BMJ Open

How do surgeons' trade-off between patient outcomes and risk of complications in total knee arthroplasty? a discrete choice experiment in Australia

Sandie Szawlowski,[1] Peter F M Choong,[2,3] Jinhu Li,[4] Elizabeth Nelson,[2] Mandana Nikpour,[5] Anthony Scott,[1] Vijaya Sundararajan,[6] Michelle M Dowsey[2,3]

To cite: Szawlowski S, Choong PFM, Li J, *et al*. How do surgeons' trade-off between patient outcomes and risk of complications in total knee arthroplasty? a discrete choice experiment in Australia. *BMJ Open* 2019;**9**:e029406. doi:10.1136/bmjopen-2019-029406

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-029406).

Received 28 January 2019
Revised 23 May 2019
Accepted 11 June 2019

For numbered affiliations see end of article.

Correspondence to
Dr Michelle M Dowsey;
mmdowsey@unimelb.edu.au

ABSTRACT

Objective To measure the trade-off between risk of complications versus patient improvement in pain and function in orthopaedic surgeons' decisions about whether to undertake total knee arthroplasty (TKA).

Methods A discrete choice experiment asking surgeons to make choices between experimentally-designed scenarios describing different levels of operative risk and dimensions of pain and physical function. Variation in preferences and trade-offs according to surgeon-specific characteristics were also examined.

Results The experiment was completed by a representative sample of 333 orthopaedic surgeons (n=333): median age 52 years, 94% male, 91% fully qualified. Orthopaedic surgeons were willing to accept substantial increases in absolute risk associated with TKA surgery for greater improvements in a patient's pain and function. The maximum risk surgeons were willing to accept was 40% for reoperation and 102% for the need to seek further treatment from a general practitioner or specialist in return for a change from postoperative severe night-time pain at baseline to no night-time pain at 12 months. With a few exceptions, surgeon-specific characteristics were not associated with how much risk a surgeon is willing to accept in a patient undergoing TKA.

Conclusion This is the first study to quantify risk-benefit trade-offs among orthopaedic surgeons performing TKA, using a discrete choice experiment. This study provides insight into the risk tolerance of surgeons.

Strengths and limitations of this study

- ► To the best of our knowledge, this study is the first to investigate the trade-offs between improvements in pain and function and risk of total knee arthroplasty surgery using a discrete choice experiment (DCE) in orthopaedic surgeons.
- ► The choice task allows researchers to quantify how surgeons weigh up their trade-offs between defined benefits and risks of surgery.
- ► This novel method reveals unique insights into the decision-making process of surgeons.
- ► The DCE may lack external validity if surgeons do not make the same choices in real life.
- ► The analysis of the DCE did not include a comparison to a 'status quo' patient.

INTRODUCTION

The decision to undertake surgery is based on a consideration of the risks of complications as well as potential benefits to patients in terms of reduction in pain and improvement in physical function. Despite the daily demand for surgeons to make risk-benefit trade-offs there is limited research on the risk tolerance of surgeons and its influence on decisions to perform surgery. It is possible that surgeons focus on the risks of complications rather than benefit, as complications are more readily observed and documented, whereas improvements in postoperative pain and function are more subjective and are less easily observed and quantified. Alternatively, surgeons may overestimate the benefits and underestimate the risks of surgery.[1]

The purpose of this study was twofold. First, to understand how orthopaedic surgeons balance the postoperative improvements in patient outcomes (pain and/or function) and risk (surgical complications) when considering patients for total knee arthroplasty (TKA). Second, we sought to identify whether surgeon characteristics are associated with preferences in terms of risk-benefit trade-offs.

Osteoarthritis (OA), one of the most disabling diseases in developed countries, affects over three million people worldwide.[2] TKA is the mainstay of treatment for end-stage knee OA. TKA can improve quality of life and reduce pain, joint deformity and

loss of function. In 2016, nearly 53 000 primary TKA surgeries were performed across Australia, an increase of 139.8% since 2003.[3] This rapid increased is witnessed throughout Organisation for Economic Co-operation and Development (OECD) countries, where on average the rate of knee replacements nearly doubled between 2000 and 2015.[4] The increased prevalence of OA and hence demand for TKA surgery is largely due to an ageing population.

METHODS

Study design

A discrete choice experiment (DCE) was administered to orthopaedic surgeons via a mailed and online survey, including orthopaedic fellows-in-training, to elicit the maximum acceptable risk they are willing to take in TKA. The survey took 30 min to complete and was divided into five sections in the following order: demographic information, surgical risk ranking, preferences and outcomes, work setting and surgeon-specific characteristics. Respondents compared a series of hypothetical but realistic scenarios describing 12 month post-TKA outcomes and risks of complications. Figure 1 gives an example of a choice pair administered to participants.

Selection and development of attributes and levels for DCE

The attributes of the DCE were designed to reflect the most salient aspects of the risks of complications and patient outcomes for TKA (table 1) using accepted methods.[5]

Six attributes, determined by an extensive literature review, face-to-face interviews with patients and orthopaedic surgeons and feedback from a panel of orthopaedics, rheumatology, primary care and health economics experts, were included in the DCE. Each attribute covered pain, physical function and risks associated with TKA surgery had three different levels.

Pain and function attributes were derived from the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC),[6] a widely-used and validated questionnaire designed specifically to evaluate patient responses to knee OA treatment. The assigned levels were determined by the 12 month post elective primary TKA surgery WOMAC scores held by the St. Vincent's Melbourne Arthroplasty (SMART) registry for patients who underwent surgery at St. Vincent's Hospital Melbourne (SVHM), a large metropolitan hospital in Australia. The SMART registry captures information from surgeons performing joint arthroplasty and participants are demographically representative of the Australian patient population.[7] Registry data collection started in 1998 and >11 000 procedures are now registered with 800 new yearly registrations. The registry has complete capture of all preoperative and postoperative encounters and achieves 98% follow-up of patient-reported outcome measures at 1 year.

The absolute risk attributes were developed by identifying the most common complications within 12 months post-TKA surgery using 2006 to 2012 SMART registry data (n=2552). The numerous types of complications were aggregated into two categories for the DCE and worded so they could be easily understood by patients for the purposes of future use in a patient cohort and patient/surgeon comparisons[8]: 'Risk of having to go back into hospital and having a second operation on your knee' and 'Risk of getting a complication that requires seeing your general practitioner (GP) or specialist for further treatment'. Patients may have to undergo reoperation on their knee if they have stiffness in the knee or for treatment of surgical site infection. If the patient suffers from a blood clot, ongoing pain or a superficial wound complication they would have to have to see their GP or specialist. The attribute levels varied by the minimum (0% for both risk attributes), median (7% for risk of reoperation and 10% for risk of a complication that requires a new specialist or GP visit) and maximum (13% for risk of reoperation and 21% for risk of a complication that requires a new specialist or GP visit) rate of the identified risks according to the registry data. Following best practice in DCE design, the risk information was presented using icon arrays as visual aid to numerical presentation (figure 1).[9 10]

Experimental design

The six attributes and their corresponding levels (shown in table 1) have a possible $3^6 = 729$ different combinations of outcome scenarios (six attributes with three levels each). All 729 scenarios were not presented to each respondent due to likely respondent fatigue and low response rates.[11] Using Ngene 1.2[12] software, a fractional factorial experimental design was used to reduce the number of scenarios while maximising the variation in the data.[13] An efficient design was used, allowing for attributes to be independently varied over scenarios while minimising predicted SEs of the parameter estimates. Specifically, we used a D-efficient design in which the D-error is minimised.[14] The final optimal design included 12 choice pairs. To reduce the cognitive burden and fatigue for the respondents, these 12 choice pairs were 'blocked' and allocated across two versions of the DCE questionnaire, each with six choice pairs. Participants were randomly allocated to one of the two versions of the questionnaire. Each choice pair consisted of two alternative scenarios (see figure 1), which were labelled 'Choice A' and 'Choice B'. Respondents chose their preferred outcome, either 'Choice A' or 'Choice B', for each of the six choice pairs presented to them. Following each choice pair, an opt-out was offered to account for the voluntary nature of elective TKA. The respondent was asked, given their choice, whether they would prefer to perform the operation or rather their patient remained in their current health state.

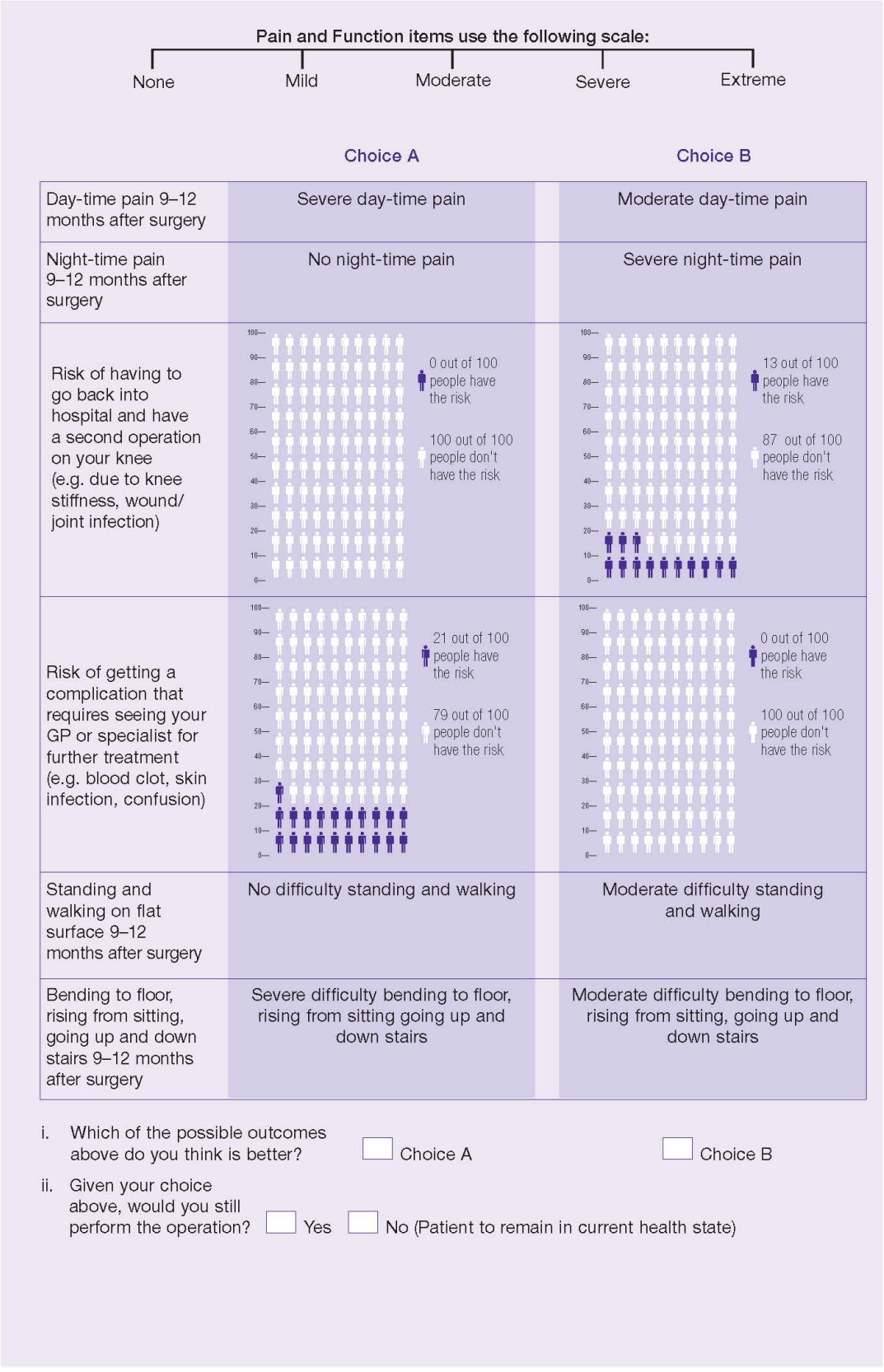

Figure 1 Example of a discrete choice experiment. GP, general practitioner; TKA, total knee arthroplasty.

Table 1 Attributes and levels included in the discrete choice experiment

Attributes	Levels	Variable coding for analysis
Pain outcomes:		
1. Daytime pain 9–12 months after surgery	None Moderate Severe	Dummy variable – 'severe daytime pain' was the omitted reference group
2. Night-time pain 9–12 months after surgery	None Moderate Severe	Dummy variable – 'severe night-time pain' was the omitted reference group
Functional outcomes:		
3. Standing and walking on a flat surface 9–12 months after surgery	No difficulty Moderate difficulty Severe difficulty	Dummy variable – 'severe difficulty standing and walking' was the omitted reference group
4. Bending to the floor, rising from sitting and going up and down stairs 9–12 months after surgery	No difficulty Moderate difficulty Severe difficulty	Dummy variable – 'severe difficulty bending from the floor, rising from sitting and going up and down stairs' was the omitted reference group
Risk of complications:		
5. Risk of having to go back into hospital and have a second operation on your knee (eg, due to knee stiffness, wound/joint infection)	0%, 7%, 13%	Continuous
6. Risk of getting a complication that requires seeing your GP or specialist for further treatment (eg, blood clot, skin infection, confusion)	0%, 10%, 21%	Continuous

GP, general practitioner.

Experimental design testing

The survey instrument underwent rigorous pretesting at the design stage to verify the appropriateness of the precise wording and framing of the attributes and their corresponding levels followed by two phases of piloting. Phase 1 involved systematic face-to-face interviews with five orthopaedic surgeons. For phase 2, 21 orthopaedic surgeons completed the full pilot version of the survey. Patients undergoing TKA at SVHM were also involved in both phases of piloting. Prior information on the regression coefficients from the analysis of the pilot were used to help generate the final experimental design. The DCE was designed with the intention of being completed by both patients and surgeons.

Data collection

All orthopaedic surgeons across Australia were invited to participate. Participants were identified using a database provided by the Australian Medical Publishing Company (AMPCo) which holds contact details for all doctors in Australia. In October 2016, 1257 orthopaedic surgeons, including fellows-in-training, were invited to participate in the study using a mixed mode of approach and completion.[15] They were contacted via mail-out and, for those with a known email address, also by email. A postal invitation included a personalised letter explaining the study, a prepaid return envelope, instructions on how to complete the survey online and a hardcopy of a randomly allocated survey. Participants chose whether to fill out the hardcopy or online version. The email invite included information about the study and a link to access their online survey. The completion of the questionnaire implied their voluntary consent to participate in the research. For surgeons who responded twice, submitting both online and hardcopy versions of the survey, the most complete entry was chosen in the analysis. If both responses were completed equally the online version was chosen to minimise the risk of administrative error in entering the data. All responses were anonymous, and all information held in the strictest of confidence.

Study size

A target sample size of 400 surgeons and registrars was defined to support effective subgroup analysis for the DCE. Our Monte Carlo simulation indicated that the minimum required sample was 200 surgeons with 12 choice pairs. However, since the 12 choice pairs were blocked into two versions of DCE, the target sample size increased to 400 surgeons.[16]

Statistical methods

The analysis of the DCE was conducted by estimating a mixed logit model using Stata 15.0. A well-defined mixed logit model can approximate any discrete choice random utility model[17] and therefore is preferred throughout the DCE literature[18] and widely applied in health economics.[11 19] Unlike other logit models, the mixed logit model can account for unobservable preference

heterogeneity by including random coefficients. These random coefficients capture how preferences for each attribute will vary over individuals, allowing for the estimation of individual-specific coefficients that follow a prespecified distribution. Hence the mixed logit model is associated with having better 'goodness of fit' than other logit models.

The DCE data contain 12 observations from six choice pairs per survey respondent. Each observation is one of the two alternatives from each of the six choice pairs presented, and with the dependent variable equal to one or zero for each choice pair. Observations from respondents with missing values of the dependent variable were excluded from the analysis. In the estimation of the model, categorical variables (ie, the attributes and associated levels) were coded as dummy variables with 'severe' as the omitted reference category. The risk attributes were considered as continuous variables in the final model. This is necessary to calculate the risk-benefit trade-offs (marginal rates of substitution). The assumption of linearity of the risk attributes was tested in a sensitivity analysis that estimated two models which relaxed the linearity assumption for each risk attribute one at a time. These models re-coded risk as a categorical variable using the levels of the attribute and comparing goodness of fit with the main model using akaike information criterion (AIC) and Bayesian information criterion (BIC). To examine the association between each attribute and surgeon characteristics, interaction terms were included in the mixed logit model. The inclusion of random coefficients in the model gives each individual their own regression coefficient.[20] The results show the mean and SD of these coefficients. A statistically significant SD shows that there is variation across individual surgeons in their preferences for the given attribute, that is, they do not 'agree' as to its relative importance.

To extract the relative importance of the attributes and their levels, the marginal rate of substitution (trade-offs) is calculated between one of the risk attributes and each quality of life attribute, by dividing the estimated coefficient of quality of life attribute (pain or function) by the estimated coefficient of risk attribute. This addresses the question of how much additional risk is equivalent to a health improvement, for example, from severe daytime pain to no daytime pain.

Surgeon-specific characteristics

Interaction terms between each attribute and the characteristics listed below allowed for the examination of surgeon-specific factors influencing preferences and trade-offs. From the literature, four characteristics were analysed. Procedure volume was analysed as a dichotomous variable where a high-volume surgeon was defined as a surgeon who performs above or equal to the median number of TKA surgeries per week in the sample (≥3.25), only surgeons who performed >0 TKA surgeries in their 'last usual working week' were included in the analysis. Experience, encompassing both age and seniority, was measured as a continuous variable by the number of years since the respondent became a Fellow of the Royal Australian College of Surgeons. Given this definition, fellows-in-training therefore had the least experience. Surgeon personality was measured using a Likert-scale approach by the Big Five Personality Index (BFI)[21], Mastery Locus of Control (LOC)[22] and Life Orientation Test-Revised (LOTR).[23] The BFI tests for a set of five broad trait dimensions (neuroticism, extraversion, openness to experience, agreeableness and conscientiousness), see online supplementary table 1 for an overview, using a 15-item questionnaire across a 5-point scale, where 1=disagree strongly to 5=agree strongly. The LOC, a 7-item questionnaire using an 11-point scale ranging from 1=strongly agree to 11=strongly disagree, evaluates the control an individual has over their everyday life and the LOTR, a 10-item questionnaire, measures optimism using a 5-point scale where 1=I agree a lot and 5=I disagree a lot. Finally, to investigate whether risk attitudes vary between surgeons who perform more TKA procedures in a public compared with private hospital, the proportion of public to private TKAs performed in a surgeon's average week was included as an interaction term with each attribute. The majority of TKA surgery is performed in the private sector where doctors are remunerated on a fee for service basis.[24] Fee for service may provide a financial incentive to surgeons and hence, could increase surgeons' propensity to overestimate the benefits and underestimate the risks.

Patient and public involvement

This study is part of a larger study which will additionally investigate the maximum acceptance of risk of patients in TKA. The DCE for both surgeons and patients were defined by the same attributes and levels. Patients were involved in the pretesting of the survey instrument. Participants had end-stage OA and were recruited at the orthopaedic preoperative assessment clinic after being consented and waitlisted for primary TKA at SVHM.

The initial pretesting phase with patients consisted of detailed face-to-face interviews with 15 patients. For the second phase, 40 patients completed the pilot survey. Patient feedback was sought for the ease of comprehension of wording and framing of the attributes and their corresponding levels, efficacy figures, icon arrays and the length of questionnaire. The main issues raised were around the language used, the wording of the attributes was consequently changed to improve understanding.

RESULTS

Among the 1257 surgeons contacted, 434 responded (34.5%). Seventy-two (16.6%) responses were refusals to complete the survey. Reasons for refusal included 'do not perform TKA' and 'being retired'. A total of 362 completed and 18 'return to sender' surveys were returned, a participation rate of approximately 29%. See online supplementary figure 1 for consort diagram. Of the 362 who returned the survey, 333 selected at least

Table 2 Respondent demographic and personality/practice characteristics

	Estimation sample (n=333)	Sample frame (n=1257)
Characteristics		
Male, no. (%)	314 (94.3)	1199 (95.4)
Age, year (IQR)	52 (44–59)	50 (42–60)
Practice status, no. (%)		
Accredited registrar	16 (4.8)	120 (9.6)
Consultant	304 (91.3)	1124 (89.4)
Other	12 (3.6)	13 (1.0)
Australian states and territories, no. (%)		
Victoria	93 (27.9)	275 (21.9)
New South Wales	92 (27.6)	408 (32.5)
South Australia	23 (6.9)	113 (9.0)
Queensland	58 (17.4)	271 (21.6)
Northern Territory	3 (0.9)	6 (0.5)
Western Australia	29 (8.7)	136 (10.9)
Tasmania	12 (3.6)	24 (1.9)
Australian Capital Territory	6 (1.8)	23 (1.8)
Residency status, no. (%)		
Australian citizen	308 (92.5)	-
Permanent resident	19 (5.7)	-
Temporary resident	3 (0.9)	-

	Mean	SD	Min	Max
Personality traits:				
BFI – extraversion	3.2	0.82	1	5
BFI – agreeableness	4.09	0.71	2	5
BFI – conscientiousness	4.45	0.54	2	5
BFI – neuroticism	2.62	0.87	1	4.67
BFI – openness	3.63	0.73	1.33	5
LOC	8.28	1.84	1.86	11
LOTR	23.84	4	11	30
Surgeon experience:				
Years of experience	19.76	10.49	1	55
Surgeon volume:				
TKA per week	3.65	4.56	0	60
Proportion of high volume surgeons	0.43	0.5	0	1
Public vs private TKA surgery:				
Proportion of public to private	0.4	0.34	0	1

Personality traits were standardised for the regression analysis, hence mean=0 and SD=1. Zero observations were excluded in the regression analysis for the high-volume and proportion of public to private procedures performed interaction effects. The median number of TKA surgeries per week was used to determine high and low volume surgeons.
BFI, Big Five Personality Index; LOC, Mastery Locus of Control; LOTR, Life Orientation Test-Revised; TKA, total knee arthroplasty.

one alternative from the each of the six choice pairs in the DCE. These 333 respondents provided 3862 observations for the analysis, out of a possible 3996 (333×12) observations. A comparison of the population of orthopaedic surgeons from the AMPCo sample frame with respondents is summarised in table 2. The median age of respondents was 52 years (IQR 44 to 59 years). Most respondents were male (94%) and fully-qualified orthopaedic surgeons (91%). The survey sample was representative of the population except for fellows-in-training who were underrepresented and surgeons performing TKA in Victoria and Tasmania were overrepresented.

Table 3 Trade-offs between risk and patient outcomes: marginal rate of substitution

	Risk of reoperation		Risk of complication requiring a new GP/specialist visit	
	Coeff	SE	Coeff	SE
Pain outcomes:				
No daytime pain	−34.06***	4.01	−87.02***	17.96
Moderate daytime pain	−25.27***	3.18	−64.54***	14.19
No night-time pain	−39.98***	4.72	−102.13***	22.08
Moderate night-time pain	−25.73***	2.86	−65.73***	13.44
Functional outcomes:				
No difficulty standing	−27.65***	5.00	−70.63***	18.62
Moderate difficulty standing	−17.33***	3.07	−44.28***	10.79
No difficulty moving	−20.62***	2.43	−52.67***	12.18
Moderate difficulty moving	−9.72***	1.56	−24.84***	6.09
Risk of complications:				
Risk of new GP/specialist visits	0.39***	0.06	–	–
Risk of reoperation	–	–	2.55***	0.41

The marginal rate of substitution is calculated between one of the risk attributes and each quality of life attribute, by dividing the estimated coefficient of quality of life attribute (pain or function) by the estimated coefficient of risk attribute. Categorical variables for pain and function were coded as dummy variables with 'severe' as the omitted reference category. The risk attributes were continuous variables.
***P≤0.01, **P≤0.05, *P≤0.1
Coeff, coefficient; GP, general practitioner.

Respondents had an average of almost 20 years of experience and performed an average of four TKAs per week. For every 10 TKAs performed in a private hospital, four were conducted in a public hospital (table 2).

The estimated mixed logit model results are presented in online supplementary table 2. It is not possible to draw direct inferences from the coefficients however, the signs are as expected and significant at the 1% level: surgeons prefer patients to suffer from less pain, have better function and for there to be less risk of adverse events occurring. Shown by the SD, there is statistically significant variation in surgeons' preferences for most attributes. The insignificant constant term illustrates no surgeon preference for 'Choice A' or 'Choice B' and tests for specification error.

The marginal rate of substitution between risk and patient outcomes are shown in table 3. Linearity of risk was confirmed (according to AIC and BIC: results available on request) by comparing models with risk re-coded as a categorical variable. The relative size of these trade-offs indicates the relative importance of each health improvement to surgeons. Surgeons believe that the alleviation of night-time pain is the most important attribute, compared with all other attributes they are willing to accept the maximum risk to achieve this. To improve a patient's night-time pain from severe-to-no pain, surgeons are willing to accept a 40% or 102% increase in the absolute risk of reoperation or the risk of a complication which requires a specialist or GP visit, respectively. Reducing pain is generally more important to surgeons than improvements in functioning. The relative importance is similar when trading off the risk of a complication that requires a new specialist or GP visit. For each attribute, surgeons are willing to accept higher risks of complications requiring GP/specialist visits, compared with risk of reoperation which they consider to be more serious. For example, surgeons are prepared to accept an 87% increase in the risk of a complication requiring a specialist or GP visit to reduce daytime pain from severe at baseline (pre-surgery) to no pain at 12 months. For the same improvement for patients they are only willing to accept a 34% increase in the risk of reoperation.

Furthermore, a 1% increase in the risk of reoperation is shown to be equal to a 2.55% increase in the risk of new GP visits within the first year after TKA. The risk of reoperation is 2.55 times more important to surgeons than the risk of a complication requiring only a specialist or GP visit. Hence surgeons are less willing to risk patients being readmitted to undergo another surgery than seeing their GP or specialist.

Table 4 summarises the direction and statistical significance of the interactions between surgeon preferences for each attribute, and the volume of TKA, personality traits, experience and public-private mix. Overall, there were only a few surgeon-specific characteristics, namely personality traits, shown to affect surgeon preferences.

A more 'open' surgeon is likely to find the ability to stand more important but the ability to move less important and an 'agreeable' surgeon finds the ability to move more important, significant at the 5% level. However, being more conscientious, neurotic or the level of control a surgeon feels they have in their everyday life

Table 4 Summary of mixed logit interaction effects

	Pain outcomes				Function outcomes				Risk of complications	
	No daytime pain	Moderate daytime pain	No night-time pain	Moderate night-time pain	No difficulty standing	Moderate difficulty standing	No difficulty moving	Moderate difficulty moving	Reoperation	New GP visits
Surgeon personality traits										
BFI - openness	[-] ns	[-] ns	[-] ns	[+] ns	[+] **	[+] **	[-] ns	[-] ***	[+] ns	[+] ns
BFI - conscientiousness	[+] ns	[+] ns	[-] ns	[-] ns	[-] ns	[-] ns	[+] ns	[+] ns	[-] ns	[+] ns
BFI - extraversion	[-] ns	[-] ns	[-] ns	[+] ns	[-] ns	[-] ns	[+] ns	[-] *	[+] ns	[+] ns
BFI - agreeableness	[+] ns	[+] ns	[+] ns	[+] ns	[-] ns	[-] ns	[+] ns	[+]**	[-] ns	[-] ns
BFI - neuroticism	[-] ns	[+] ns	[-] ns	[-] ns	[+] ns	[+] ns	[-] ns	[-] ns	[+] ns	[+] ns
LOC	[-] ns	[-] ns	[-] ns	[-] ns	[-] ns	[-] ns	[-] ns	[+] ns	[-] ns	[+] ns
LOTR	[+] ns	[+] ns	[+] ns	[+] ns	[+] ns	[+] ns	[+] ns	[+] *	[-] *	[-] ns
Surgeon experience										
Years of experience	[-] ns	[-] ns	[-] ns	[-] ns	[-] ns	[-] ns	[-] ns	[-] *	[+] ns	[+] ns
Surgeon volume										
High volume surgeons	[+] ns	[+] ns	[+] ns	[-] ns	[+] ns	[+] ns	[-] ns	[-] ns	[+] ns	[+] ns
Public vs private TKA surgery										
Proportion of public to private	[-] ns	[+] ns	[-] ns	[-] ns	[+] ns	[+] ns	[-] ns	[-] ns	[+] ns	[+] ns

***P≤0.01, **P≤0.05, *P≤0.1, sign of coefficient in square brackets.
The sample size of each model varies from between 2892 and 3680 observations, from between 245 and 310 respondents.
BFI, Big Five Personality Index; GP, general practitioner; LOC, Mastery Locus of Control; LOTR, Life Orientation Test-Revised, ns, not significant; TKA, total knee arthroplasty.

has no effect on any of the outcomes. Neither does a surgeon's public-private mix or procedure volume. Weak negative associations between a patient improvement from severe difficulty to moderate difficulty moving with a surgeon's experience and level of extraversion are illustrated in table 4. The LOTR variable, measuring surgeon optimism, also illustrates a relationship at 10% level. A more optimistic surgeon places greater weight on the importance of a patient's improvement in function from severe-to-moderate difficulty moving and places a lower weight on the importance of risk of reoperation than a less optimistic surgeon.

DISCUSSION

This study is the first of its kind to investigate the trade-offs between improvements in pain and function and risk of TKA surgery using a DCE in orthopaedic surgeons. The choice task allows the elicitation of risk tolerance to be quantified by weighing up the different outcome alternatives (pain, function and risk).

Surgeons are willing to accept a large increase in the absolute risk of complication requiring a return to hospital for a follow-up knee operation, up to a maximum of 40%, to eliminate night-time pain (improvement from severe-to-none 12 months after the procedure). This figure is 102% for a complication that requires a GP or specialist visit for further treatment. With regards to improvements in a patients' function, a surgeon is willing to accept a 10% and 21% increase in the risk of reoperation for an improvement from severe difficulty walking to moderate and no difficulty, respectively. These trade-offs show that across all attributes, surgeons are willing to accept higher absolute risks of GP/specialist visits in comparison to reoperation. This is unsurprising as complications requiring reoperation are likely to be much more severe than those that can be treated in an ambulatory visit.

Surgeons were willing to accept the same amount of risk for improvements in each attribute regardless of personality type, experience, procedure volume or whether a surgeon performed TKA surgery in a public or private setting. Suggesting that their preferences for risk and patient outcomes, and how they trade them off, do not vary along these dimensions, though preferences do vary due to other unobserved factors. With regards to surgeon personality, the literature is conflicted. Despite evidence that surgeon personality influences risk tolerance[25] and decision making,[26] the 'surgical personality'[27–30] suggests that all surgeons have inherent personality traits that are different to non-surgeons. Hence there may be less variation within surgeons, especially within specialities such as orthopaedic surgeons. The 'surgical personality' is a consequence of surgeons' self-selection into the profession and their continual rigorous standardised training throughout their career. Though table 2 suggests some variation in personality, this may not have been sufficient variation to influence their preferences.

The finding that neither experience nor volume of TKAs influenced their preferences, suggests that surgeons are homogenous with respect to the importance they place on risk and patient outcomes. Though the risk of adverse events is associated with volume[31–34] and experience[35] through a broader and more refined skillset of high-volume surgeons compared with low-volume surgeons,[31 36] surgeons may be unaware of this relationship such that the importance of risk does not vary. We were not able to collect data on the extent to which respondents had patients who had experienced adverse events.

Our hypothesis that surgeons in the private sector may overestimate the benefits and underestimate the risks was not supported. It is uncommon for surgeons to exclusively operate in either a public or private hospital in Australia and unlikely that individual surgeons have specific 'public' and 'private' surgeon behaviours which are different. Additionally, evidence suggests that the quality of care among TKA patients is not compromised regardless of whether the surgery is performed by a public or private healthcare provider.[37]

There are several limitations to this study. First, the DCE may lack external validity if surgeons do not make the same choices in real life. Despite the outcome choices presented in the DCE being realistic and based on real data, the choice task was hypothetical. However, a recent systematic review and meta-analysis showed that choice experiments provide a reasonable approximation to actual choices.[38] DCEs are especially useful in situations where data on actual choices are difficult to collect.

Another limitation may be that data were also collected on whether the surgeons, conditional on their choice of A or B, would rather not perform the operation (figure 1). These data were not analysed in this paper which was focused more on the trade-offs between risk and patient outcomes. This option was not included as a potential third 'status quo' alternative in the analysis since no specific attribute levels could be assigned to this. In addition, the question was framed as an additional question (conditional on choice of A or B), rather than being included as a third mutually exclusive alternative.

The response rate of 34.4% may be considered as an additional limitation. However, physician response rates are notably lower than the general population.[39] Our survey compares favourably with the Medicine in Australia: Balancing Employment and Life survey which has had response rates varying from 20.6% to 33.9%, between 2010 to 2017, for specialists who have not previously completed the survey.[40] The sample analysed in this paper is representative of the population in terms of age and gender, except for fellows in training who were underrepresented and surgeons performing TKA in Victoria and Tasmania who were slightly overrepresented, see table 2. Moreover, a high response rate is not the only indicator of survey quality, since response bias may still be a cause for concern in surveys with high response rates if certain sectors of the population fail to respond.

Finally, despite the expectation of risk to be non-linear, the estimated mixed logit model included the risk attributes as continuous variables. The sensitivity analysis conducted supported the linearity assumption of risk. However, the evidence of linearity may be a consequence of the DCE design. During the design phase risk was included as continuous variable to reduce the number of questions a surgeon would have to answer, and the sample size required. Increasing the number of questions would have decreased the response rate by increasing the time burden on surgeons. There is, therefore, potential that there is insufficient variation in the data to show non-linearity and properly test this assumption.

This study is part of a larger project exploring risk-preferences of surgeons and patients. Moving forward, research into risk-benefit trade-offs of patients considering TKA as a treatment option for end-stage OA will be undertaken. This research has implications for both clinicians and policymakers. Anecdotal evidence suggests that surgeon and patient expectations of surgery are often misaligned; our findings will help improve the shared decision-making process, vital to providing high quality patient-centred healthcare. In turn, this will allow for improvements in surgical outcomes and greater patient satisfaction.

Author affiliations
[1]Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, Parkville, Victoria, Australia
[2]Department of Surgery, St Vincent's Hospital, The University of Melbourne, Fitzroy, Victoria, Australia
[3]Department of Orthopaedics, St Vincent's Hospital, Melbourne, Victoria, Australia
[4]Department of Economics, Deakin University, Burwood, Victoria, Australia
[5]Department of Medicine, St Vincent's Hospital, The University of Melbourne, Fitzroy, Victoria, Australia
[6]Department of Public Health, La Trobe University, Melbourne, Victoria, Australia

Acknowledgements We acknowledge all the participants in the survey. We also acknowledge the surgeons, patients and other medical professionals who took part in the pre-testing phases and gave their time to the project. Associate Professor Michelle Dowsey holds an NHMRC Career Development Fellowship (APP1122526). Associate Professor Mandana Nikpour holds an NHMRC Career Development Fellowship (APP1126370). Professor Peter Choong holds an NHMRC Practitioner Fellowship (APP1154203). Dr Jinhu Li holds an ARC Discovery Early Career Researcher Awards (Project ID: DE170100829).

Contributors SS was an investigator, conducted the literature search and statistical analysis, contributed to data interpretation and drafted and revised the paper. PC was a chief investigator, was involved in the design of the study, provided management oversight of the whole trial, contributed to data interpretation and drafted and revised the paper. JL was an investigator, wrote the statistical analysis plan, conducted statistical analysis, contributed to data interpretation and drafted and revised the paper. EN was the study coordinator, responsible for participant recruitment, provided technical support to participants, monitored data collection for the whole trial, drafted and revised the paper. MN and VS were chief investigators, designed the study, contributed to data interpretation and drafted and revised the paper. AS was a chief investigator, designed the study, provided management oversight over the statistical analysis, contributed to data interpretation and drafted and revised the paper. MD was the lead investigator, initiated the collaborative project, designed the study, monitored data collection for the whole trial, provided management oversight of the whole study, contributed to data interpretation, drafted and revised the paper and is the guarantor. All authors contributed to redrafts of the report. All authors had full access to the study data and take responsibility for the integrity of the data and the accuracy of the data.

Funding Financial support for this study was provided entirely by a grant from the National Health and Medical Research Grant Project, Grant no. APP1058438 (www.nhmrc.gov.au, nhmrc@nhmrc.gov.au, phone: +61 2 6217 9000). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing and publishing the report.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the St. Vincent's Human Research Ethics Committee (HREC-A 177/15).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No data are available.

REFERENCES

1. Hoffmann TC, Del Mar C. Clinicians' Expectations of the Benefits and Harms of Treatments, Screening, and Tests: A Systematic Review. *JAMA Intern Med* 2017;177:407–19.
2. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017;390:1211–59.
3. Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). *Hip, Knee & Shoulder Arthroplasty: 2017 Annual Report*. Adelaide: AOA, 2017.
4. OECD. *Health at a Glance 2017: OECD Indicators*. Paris: OECD Publishing, 2017.
5. Coast J, Al-Janabi H, Sutton EJ, *et al*. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ* 2012;21:730–41.
6. Bellamy N, Buchanan WW, Goldsmith CH, *et al*. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
7. Dowsey MM, Choong PFM, Paxton EW, *et al*. Body Mass Index Is Associated With All-cause Mortality After THA and TKA. *Clin Orthop Relat Res* 2018;476:1139–48.
8. Dowsey MM, Scott A, Nelson EA, *et al*. Using discrete choice experiments as a decision aid in total knee arthroplasty: study protocol for a randomised controlled trial. *Trials* 2016;17:416.
9. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Making* 2007;27:696–713.
10. Harrison M, Rigby D, Vass C, *et al*. Risk as an attribute in discrete choice experiments: a systematic review of the literature. *Patient* 2014;7:151–70.
11. Clark MD, Determann D, Petrou S, *et al*. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics* 2014;32:883–902.
12. ChoiceMetrics. *Ngene 1.2 User Manual & Reference Guide*, 2018.
13. Louviere JJ, Flynn TN, Carson RT. Discrete Choice Experiments Are Not Conjoint Analysis. *Journal of Choice Modelling* 2010;3:57–72.
14. Rose JM, Bliemer MCJ. Sample size requirements for stated choice experiments. *Transportation* 2013;40:1021–41.
15. Scott A, Jeon SH, Joyce CM, *et al*. A randomised trial and economic evaluation of the effect of response mode on response rate, response bias, and item non-response in a survey of doctors. *BMC Med Res Methodol* 2011;11:126.
16. de Bekker-Grob EW, Donkers B, Jonker MF, *et al*. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. *Patient* 2015;8:373–84.
17. McFadden D, Train K. Mixed MNL models for discrete response. *J Appl Econ* 2000;15:447–70.
18. Hole AR, Kolstad JR. Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. *Empir Econ* 2012;42:445–69.

19. Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law* 2009;4:527–46.
20. Hall J, Fiebig DG, King MT, *et al*. What influences participation in genetic carrier testing? Results from a discrete choice experiment. *J Health Econ* 2006;25:520–37.
21. John OP, Donahue EM, Kentle RL. *The big five inventory—versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research, 1991.
22. Pearlin LI, Schooler C. The structure of coping. *J Health Soc Behav* 1978;19:2–21.
23. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J Pers Soc Psychol* 1994;67:1063–78.
24. Australian Orthopaedic Association National Joint Replacement Registry. *Annual Report*. Adelaide: AOA, 2015.
25. Contessa J, Suarez L, Kyriakides T, *et al*. The influence of surgeon personality factors on risk tolerance: a pilot study. *J Surg Educ* 2013;70:806–12.
26. Pauley K, Flin R, Yule S, *et al*. Surgeons' intraoperative decision making and risk management. *Am J Surg* 2011;202:375–81.
27. McGreevy J, Wiebe D. A preliminary measurement of the surgical personality. *Am J Surg* 2002;184:121–5.
28. Schwartz RW, Barclay JR, Harrell PL, *et al*. Defining the surgical personality: a preliminary study. *Surgery* 1994;115:62–8.
29. Foster KN, Neidert GP, Brubaker-Rimmer R, *et al*. A psychological profile of surgeons and surgical residents. *J Surg Educ* 2010;67:359–70.
30. Drosdeck JM, Osayi SN, Peterson LA, *et al*. Surgeon and nonsurgeon personalities at different career points. *J Surg Res* 2015;196:60–6.
31 Morche J, Mathes T, Pieper D. Relationship between surgeon volume and outcomes: a systematic review of systematic reviews. *Syst Rev* 2016;5:204.
32. Ravi B, Jenkinson R, Austin PC, *et al*. Relation between surgeon volume and risk of complications after total hip arthroplasty: propensity score matched cohort study. *BMJ* 2014;348:g3284.
33. Katz JN, Barrett J, Mahomed NN, *et al*. Association between hospital and surgeon procedure volume and the outcomes of total knee replacement. *J Bone Joint Surg Am* 2004;86:1909–16.
34. Katz JN, Mahomed NN, Baron JA, *et al*. Association of hospital and surgeon procedure volume with patient-centered outcomes of total knee replacement in a population-based cohort of patients age 65 years and older. *Arthritis Rheum* 2007;56:568–74.
35. Authen AL, Dybvik E, Furnes O, *et al*. Surgeon's experience level and risk of reoperation after hip fracture surgery: an observational study on 30,945 patients in the Norwegian Hip Fracture Register 2011–2015. *Acta Orthopaedica* 2018:1–7.
36. Lau RL, Perruccio AV, Gandhi R, *et al*. The role of surgeon volume on patient outcome in total knee arthroplasty: a systematic review of the literature. *BMC Musculoskelet Disord* 2012;13:250.
37. Holom GH, Hagen TP. Quality differences between private for-profit, private non-profit and public hospitals in Norway: a retrospective national register-based study of acute readmission rates following total hip and knee arthroplasties. *BMJ Open* 2017;7:e015771.
38. Quaife M, Terris-Prestholt F, Di Tanna GL, *et al*. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *Eur J Health Econ* 2018;19:1053–66.
39. Taylor T, Scott A. Do Physicians Prefer to Complete Online or Mail Surveys? Findings From a National Longitudinal Survey. *Eval Health Prof* 2019;42:41–70.
40. Szawlowski S, Taylor T, Scott A, *et al*. *MABEL User Manual: Wave 10 Release. Melbourne Institute of Applied Economic and Social Research*. The University of Melbourne, 2019.