panel Assay of cardiac markers (RATPAC) trial. Acad Emerg Med Off J Soc Acad Emerg Med. 2011 May;18(5):488–95.

11. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ. 2005;14(5):487–96.

12. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada, 3rd edition. 2006.

13. Grant AM, Boachie C, Cotton SC, Faria R, Bojke L, Epstein DM, et al. Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease:

5-year follow-up of multicentre randomised trial (the REFLUX trial). Health Technol Assess Winch Engl. 2013 Jun;17(22):1–167.

14. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Sep 1;338(jun29 1):b2393–b2393.

15. Black WC. The CE Plane A Graphic Representation of Cost-Effectiveness. Med Decis Making. 1990 Aug 1;10(3):212–4.

16. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. Health Econ. 1997;6(4):327–40.

Appendix 1: CHEERS Checklist

Section/item	Item No	Recommendation
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.
Introduction		
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions
Methods		
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.
Measurement of effectiveness	11a	*Single study-based estimates*: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.
	11b	*Synthesis-based estimates*: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.
Estimating resources and costs	13a	*Single study-based economic evaluation*: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
	13b	*Model-based economic evaluation*: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
Currency, price date,	14	Report the dates of the estimated resource quantities and unit costs. Describe

and conversion		methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.
Results		
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.
Characterising uncertainty	20a	*Single study-based economic evaluation*: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).
	20b	*Model-based economic evaluation*: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.
Discussion		
Study findings, limitations, generalizability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalizability of the findings and how the findings fit with current knowledge.
Other		
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.

Husereau et al. 2013

Appendix 2: Derivation of resource utilization and costs from Ontario administrative databases

Database	Description	Use for this analysis
OHIP	The insured services covered under OHIP include diagnostic, preventive, and rehabilitation services provided by both generalists and specialists, as well as services provided by community laboratories.	Physician services and laboratory services.
DAD	The Discharge Abstract Database (DAD) captures administrative, clinical and demographic information on hospital discharges (including deaths, sign-outs and transfers). Some provinces and territories also use the DAD to capture day surgery.	Acute inpatient hospitalizations.
ODB	The Ontario Drug Benefit (ODB) program database contains prescription claims data for eligible recipients of the program. The HNS provides on-line prescription claims processing and adjudication for the reimbursement of prescription medications for eligible recipients.	Cost of prescription medications (provided outside of hospital) for those over age 65 or those in provincial disability programs.
NACRS	The National Ambulatory Care Reporting System (NACRS) contains data for all hospital-based and community-based ambulatory care for day surgery, outpatient clinics, and emergency department visits.	Same day surgery, outpatient clinic visits, and ED visits.
CCRS	The Continuing Care Reporting System (CCRS), launched in 2003–2004, contains demographic, clinical, functional and resource utilization information on individuals receiving continuing care services in hospitals or long-term care homes in Canada.	Complex continuing care.
NRS	Rehabilitation care includes a comprehensive range of inpatient and outpatient services and complex continuing care for adults who experience debilitating illness or injury. We facilitate the collection of standardized information on inpatient clients, mainly through the National Rehabilitation Reporting System (NRS).	Inpatient rehabilitation.
HCD	The HCD captures all services provided by or coordinated by Ontario's Community Care Access Centres (CCACs). These organizations were established by the Ministry of Health and Long-Term Care to provide access to government-funded home and community services and long-term care homes.	Home care
RPDB	The Registered Persons Database (RPDB) is a Personal Information Bank that collects organized personal information that can be retrieved using an individual's Health Number.	Demographic information and date of death.
		Wodchis et al., 2013

Appendix 2.2: CHEERS Checklist

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Page 40
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	NA
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Page 40
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Page 42
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Page 42
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Page 42
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Page 42
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Page 42
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 42
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Pages 48 - 50
Measurement of effectiveness	11b	*Single study–based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	NA

Section/item	Item No	Recommendation	Reported on page No/line No
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Page 45
Estimating costs and resources	13b	*Single study–based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs	Page 44
Currency, price date and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Page 44
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	N/A This is a within trial economic evaluation.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	N/A This is a within trial economic evaluation.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Page 46
Results			

Section/item	Item No	Recommendation	Reported on page No/line No
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Table 2.2
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Table 2.3
Characterising uncertainty	20b	*Single study–based economic evaluation:* Describe the effects of sampling uncertainty for estimated incremental cost, incremental effectiveness, and incremental cost-effectiveness, together with the impact of methodological assumptions (such as discount rate, study perspective)	Table 2.4 Figure 2.3 Figure 2.4
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Page 47
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Page 54
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 40

Section/item	Item No	Recommendation	Reported on page No/line No
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	NA

Appendix 2.3: Additional methodological information

Costs

The costs for each individual patient in trial were calculated by multiplying their use of health-care resources by the associated unit costs. The cost of the decision aid arm was calculated based on the time required to compile the preference report (assumed to be 15 minutes of research assistant time at an hourly rate of $25) in addition to the cost of the DVD and booklet (assumed to be $10) and a surgical consultation. Health care resource utilization was captured prospectively using a patient diary. The patient diary asked patients to record the number of visits and dates of visits to healthcare professionals, including: "family doctor", "surgeon", and "other (e.g., physiotherapist)." In all cases, it specified that these visits should be related to their joint problem. Average Ontario unit costs were used for each resource utilization category. Costs for physician/specialist services were based on OHIP billing codes, while nurse visits were based on the 2014 Collective Agreement between hospitals and the nurses' union, at a pay rate for a nurse with 6 years' experience and assuming 13% in lieu of benefits. Non-physician services included primarily physiotherapy and massage therapy and were costed based on a review of Ontario specific websites for a one-hour session.

The patient diary also asked patients to track usage of: "prescription medications", "other treatments", and "other expenses." Given that patients may experience different levels of coverage for prescription medications based on extended health insurance plans

and government assistance, self-reported medication costs were not used in the analysis. Instead, self-reported medication prescriptions were used and multiplied by average Ontario formulary costs. In most cases prescriptions were assumed to be for 30 days, except for opioid-based narcotics and antibiotics (7 days).

Chart review was used to determine if/when patients underwent surgery. Chart review was only available at the two participating hospitals (TOH and QCH). In a small number of cases it was clear that the patient had undergone surgery at another institution. In such cases, it was assumed that patients had surgery at either 6-months post enrollment in the trial (if they had stayed on the waiting list continuously) or 18-months post enrollment (if they had been removed from the waiting list and gone back on it after). The unit cost of hip and knee surgery were taken from the CIHI Patient Cost Estimator for unilateral hip and knee replacement in Ontario for 60-79-year-old patients (2012 CAD$). Costs were inflated to 2014 CAD$ using the health care component of the consumer price index.

If patients reported physician visits, surgeon visits, other visits, or pharmaceutical costs, but did not report the number of visits or prescriptions, it was assumed that they had the median costs for all individuals that did have costs. If patients reported resource utilization for one resource category (e.g., physician visits) but had missing values for others (e.g.,, prescriptions), it was assumed that there were no costs (rather than assuming these values were missing).

Outcomes: QALYs

Health outcomes were expressed in quality-adjusted life-years (QALYs). During the trial, health-related quality-of-life (HRQoL) was assessed at baseline, 6, 12, 18, and 24 months using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). The WOMAC was mapped to the EQ-5D based on an algorithm developed by Wailoo et al. The total number of questions in WOMAC is 24. In cases where there were missing data in the WOMAC, the following rules were applied. If an individual was missing 6 or fewer answers (of the 24) it was assumed that there were 'no issues' for this question. This cut-off was used as analysis found that most with missing values had 5 or fewer. Individuals who had more than six missing answers were assumed to be missing the

Missing data and uncertainty

Missing data was a concern. For the economic evaluation, this included WOMAC scores that were used to calculate QALYs (see below). In addition, there were missing data on health resource utilization, such as physician visits, physiotherapy visits, and medication.

Number (and proportion) of individuals for which follow-up data was available

Visit	QALYs (WOMAC)	Cost (Health Resource Utilization)
Baseline	328 (98%)	N/A
6-months	260 (78%)	260 (78%)
12-months	248 (74%)	248 (74%)
18-months	224 (67%)	225 (67%)
24-months	215 (64%)	215 (64%)

Since there are missing data for both costs and QALYs, multiple imputation using chain equations (MICE) was employed. This involves individually predicting each variable using its own regression model. Predictive mean matching was used (PMM), which imputes an observed value from an individual that is similar based on the predictor characteristics. Categorical costs (e.g., prescription, physician) at period *t* were predicted by costs in other categories and current QoL, lagged costs and QoL from period *t-1* and demographic and clinical characteristics such as joint and gender. QoL was predicted from costs in period *t*, lagged costs and QoL from period *t-1,* and demographic and clinical characteristics.

Appendix 3.1: Sample characteristics

		Decision aid arm (n=161)	Usual care arm (n=163)
Age (yrs), mean (SD)		66.1 (9.8)	67.0 (9.9)
Joint (n)	Hip	45	43
	Knee	116	120
HKPT* (total 80), mean (SD)		45.2 (13.7)	45.5 (13.4)
WOMAC* (total 96), mean (SD)		56.0 (17.2)	52.9 (15.9)
Sex (n)	Men	77	62
	Women	84	101
BMI, mean (SD)		30.8 (6.4)	31.7 (6.1)
Language (n)	English	151	156
	Other	10	7
Education (n)	< HS	10	12
	HS/TS	74	69
	College	31	24
	University	28	39
	Graduate School	18	19
Living	Alone	36	42
arrangement (n)	With someone	125	121
Employment	full time(n)	31	32
	part time (n)	11	12
	retired (n)	101	105
	other(n)	18	14
Household income	<$20,000	14	10
	to $39,999	25	34
	to $59,999	38	35
	to $79,999	33	22
	to $99,999	15	16
	>$100,000	27	31
	no response	9	15

Appendix 6.1: Consent document for think-aloud interviews

Information & Consent Form

Public preferences for arthritis care: A discrete choice experiment

Principal Investigator: Nick Bansback, PhD, Assistant Professor
UBC School of Population and Public Health

Co-Investigators: Stirling Bryan, PhD, Professor
UBC School of Population and Public Health

Logan Trenaman, MSc, PhD Student
UBC School of Population and Public Health

Project Manager:

Welcome to the Public Preference survey

Thank you for your interest in completing the Public Preference survey. Before you decide whether to complete this survey it is important that you understand why the survey is being carried out and what it will involve. Please read the information provided carefully.

Who is conducting the study?
This study is being conducted by researchers at the University of British Columbia.

Who is funding this study?
The study is funded by the Canadian Institutes of Health Research.

Why are we doing this study?
The purpose of this project is to find out which aspects of arthritis care are most important to members of the general public. This information will help policy makers understand how care can be improved.

How is the study done?
Taking part in this study will involve taking part in **one** interview with a researcher and will take about **60 minutes.** The researcher will ask you to go through a survey and talk out loud about how you interpret the survey instructions and questions. The researcher will also take notes and listen to what you say. The interview will be audio recorded to the researcher can listen to it at a later time.

What will happen with the results of the study?
The results of the study will be published in journal articles, and presented at conferences. Nobody will be able to identify you in any reports or publications as only aggregate results be published. If you would like a copy of the results once the study has finished, please contact nick.bansback@ubc.ca

What are the risks of participating?
We do not think you face any risks by participating in this study. You do not have to answer a question if you do not want to. Please tell a research team member if you have any concerns.

What are the benefits of participating?
There are no anticipated benefits to you personally, but the results of the study will be used by policy makers to make better decisions about health care spending.

How will your identity be protected?
Your confidentiality will be respected. No information that discloses your identity will be released or published without your consent unless required by law. All documents will be identified only by code number. Anonymized data will be stored on a network accessible through password protected computers.

Who can I contact if I have a question or concern about the study?

If you have any questions about what we are asking of you, please contact the Principal Investigator or Project Manager. The names and telephone numbers are listed at the top of the first page of this form. If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Participant consent:

Taking part in this study is entirely up to you. You have the right to refuse to participate in this study. If you decide to take part, you may choose to pull out of the study at any time without giving a reason.

- Your signature below indicates that you have received a copy of this consent form for your own records.
- Your signature indicates that you consent to participate in this study.
- Your signature indicates that you consent to have your anonymized quotes published in a PhD thesis, in conference posters, and peer-reviewed manuscripts.

Participant signature: _____

Participant name (please print): _____

Appendix 6.2: Guide for think-aloud interviews

Pilot testing protocol

Guidelines for interviewer:

- Review questionnaire, determine appropriate probes for questions, list additional probes to encourage participants to engage in 'thinking aloud'.
- Before you begin interview/testing, explain process of interview to participant, using examples and exercises to communicate meaning of approach, i.e. how it operates in practice.

Script to guide interview:

- Thank participant for taking part and for their time.
- Briefly describe the aims and purpose of the research and why this pilot testing phase is important.
- Explain how this testing differs from traditional interview/survey testing:

During this interview, you will be working your way through the online survey. However, this process differs from a typical interview or a survey testing session, as I'd like you to read instructions and questions aloud as you go, as well as think aloud, reporting your interpretations and reactions to questions, as well as what details and information you consider in reaching your answer to a question. Our goal with this approach is to get a better idea of how the questions are working and are being interpreted and this will help us to ensure there are not problems around questions being misleading, too vague or confusing. So, by speaking aloud your thought processes during the survey, we'll be able to assess this. For this testing, we are not so much interested in your answer but rather your understanding of the questions and situations described and how you react and think through information to reach an answer.

At times I may stop and ask you more questions about certain questions or items and I may take occasional notes. Also, with your permission, it would be helpful to audio record the testing session.

Please keep in mind that I want to hear all of your opinions and reactions. Don't hesitate to mention if a question is unclear or problematic for whatever reason.

Finally, the online survey should take no longer than 45 minutes to complete. Do you have any questions before we start?

Think-aloud practice exercise

To give you an idea of what the approach is like and help you adapt to it, we'll just start with an example. I'll go first with this one and then you can have a go.

- How many windows are in your house or apartment?
- How did you go about coming up with that answer?
- How difficult was it to arrive at that answer?

Ask if there are any final questions before beginning survey. Remind participant again to think aloud and that interviewer will stay mostly silent during testing.

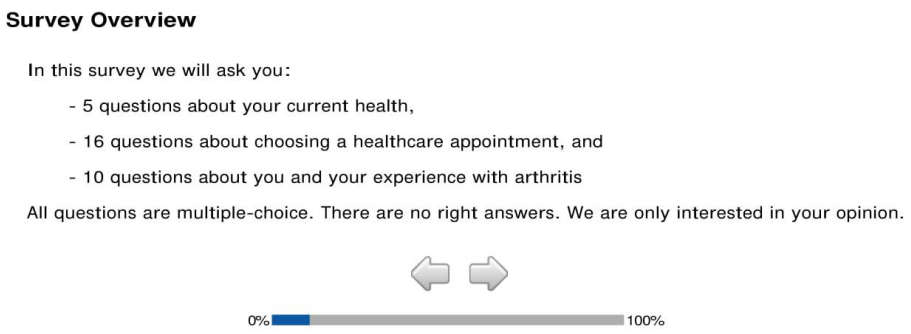
List of general probes:

- How easy is it for you to imagine yourself in this state?
- Feature 1: What do you think of these wait times? Prefer weeks or months or both?
    - How long would you expect to wait to see a specialist?
- Feature 2: What would be the benefit to you of having a specialist that put in 'every' effort compared to 'some' effort?
- Feature 3: What would be the benefit to you of having a specialist that put in 'every' effort compared to 'some' effort?
- Feature 4: What do you think of the chance of pain improvement?
    - Is this better or worse than you would have expected?
    - What do you think happens to those that do not improve?
- DCE question:
    - In your own words, what is this question asking?
    - What does the term 'arthritis specialist' mean to you?
    - What do you consider when answering this question?
    - How did you arrive at your answer?
    - What did you consider in reaching your answer?
    - How difficult was it to arrive at that answer?
- What made you hesitate in giving your answer?
- Can you repeat what you've just read in your own words?

After completing DCE (question 12)

- Consider 2 additional questions provided (for the appropriate scenario)
- Consider alternative scenario
- How do these scenarios compare to you?
- How would this scenario change your answer?

Appendix 6.3: Online DCE survey



About your current health

Under each heading, plese click the ONE that best describes your health TODAY.

Mobility

- ◯ I have no problems in walking about
- ◯ I have some problems in walking about
- ◯ I am confined to bed

Self-care

- ◯ I have no problems with self-care
- ◯ I have some problems washing or dressing myself
- ◯ I am unable to wash or dress myself

Usual activities

(e.g., work, study, housework, family or leisure activities)

- ◯ I have no problems with performing my usual activities
- ◯ I have some problems with performing my usual activities
- ◯ I am unable to perform my usual activities

Pain/discomfort

- ◯ I have no pain or discomfort
- ◯ I have moderate pain or discomfort
- ◯ I have extreme pain or discomfort

Anxiety/depression

- ◯ I am not anxious or depressed
- ◯ I am moderately anxious or depressed
- ◯ I am extremely anxious or depressed

Choosing a healthcare appointment

We want you to imagine you have knee osteoarthritis. Your family doctor has referred you to an arthritis specialist to decide between these options:

- Have knee replacement surgery, or

- Do NOT have knee replacement surgery. Instead you would use other treatments like exercise, weight loss, or medicines.

0% 100%

Key points to remember

- The decision you and the arthritis specialist make depends on your age, health, and activity level, and on how much pain and disability you have.

- Most people have knee replacement only when they can no longer control arthritis pain with medicine and other treatments and when the pain really interferes with their lives.

- Rehabilitation after knee replacement requires daily exercises for several weeks.

- Most knee replacements last for at least 15 years. Some people need to have the knee replaced again.

- If you wait so long to have a knee replacement that you have already lost much of your strength, endurance, and ability to be active, then after the surgery you might have a harder time returning to your normal activities.

Background

You will be asked about appointments with two different arthritis specialists.

For each question, you will be asked to imagine that you are experiencing either moderate OR extreme pain or discomfort while you wait for an appointment.

The appointments will be different. These are the features that will change:

- Your waiting time for the appointment

- Your experience of shared decision-making

- Your chance of improvement in pain or discomfort (we would expect this improvement to last for at least two years)

We will now describe the features of your appointment.

Feature 1 of 3

Waiting time

This is how long you must wait to have an appointment with the arthritis specialist.

The options you will see in this survey will be:

- 4 months
- 6 months
- 8 months

0% 100%

Feature 2 of 3

Shared decision-making

This is how much effort the arthritis specialist puts into:

- Helping you understand your health issues,
- Listening to the things that matter most to you about your health issues, and
- Including what matters most to you in choosing what to do next

The options you will see in the survey will be:

-No effort

-Some effort

-Every effort

Feature 3 of 3

Chance of improvement in pain or discomfort

This is the number of patients treated by the arthritis specialist who improve from moderate OR extreme pain or discomfort to NO pain or discomfort 1 year after the appointment

The options you will see in the survey will be:

- 50 out of 100 patients improve (50%)
- 60 out of 100 patients improve (60%)
- 70 out of 100 patients improve (70%)

Background

We would also like you to assume that some things do not differ between the arthritis specialists.

Both arthritis specialists are from the same hospital.

They have the same influence on their patients' feelings of anxiety/depression and ability to:

- Walk about,
- Engage in self-care, and
- Participate in usual activities.

We will now start the survey. The first part has 7 questions.

When you answer the questions, imagine you are experiencing **Moderate pain**.

Question (1 of 10)

Which arthritis specialist (A or B) would you choose for your appointment to make a treatment decision?

	Arthritis Specialist A	Arthritis Specialist B
Pain while you wait *Please imagine that you are experiencing -*	Moderate pain	Moderate pain
Waiting time *How long you must wait for an appointment.*	8 months	4 months
Shared decision-making *Effort made by the specialist to:* *-help you understand your health issues* *-listen to the things that matter most to you* *-include what matters most to you in choosing what to do next*	No effort	Some effort
Chance of improvement in pain or discomfort *Number of patients who improve to NO pain or discomfort 1 year after the appointment*	60 out of 100 patients improve (60%)	50 out of 100 patients improve (50%)
I would choose:	○	○

0% 100%

We will now ask you some final questions

Please choose one of the following statements that best describes how you like treatment decisions to be made:

- I prefer to make the final treatment decision.
- I prefer to make the final treatment decision after seriously considering my doctor's opinion.
- I prefer that my doctor and I share responsibility for deciding which treatment is best.
- I prefer that my doctor makes the final treatment decision, but seriously considers my opinion.
- I prefer to leave all treatment decisions to my doctor.

What is the highest grade or level of school that you have completed?

- 8th grade or less
- Some high school, but did not graduate
- High school or high school equivalency certificate
- College, CEGEP or other non-university certificate or diploma
- Undergraduate degree or some university
- Post-graduate degree or professional designation

What is your gender?

- Male
- Female
- Other
- Prefer not to say

What is your age?

- 40-44
- 45-49
- 50-54
- 55-59
- 60-64
- 65-69
- 70-74
- 75-79
- 80+

We will now ask you some final questions

Have you ever been diagnosed with arthritis?

- Yes
- No

Do you have a friend or family member who has been diagnosed with arthritis?

- Yes
- No

Have you ever had a hip or knee replacement?

- Yes
- No

Do you have a friend or family member who has had a hip or knee replacement?

- Yes
- No

Thank you for participating! Please click to exit the survey.

Appendix 6.4: Additional DCE analysis

Table A.1: Conditional logit model with dummy variables for scenario order

	β	SE	p
Wait			
4 months	Ref		
6 months	-0.22	0.06	0.00
8 months	-0.67	0.06	0.00
Shared decision-making			
No effort	-0.77	0.06	0.00
Some effort	Ref		
Every effort	0.36	0.06	0.00
Pain			
50%	Ref		
60%	0.82	0.06	0.00
70%	1.51	0.06	0.00
Wait			
4 months * Extreme	Ref		
6 months * Extreme	-0.62	0.10	0.00
8 months * Extreme	-1.13	0.12	0.00
Shared decision-making			
No effort * Extreme	0.34	0.09	0.00
Some effort * Extreme	Ref		
Every effort * Extreme	-0.08	0.10	0.44
Pain			
50% * Extreme	Ref		
60% * Extreme	-0.11	0.10	0.30
70% * Extreme	-0.21	0.10	0.04
Interactions (Order)			
Wait			
4 months * Order	Ref		
6 months * Order	-0.33	0.10	0.00
8 months * Order	-0.40	0.11	0.00
Shared decision-making			
No effort * Order	0.36	0.09	0.00
Some effort * Order	Ref		
Every effort * Order	0.06	0.10	0.52
Pain			
50% * Order	Ref		
60% * Order	0.13	0.10	0.21
70% * Order	0.21	0.10	0.04
Wait			
4 months * Extreme * Order	Ref		
6 months * Extreme * Order	0.66	0.14	0.00
8 months * Extreme * Order	1.01	0.17	0.00
Shared decision-making			
No effort * Extreme * Order	-0.59	0.12	0.00
Some effort * Extreme * Order	Ref		
Every effort * Extreme * Order	0.02	0.14	0.91
Pain			

50% * Extreme * Order	Ref		
60% * Extreme * Order	-0.13	0.15	0.37
70% * Extreme * Order	-0.17	0.14	0.23
Log-likelihood	-9083		
AIC	14,559		
BIC	14,756		
Observations	26,208		

Table A.2: Heteroskedastic conditional logit model*

	β	SE	p
Main effects			
Wait (per month)	-0.28	0.01	0.00
Shared decision-making			
No effort	-0.57	0.04	0.00
Some effort	Ref		
Every effort	0.26	0.04	0.00
Pain (per %)	0.07	0.00	0.00
Scale term (Extreme)	-0.02	0.04	0.61
Log-likelihood	-7412		
Observations	26,208		

* model includes dummy variables to control for order

Table A.3: Latent class models for 2 to 6 classes*

	2 Classes		3 Classes			4 Classes				5 Classes					6 Classes					
	1	2	1	2	3	1	2	3	4	1	2	3	4	5	1	2	3	4	5	6
Wait (per month)	-0.41	-0.07	-0.41	-0.22	-0.13	-0.36	-0.13	-0.31	-0.01	-0.46	-0.68	-0.08	-0.30	9.81	-1.10	-0.79	-0.13	-0.04	-0.30	-1.07
Shared decision-making																				
No effort	-0.45	-0.98	-0.58	-0.08	-2.03	-0.44	0.30	-1.47	-2.23	-0.38	0.46	-0.51	-2.25	-22.35	-0.41	0.61	-0.72	-0.54	-2.28	-14.44
Some effort	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Every effort	0.40	0.46	0.41	0.51	1.17	0.26	0.42	0.94	1.66	0.32	0.45	0.09	1.61	11.26	0.99	0.40	0.37	0.13	1.40	13.42
Pain (per %)	0.06	0.09	0.06	0.19	0.05	0.05	0.21	0.11	0.02	0.06	0.37	0.02	0.14	-1.93	0.05	0.41	0.13	0.02	0.15	0.41
Wait (per month) * Extreme	-1.17	-0.29	-1.06	-0.28	-0.17	-0.52	-0.30	-0.88	-0.27	-1.34	0.18	-0.14	-0.70	-10.19	-0.33	0.32	-1.50	-0.14	-0.76	0.66
Shared decision-making																				
No effort * Extreme	0.46	0.17	0.32	-0.40	0.59	0.91	-0.77	0.63	-0.11	0.23	-1.02	0.15	1.20	18.25	0.47	-1.08	0.73	-0.02	1.32	10.25
Some effort * Extreme	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Every effort * Extreme	0.14	0.09	-0.07	-0.40	0.16	0.24	-0.47	-0.14	0.61	0.35	-0.54	0.32	-0.72	-7.39	-0.55	-0.44	0.42	0.30	-0.52	-9.27
Pain (per %) * Extreme	0.00	0.02	0.02	-0.02	0.03	-0.06	-0.03	0.09	0.09	-0.03	-0.18	0.03	0.03	2.15	-0.01	-0.20	-0.10	0.03	0.03	-0.17
Class Share	30%	70%	37%	35%	28%	32%	25%	28%	15%	24%	25%	16%	25%	10%	17%	19%	14%	16%	24%	10%
BIC	13,540		12,609			12,411				12,389					12,379					

* models include dummy variables to control for order

Table A.3: Latent class models for 7 to 8 classes*

	7 classes							8 classes							
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8
Wait (per month)	-1.09	-0.78	-0.07	-0.03	-0.67	-0.12	10.07	-1.13	-0.76	-0.48	-0.29	-0.04	-0.21	-0.11	-4.17
Shared decision-making															
No effort	-0.41	0.61	-0.33	-3.18	-1.62	-0.87	-22.78	-0.41	-1.24	0.90	-23.09	-0.91	-3.13	-0.67	-57.36
Some effort	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Every effort	0.90	0.46	-0.06	0.64	3.33	0.31	7.05	0.99	2.02	0.72	-1.04	0.24	1.46	0.36	52.90
Pain (per %)	0.06	0.40	0.00	0.19	0.15	0.13	-1.99	0.07	0.19	0.23	2.53	0.01	0.16	0.12	1.75
Wait (per month) * Extreme															
Shared decision-making	-0.36	0.31	-0.10	-0.63	-2.43	-1.49	-10.39	-0.33	-3.93	-0.09	-0.43	-0.16	-0.81	-3.03	3.76
No effort * Extreme	0.42	-1.11	-0.09	2.40	-1.52	0.74	19.23	0.36	-4.47	-1.24	21.68	0.36	1.70	7.63	52.05
Some effort * Extreme	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Every effort * Extreme	-0.51	-0.52	0.47	0.75	-4.03	0.35	-3.75	-0.67	-4.78	-0.63	1.57	0.23	0.68	8.05	-47.69
Pain (per %) * Extreme	-0.02	-0.20	0.05	-0.05	0.29	-0.09	2.15	-0.02	0.47	-0.11	-2.08	0.02	0.07	-0.34	-1.46
Class Share	17%	21%	11%	17%	11%	14%	9%	17%	7%	14%	13%	15%	16%	10%	7%
BIC	12,431							12,542							

* models include dummy variables to control for order

Table A.4: Characteristics of respondents by latent class

	Class 1 (n=445)		Class 2 (n=358)		Class 3 (n=429)		Class 4 (n=224)	
Age group, n (%)	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
60-64	186	42%	143	40%	178	41%	94	42%
65-69	134	30%	97	27%	126	29%	64	29%
70-74	72	16%	71	20%	90	21%	41	18%
75-79	43	10%	34	9%	27	6%	18	8%
80+	10	2%	13	4%	8	2%	7	3%
Gender, n (%)								
Male	237	53%	186	52%	171	40%	98	44%
Female	208	47%	172	48%	258	60%	126	56%
Education, n (%)								
8th grade or less	5	1%	2	1%	2	0%	1	0%
Some high school, but did not graduate	24	5%	22	6%	17	4%	12	5%
High school or high school equivalency certificate	117	26%	89	25%	110	26%	52	23%
College, CEGEP or other non-university certificate or diploma	144	32%	114	32%	136	32%	72	32%
Undergraduate degree or some university	97	22%	82	23%	113	26%	62	28%
Post-graduate degree or professional designation	58	13%	49	14%	51	12%	25	11%
Preference for involvement in decision-making, n (%)								
…to make the final treatment decision.	22	5%	24	7%	15	3%	17	8%
…to make the final treatment decision after seriously considering my doctor's opinion.	181	41%	137	38%	175	41%	95	42%
…that my doctor and I share responsibility for deciding which treatment is best.	191	43%	161	45%	186	43%	91	41%
…that my doctor makes the final treatment decision, but seriously considers my opinion.	39	9%	28	8%	45	10%	19	8%
…to leave all treatment decisions to my doctor.	12	3%	8	2%	8	2%	2	1%
Has been diagnosed with arthritis, n (%)	200	45%	166	46%	225	52%	100	45%
Friend/family member has been diagnosed with arthritis, n (%)	328	74%	262	73%	342	80%	162	72%
Has had a THA or TKA, n (%)	29	7%	18	5%	18	4%	15	7%
Friend/family member has had a THA or TKA, n (%)	283	64%	251	70%	298	69%	144	64%

Table A.5: Conditional logit models for alternative samples*

	Sample 1 (base case)			Sample 2			Sample 3			Sample 4			Sample 5		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Wait (per month)	-0.16	0.01	0.00	-0.19	0.02	0.00	-0.19	0.02	0.00	-0.19	0.02	0.00	-0.20	0.02	0.00
Shared decision-making															
No effort	-0.74	0.05	0.00	-0.79	0.06	0.00	-0.80	0.07	0.00	-0.81	0.07	0.00	-0.84	0.07	0.00
Some effort	Ref			Ref			Ref			Ref			Ref		
Every effort	0.39	0.06	0.00	0.48	0.07	0.00	0.56	0.08	0.00	0.56	0.08	0.00	0.59	0.08	0.00
Pain (per %)	0.08	0.00	0.00	0.08	0.00	0.00	0.08	0.00	0.00	0.08	0.00	0.00	0.09	0.00	0.00
Wait (per month) * Extreme	-0.30	0.03	0.00	-0.36	0.03	0.00	-0.40	0.04	0.00	-0.40	0.04	0.00	-0.42	0.04	0.00
Shared decision-making															
No effort * Extreme	0.31	0.08	0.00	0.46	0.10	0.00	0.47	0.11	0.00	0.45	0.11	0.00	0.47	0.12	0.00
Some effort * Extreme	Ref			Ref			Ref			Ref			Ref		
Every effort * Extreme	-0.10	0.09	0.28	-0.07	0.11	0.49	-0.05	0.12	0.66	-0.07	0.12	0.56	-0.09	0.13	0.49
Pain (per %) * Extreme	-0.01	0.00	0.12	-0.02	0.01	0.00	-0.01	0.01	0.09	-0.01	0.01	0.07	-0.01	0.01	0.09
Log-Likelihood	-7263			-5084			-4105			-4073			-3673		
AIC	14,559			10,200			8,244			8,179			7,380		
BIC	14,690			10,326			8,367			8,302			7,502		
Observations	26,208			19,404			16,218			16,110			14,904		

* models include dummy variables to control for order

Sample 1: base case excludes only those that chose the same alternative every time; Sample 2: same as base case plus those with lexicographic preferences; Sample 3: same as sample 2 + those that failed the consistency check; Sample 4: same as sample 3 + those that spent less than 5 seconds per DCE question on average; Sample 5: same as sample 3 + those that spent less than 10 seconds per DCE question on average.

Table A.6: Mixed logit models for alternative samples*

	Sample 1 (base case)			Sample 2			Sample 3			Sample 4			Sample 5		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Mean															
Wait (per month)	-0.28	0.02	0.00	-0.24	0.02	0.00	-0.24	0.02	0.00	-0.24	0.02	0.00	-0.26	0.02	0.00
Shared decision-making															
No effort	-1.07	0.08	0.00	-0.97	0.08	0.00	-1.01	0.08	0.00	-1.03	0.09	0.00	-1.05	0.09	0.00
Some effort	Ref			Ref			Ref			Ref			Ref		
Every effort	0.68	0.08	0.00	0.61	0.08	0.00	0.71	0.09	0.00	0.74	0.09	0.00	0.75	0.10	0.00
Pain (per %)	0.12	0.01	0.00	0.11	0.01	0.00	0.10	0.01	0.00	0.11	0.01	0.00	0.12	0.01	0.00
Wait (per month) * Extreme	-0.48	0.04	0.00	-0.48	0.04	0.00	-0.60	0.05	0.00	-0.58	0.05	0.00	-0.59	0.05	0.00
Shared decision-making															
No effort * Extreme	0.42	0.12	0.00	0.50	0.12	0.00	0.49	0.14	0.00	0.49	0.14	0.00	0.45	0.15	0.00
Some effort * Extreme	Ref														
Every effort * Extreme	-0.16	0.13	0.21	-0.09	0.14	0.52	-0.03	0.16	0.84	-0.06	0.16	0.72	-0.16	0.17	0.32
Pain (per %) * Extreme	-0.02	0.01	0.00	-0.02	0.01	0.00	-0.01	0.01	0.40	-0.01	0.01	0.11	-0.01	0.01	0.10
Standard Deviation															
Wait (per month)	0.36	0.02	0.00	0.15	0.03	0.00	-0.10	0.04	0.01	-0.17	0.03	0.00	0.16	0.03	0.00
Shared decision-making															
No effort	1.18	0.06	0.00	-0.76	0.06	0.00	0.75	0.07	0.00	0.79	0.08	0.00	0.81	0.08	0.00
Some effort	Ref			Ref			Ref			Ref			Ref		
Every effort	0.36	0.09	0.00	0.24	0.11	0.03	-0.48	0.10	0.00	-0.34	0.12	0.01	0.44	0.09	0.00
Pain (per %)	0.11	0.00	0.00	0.06	0.00	0.00	0.06	0.00	0.00	0.06	0.00	0.00	0.07	0.00	0.00
Wait (per month) * Extreme	0.20	0.05	0.00	-0.21	0.04	0.00	0.34	0.04	0.00	-0.28	0.05	0.00	-0.27	0.05	0.00
Shared decision-making															
No effort * Extreme	0.15	0.09	0.08	0.04	0.10	0.67	-0.17	0.11	0.12	0.01	0.10	0.92	-0.12	0.12	0.32
Some effort * Extreme	Ref			Ref			Ref			Ref			Ref		
Every effort * Extreme	0.15	0.09	0.08	-0.02	0.10	0.82	-0.11	0.13	0.40	0.28	0.15	0.06	0.12	0.13	0.37
Pain (per %) * Extreme	0.01	0.01	0.14	0.02	0.01	0.05	0.03	0.01	0.00	-0.03	0.01	0.00	0.02	0.01	0.08
Log-Likelihood	-6392			-4880			-3923			-3866			-3504		
AIC	12,848			9,825			7,910			7,797			7,073		

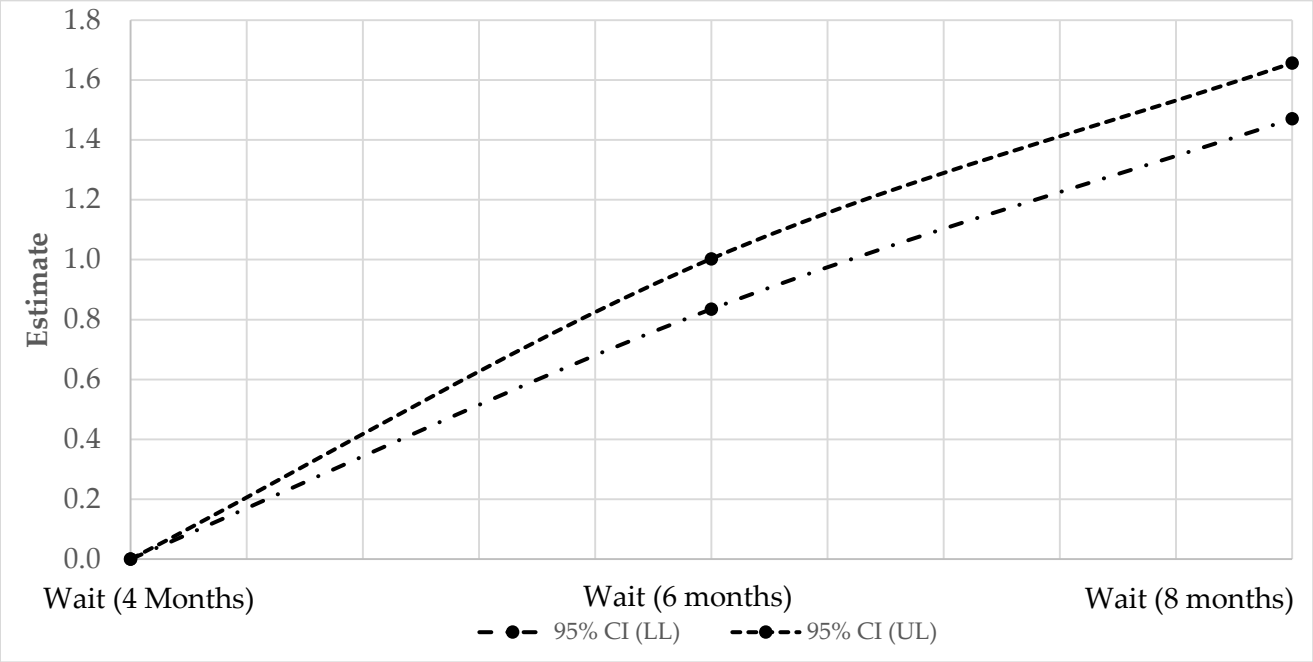
BIC	13,110	10,077	8,156	8,043	7,316
Observations	26,208	19,404	16,218	16,110	14,904

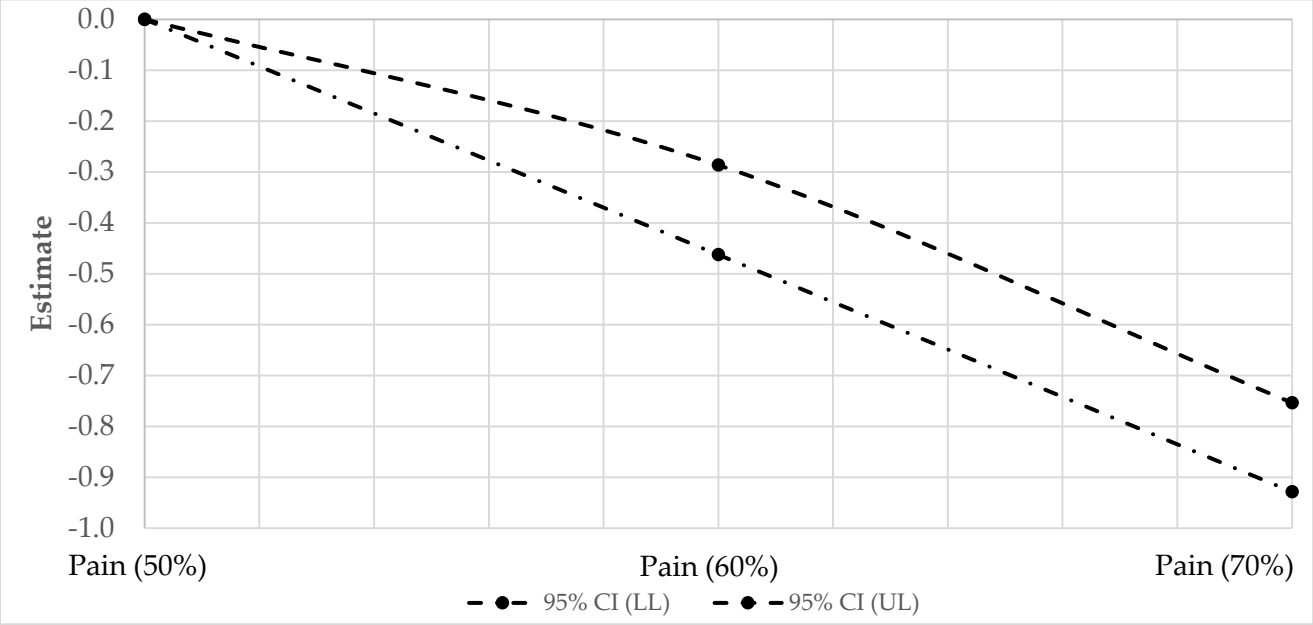
* models include dummy variables to control for order

Sample 1: base case excludes only those that chose the same alternative every time; Sample 2: same as base case plus those with lexicographic preferences; Sample 3: same as sample 2 + those that failed the consistency check; Sample 4: same as sample 3 + those that spent less than 5 seconds per DCE question on average; Sample 5: same as sample 3 + those that spent less than 10 seconds per DCE question on average.

Appendix 6.5: Testing linearity assumption for wait and pain

Testing the linearity assumption for both the *wait* and *pain* attributes was assessed by plotting upper and lower 95% confidence intervals for categorical coefficient estimates. Both plots suggest that a linear relationship is plausible.





Appendix 6.6: Estimating QALY gain

In the trial, decision quality was 44.5% in the usual care arm, compared to 56.1% in the decision aid arm, representing an 11.6% absolute increase. Multiplying this absolute increase (0.116) by the health state utility value of SDM estimated in this study for every effort (relative to 'some effort') (0.002), and assuming this benefit last one-year (1), results in an estimated QALY gain of 0.0002 per patient (0.116*0.002*1).