Social Science & Medicine 294 (2022) 114730

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Do patients and other stakeholders value health service quality equally? A prospect theory based choice experiment in cataract care

Aline C. Stolk-Vos [a,b,c,*], Arthur E. Attema [b], Michele Manzulli [d], Joris J. van de Klundert [b,e]

[a] Rotterdam Ophthalmic Institute, Rotterdam, the Netherlands
[b] Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[c] Department of Psychiatry, Section Medical Psychology and Psychotherapy, Erasmus Medical Center, Rotterdam, the Netherlands
[d] Rotterdam Eye Hospital, Rotterdam, the Netherlands
[e] Prince Mohammad Bin Salman College of Business & Entrepreneurship, King Abdullah Economic City, Saudi Arabia

ARTICLE INFO

Keywords:
Quality of health care
Prospect theory
Healthcare quality indicators
Shared decision making

ABSTRACT

Objectives: This study aims to compare the valuation of health service quality by patients and other stakeholders through a case study in cataract care.
Methods: The valuation of health service quality by Dutch patients, ophthalmologists and healthcare purchasers involved in cataract care are elicited by a prospect theory-based measurement task. Respondents stated preferences for probabilities and scores for the clinical indicator Complication (posterior capsular rupture with vitreous loss) and the patient-reported experience measure Information Provisioning (the ophthalmologist provides sufficient information about risks of cataract surgery to the patient). Our subject pool (n = 256) consisted of 90 ophthalmologists, 125 cataract patients, and 41 healthcare purchasers employed by health insurance companies.
Results: Following prospect theory, respondents were loss averse, and risk averse for gains. However, utilities differed from prospect theory, especially the concave utility for losses. Patients were significantly more loss averse than the other respondents, more subject to a pessimistic view on losses, and had significantly more concave utility for losses, especially for the clinical quality indicator Complications. For each of the stakeholders, the results differed significantly between the two essentially different quality indicators.
Conclusions: The heterogeneous valuations of patients and other stakeholders invalidate commonly applied cataract care quality assessment frameworks. Incorporating loss aversion, pessimism and concave utility for losses can remedy existing shortcomings. The valuation differences between patients and other stakeholders emphasize the need for communication and shared decision making in patient-centered treatment, purchasing and policy.

1. Introduction

Patients and other stakeholders in healthcare increasingly consider healthcare quality indicator scores when choosing and evaluating health services and providers. In many countries, forms of public reporting and ranking have been made available, based on Patient Reported Outcome Measures (PROMs) and Patient Reported Experience Measures (PREMs) in addition to clinical and health outcome measures. However, patients attach difference importance to the quality indicators figuring in these reporting frameworks and rankings than other stakeholders, such as physicians, management, health insurance companies, and policy makers (Stolk-Vos et al., 2017). This hampers the validity of current quality frameworks that express healthcare quality as a weighted sum of indicator scores, as for instance used by insurers for purchasing purposes and in quality rankings published to guide patient decision making (ealth. Implementing, 2020; Klasa et al., 2018; Steenhuis et al., 2020). The methodological shortcomings of the underlying linear additive logics based on expected utility are well documented (Caroff et al., 2020; Hota et al., 2016; Hofstede et al., 2019; Lingsma et al., 2010; Roessler et al., 2019; Lichtman et al., 2019), as are the disproportional effects on hospital reputation (Cua et al., 2017; Mehta et al., 2020).

This study investigates alternative valuation methods of healthcare quality with higher validity, especially with regard to valuation by patients. To this purpose we conduct a case study investigating quality

* Corresponding author. P.O. Box 70030, 3000 LM, Rotterdam, the Netherlands.
E-mail address: a.stolk@oogziekenhuis.nl (A.C. Stolk-Vos).

https://doi.org/10.1016/j.socscimed.2022.114730
Received 11 June 2021; Received in revised form 23 November 2021; Accepted 14 January 2022
Available online 19 January 2022

evaluations of patients receiving cataract care, as well as the evaluations of two other important stakeholder groups of cataract care, ophthalmologists and purchasers employed by health insurance companies (henceforth purchasers). Cataract surgery is selected for our study purposes as it is one of the oldest and most frequently performed surgical procedures worldwide (Liu et al., 2017). The condition is still responsible for 5% of blindness in developed countries and 50% of blindness in low- and middle-income countries (Liu et al., 2017). Several quality frameworks and indicator sets for cataract care have been developed and are widely implemented by stakeholders across the globe (Mahmud et al., 2015; Stolk-Vos et al., 2021).

There is a growing body of evidence showing that the valuation of health outcomes partially follows Prospect Theory (PT, see Text Box) and therefore is not a linear function of health outcomes. These studies have focused on utility and/or probability weighting of health outcomes, such as QALYs and life expectancy, and involved respondents from proxy groups, the general population, and patients [e.g., 16–19]. Instead of such general health outcome indicators, our study considers two commonly adopted indicators which are of specific importance in cataract care quality frameworks.

In Prospect Theory (PT), every individual has an initial condition for assessment, a *reference point*, which may for instance represent the current status. The valuation of increases above the reference point, called *gains*, and of decreases below the reference point, called *losses*, varies with the proximity to the reference point (*reference dependence*) (Tversky and Kahneman, 1992). More specifically, PT assumes that the value function is S-shaped with the reference point in the flection (Tversky and Kahneman, 1981). Furthermore, PT proposes a kink at the reference point, reflecting a steeper value function for losses than for gains. This phenomenon is known as *loss aversion* and indicates that losses loom larger than gains of the same magnitude (Kahneman and Tversky, 1979). Finally, individuals tend to transform probabilities into decision weights in a nonlinear fashion, usually underweighting high probabilities and overweighting low probabilities (*probability weighting*). The resulting probability weighting function has an inverse S-shape and may be different for gains and losses (Liu et al., 2017).

The aim of our study is to provide a deeper and more accurate understanding of the valuations of quality indicators scores by relevant stakeholders and of the differences in valuation between patients and other relevant stakeholders. We therefore conduct a case study on the valuation of cataract care quality involving patients, ophthalmologists and purchasers in the Netherlands. More specifically, our research tests the hypotheses that the valuations of these stakeholders are different and follow PT.

By providing a more accurate scientific understanding of the (differences in) valuation of the quality in cataract care by patients and other stakeholders, the results are intended to contribute to resolving the aforementioned shortcomings of quality frameworks commonly applied in practice and corresponding negative effects on patient centeredness of cataract care.

2. Methods

2.1. Study design

The preferences of the stakeholders regarding the valuation of healthcare quality measures are elicited by a bisection procedure. In the bisection procedure, respondents are repeatedly asked to choose between two options to elicit their valuations (Abdellaoui et al., 2008). We now first describe the selection of indicators and subsequently turn to the bisection procedure and the analysis methods.

We selected the indicators and their corresponding levels based on a previously determined list of 125 items to measure quality in cataract care. The 125-item list was generated through a systematic search of the scientific and grey literature in Embase, PubMed, Scopus, and Google (Stolk-Vos et al., 2017). The 125 items were previously clustered into seven quality dimensions using Concept Mapping (Stolk-Vos et al., 2017) and their importance was rated on a scale from 1 to 5 by all relevant stakeholders. In a first round, indicators were selected from that list according to five criteria:

1. *Interpretability.* All respondents are competent to interpret the indicator. For example, the indicator does not contain medical terminology.
2. *Importance.* The selected indicator must have an average importance score of 4.5 or higher for each of the three stakeholder groups involved (as reported in (Stolk-Vos et al., 2017)).
3. *Continuous outcome.* The indicator must be a continuous outcome measure to facilitate bisection.
4. *Variation.* The indicator value can vary across settings, to facilitate realistic differences in values between settings in the choices made during the bisection process.
5. *Data availability.* Empirical data is available for the indicator to construct realistic choice sets.

Nine of the 125 items on the list met the selection criteria. Six of these nine items regarded the quality dimension 'patient experience' (Stolk-Vos et al., 2017). Out of these six items, we selected the item rated as most important within the dimension 'patient experience' (Stolk-Vos et al., 2017), which was the PREM addressing the information provision to a patient by her/his ophthalmologist. Of the other three indicators, we selected an indicator of a very different, more technical, nature, from the quality dimension 'clinical outcomes'. This indicator regarding complications relates to treatment effectiveness and is of importance in value-based healthcare (Mahmud et al., 2015). The resulting two selected distinct indicators covering patient-reported data and clinical data are:

1. Complication: posterior capsular rupture with vitreous loss
2. Ophthalmologist gives sufficient information about risks of cataract surgery to patient

While they are distinct and from different quality dimensions, these two indicators together are not intended to form a proxy of the much wider (seven dimension) construct of quality of cataract care. These indicators are further referred to as 'Complications' and 'Information Provision'.

We investigate preferences over two-outcome lotteries (x,p;y,1-p), giving outcome x with probability p ($0 < p < 1$) and outcome y with probability 1-p. We assume respondents behave according to PT, with reference-dependent preferences with respect to a reference point r. Gains are outcomes that are strictly preferred to r and losses are outcomes strictly less preferred to r. Gain prospects involve no losses, loss prospects involve no gains, and mixed prospects involve both a gain and a loss.

We use commonly adopted parametric shapes to model the utility function and the probability weighting function. For utility, we estimate the power function ($U(x) = x^{\alpha}$ for gains and $U(x) = -(-x)^{\beta}$ for losses), with $U(x)$ the utility of outcome x, $\alpha, \beta > 0$, and $\alpha < 1$ [> 1] implying a concave [convex] utility for gains, $\beta < 1$ a convex [concave] utility for losses, while α, $\beta = 1$ implies linear utility.

We model probability weighting with Prelec's (Prelec, 1998) one-parameter function: $w^{i}(p) = \exp\{-(-\ln(p))^{j}\}$, where $w^{i}(p)$ represents the decision weight given to probability p, $i = +$,- (i.e., we have separate weighting functions for gains and losses), and $j = \gamma$ for gains and $j = \delta$ for losses. For $0 < j < 1$, this function has an inverse S-shape, with overweighting of small probabilities and underweighting of large probabilities. For complications, this implies that respondents would give too much weight to a small probability of fewer complications, too little weight to higher probabilities, and that they are not very sensitive to changes in intermediate probabilities. Hence, for $0 < j < 1$, this function causes insensitivity to probabilities in the middle, and extreme sensitivity to changes from impossible to possible (e.g., a slight change from $p = 0$ to $p = 0.01$) and from possible to certain (e.g., from $p = 0.99$ to $p = 1$). Expected utility theory (Neumann and Morgenstern, 1947) is the special case of this function when $j = 1$, in which case there is no probability weighting.

Loss aversion is modelled by multiplying the utility of losses by the loss aversion index λ. Respondents are classified as loss averse if $\lambda > 1$, gain seeking if $\lambda < 1$, and loss neutral if $\lambda = 1$. Here, $\lambda > 1$ implies that respondents give more weight to deteriorations in Complications and Information Provision than to comparable improvements. Appendix A gives a derivation of our regression equations that result from these models.

2.2. Data collection

Data collection took place by digital and paper surveys. The digital version of the survey was built using Qualtrics. The survey started with background information explaining the study rationale, indicators, and levels. Sociodemographic data (age, sex) were collected to assess if these factors influenced stated preferences. Completion of the survey took approximately 15–30 min.

The draft survey was administered to a sample of ophthalmologists, policy makers of an eye hospital and patients from the patient council of an eye hospital. The resulting feedback on the length, lay out and wording led to adjustments by consensus among the researchers. The piloted draft survey with the background information confirmed the interpretability of the somewhat technical complications indicator by patients.

Our subject pool (n = 256) consisted of 90 ophthalmologists, 125 cataract patients, and 41 purchasers (115 women, 141 men) recruited between September 2018 and Augustus 2019. The Committee of the Rotterdam Eye Hospital has formally considered the application and reached the committee decision that explicit approval by the medical ethical committee was not required, according to Dutch regulations, and hence that the committee had no objections to the study. All participants gave informed consent to participate.

Adult cataract patients were recruited at the outpatient clinic of the Eye Hospital Rotterdam by one ophthalmologist (MM). The ophthalmologist handed out envelopes containing an invitation letter, the paper version of the survey, a return envelope and a reimbursement form for a book receipt. 250 patients received an envelope with questions regarding indicator I and 200 patients received an envelope with questions regarding indicator II. The ophthalmologist registered who received an envelope. Patients received a phone call from a research assistant to remind them to complete the survey. Patients were offered a book voucher of €10 for participation.

Ophthalmologists and ophthalmologists in training at the Eye Hospital Rotterdam (n = 65) were invited by a researcher (AS). After a presentation about the study during a clinical meeting, they received an email with a link to the survey. They were reminded to participate by email and in person (by AS and MM). All other ophthalmologists in The Netherlands (n = 675) were recruited by post letter. The letter contained a QR-code and a short link to fill in the digital version of the full survey. Ophthalmologists were reminded to participate by post letter.

Purchasers employed by health insurer companies were recruited by contacting one or more employees at the health insurer company who subsequently invited healthcare purchasing professionals within their own organization by email with a survey link. Five companies were contacted and willing to participate. The combined market share of the health insurer companies included is around 90%.

Ophthalmologists and purchasers were incentivized to participate by a donation of €10 to Aravind Eye Hospital India made by the researchers for every fully completed survey.

The bisection procedure is a common way to elicit preferences in economic experiments (Abdellaoui et al., 2008; Bostic et al., 1990). In short, a bisection procedure elicits indifferences between two options by requesting several iterative choices. One of the options remains fixed throughout the entire list, while the other becomes more attractive or less attractive, conditional upon the previous choice. The resulting indifference point gives an indication of the preferences of the respondent.

The experiments always started with the task on Complications. Since complications are a bad outcome, i.e., people generally prefer to have fewer complications, a reduction in complications is considered a gain, while an increase is seen as a loss. In the task, respondents were instructed to consider two hospitals. One hospital had one full-time ophthalmologist the respondent would see for sure when choosing that hospital. The other hospital had two ophthalmologists both working part-time. When choosing that hospital, the respondent would be assigned to either of these two ophthalmologists depending on chance. The probability p of being assigned to either of the two ophthalmologists equalled the fraction of the week (s)he was on duty.

For each ophthalmologist, a specified number of complications per 1000 surgeries was given. The instructions mentioned that the national average number of complications was 100 per 1000 surgeries, which was used as the reference point r. In the gain part, the number of complications was always smaller than or equal to 100 per 1000. In the loss task the number of complications was always larger than or equal to 100 per 1000 (with a maximum of 200).

For example, let Hospital A have part-time Ophthalmologist 1 on duty 2 days per week with 50 complications out of 1000 (a gain of 50 relative to the reference point of 100 complications per 1000) and Ophthalmologist 2 on duty the other 3 weekdays, with 80 complications out of 1000 (a gain of 20). The best outcome is to see Ophthalmologist 1, which has probability $p = 0.4$ (2/5). The worst outcome is to see Ophthalmologist 2, with probability $1 - p = 0.6$ (3/5). The alternative choice is to select Hospital B where the respondent sees Ophthalmologist 3 for sure (i.e., $p = 1$). The complication rate of Ophthalmologist 3 in Hospital B varies in the experiment between the complication rates of the two ophthalmologists in Hospital A. In our example, the complication rate for Hospital B varies between 80 per 1000 and 50 per 1000 and therefore the gain varies between 20 and 50. Now, the lower the gain in Hospital B for which the respondent would be indifferent between Hospitals A and B, the more risk averse is the respondent. A respondent that prefers Hospital B if it offers a complication rate of 70 per 1000 (gain of 30) to avoid the risk of being assigned to Ophthalmologist 2 in Hospital A (with complication rate 80 per 1000) and forego the opportunity to see Ophthalmologist 1 in Hospital A (complication rate 50 per 1000), is more risk averse than a respondent who only prefers Hospital B if the complication rate in Hospital B is 60 per 1000 or less (and therefore willing to take the risk of being assigned to Ophthalmologists 2 in Hospital A if the complication rate in Hospital B is 70 per 1000).

We asked 5 choice questions in the gain part and 5 in the loss part, in both tasks. This number sufficed to enable estimation of the parameters and was not too cognitively demanding.

We elicited fulltime equivalents (FE's) from these binary choices. Both for Complications and for Information Provision, FE's describe the outcome in Hospital B that a respondent accepts to be indifferent between the outcomes of Hospitals A and B. For each question, an FE was estimated as the mean of the largest sure gain that was turned down and the smallest sure gain that was preferred to Hospital A. If all sure gains were chosen, the FE was estimated as the mean of the smallest sure gain of Hospital B and the worst possible outcome of Hospital A. If no sure gain was chosen, the FE was estimated as the mean of the largest gain of Hospital B and the best possible outcome of Hospital A.

The loss part was the same as the gain part except that the two possible outcomes in Hospital A were worse than the national average. To elicit loss aversion, a mixed-prospect bisection procedure was used. Table 1 presents the outcomes offered for the Complications task.

The design of the Information Provision task was similar to the design of the complications task. Respondents were asked to imagine the same situation with two hospitals, one with one ophthalmologist and

Table 1
Stimuli for the "Complications" task.

Number	Hospital A (p,x; 1-p, y)	Hospital B
Losses		
1	(0.5, 200; 0.5, 100)	CL1
2	(0.5, 200; 0.5, 150)	CL2
3	(0.9, 175; 0.1, 100)	CL3
4	(0.7, 200; 0.3, 100)	CL4
5	(0.35, 175; 0.65, 100)	CL5
Gains		
6	(0.5, 0; 0.5, 100)	CG1
7	(0.5, 50; 0.5, 100)	CG2
8	(0.1, 0; 0.9, 75)	CG3
9	(0.3, 0; 0.7, 100)	CG4
10	(0.65, 0; 0.35, 75)	CG5
Mixed		
11	(0.5, CM; 0.5, 150)	100

one with two part-time ophthalmologists. Now, the outcome was replaced by the relative number of times sufficient information was provided. Again, five indifferences were elicited, for each of the gain, loss and mixed tasks. Table 2 presents the stimuli for the Information Provision task.

2.3. Data analysis

The outcomes were normalized to facilitate comparison between the tasks. For Complications, outcomes were divided by 100, resulting in a normalized value in the range [−1,1]. For Information Provision, we divided outcomes by 250.

The parameters of functions 1 and 2 were estimated by nonlinear regression (Attema et al., 2013). The gain parameters α and γ were estimated simultaneously using the responses to questions 6 to 10 (from Tables 1 and 2). The same was done for the loss parameters β and δ with the responses to questions 1 to 5 (from Tables 1 and 2). The loss aversion coefficient λ was assessed by means of the indifference value obtained from the responses in the mixed prospect together with the other parameters obtained (see. Appendix A).

3. Results

3.1. Reference dependence, probability weighting, loss aversion

For all respondents together, Table 3 shows the medians and interquartile ranges for the five parameters α, β, γ, δ, λ (medians are shown instead of averages because of some outliers). To facilitate interpretation of the results, let us recall that PT hypothesizes $\alpha < 1$, i.e., concave utility for gains and $\beta < 1$, i.e., convex utility for losses. Moreover, it proposes $\gamma, \delta < 1$, representing overweighting of small probabilities and underweighting of large probabilities, and $\lambda > 1$, reflecting loss aversion.

The results in the first two columns of Table 3 reject the hypothesis that the valuation follows PT. For gains, the utility power estimates of the reference dependent valuations are not significantly different from 1 ($p = 0.364$ for Complications and $p = 0.227$ for Information Provision). Interestingly, for losses, the estimates are significantly higher than 1 (instead of less than 1), confirming reference dependence yet contradicting PT. Comparing the two tasks, we see that respondents have a more concave utility function for losses for Information Provision than for Complications ($p < 0.01$).

Columns 3 and 4 of Table 3 confirm PT regarding probability weighting for gains and for losses for both tasks ($p < 0.01$), indicating probabilistic pessimism (Wakker, 1994). Column 5 confirms PT as loss aversion indices are higher than 1 for both tasks ($p < 0.01$). Respondents were more loss averse for Information Provision than for Complications ($p < 0.01$), while there are no differences in the probability weighting ($p = 0.888$ for losses and $p = 0.652$ for gains).

Table 4 presents the median parameters estimates for patients, ophthalmologists and purchasers separately. It suggests several differences between patients and the other stakeholders, as is confirmed by statistical tests. Patients have more concave utility for losses ($p < 0.01$ for ophthalmologists vs. patients, $p = 0.02$ for purchasers vs. patients), are more subject to probabilistic pessimism for losses ($p < 0.01$ for both comparisons) and are more loss averse ($p < 0.01$ for ophthalmologists vs. patients, $p = 0.056$ for purchasers vs. patients). We find no such differences for gains (only marginally significantly higher γ for purchasers, $p = 0.072$). Moreover, no differences are observed between ophthalmologists and purchasers.

In addition, interesting differences emerge when comparing the two tasks for each of the respondent groups. For both ophthalmologists and purchasers, we find that utility for losses is more concave for Information Provision (Wilcoxon signed ranks test, $p < 0.03$) and loss aversion is higher for Information Provision ($p < 0.01$). For patients however, utility of Complications was significantly more concave ($p < 0.04$), even though the number of respondents who did both tasks was low ($n = 13$). A between-subjects test in which we could include more respondents did not confirm this finding (Mann-Whitney test, $p = 0.34$), although we did find significantly more probability weighting for losses for the Complications task there ($p < 0.01$).

In addition to the above, we found a positive correlation between age and loss aversion for Complications ($p < 0.05$). Moreover, older respondents have more convex gain utility and more concave loss utility for Complications ($p < 0.01$). Older people also have more probability weighting for losses ($p < 0.02$ for Complications and $p < 0.05$ for Information Provision). No gender effect is present, except for probability weighting for gains in Information Provision, where women show marginally significantly more probability weighting ($p < 0.06$).

Finally, we ran ordinary least squares regressions where we combined these explanatory variables in one model (Table 5). The results revealed significant effects for patients on β and λ for Complications. For age there were some marginal effects. This suggests that the significant differences between patients and other respondents were driven by their respondent type rather than by their older age.

3.2. Risk aversion

The majority of choices (around 60%) were risk averse for both Complications and Information Provision, both for gains and for losses. However, there were some significant differences in risk aversion within tasks, depending on the probabilities (Friedman tests, $p < 0.01$). These differences were consistent with the usual pattern predicted by PT, with more [less] risk aversion for higher [lower] probabilities of the best outcome. This phenomenon is known as the fourfold pattern of risk

Table 2
Stimuli for the Information Provision task.

Number	Hospital A (p,x; 1-p, y)	Hospital B
Losses		
1	(0.5, 650; 0.5, 750)	IL1
2	(0.5, 675; 0.5, 750)	IL2
3	(0.1, 650; 0.9, 700)	IL3
4	(0.3, 650; 0.7, 750)	IL4
5	(0.65, 650; 0.35, 700)	IL5
Gains		
6	(0.5, 1000; 0.5, 750)	IG1
7	(0.5, 1000; 0.5, 875)	IG2
8	(0.9, 950; 0.1, 750)	IG3
9	(0.7, 1000; 0.3, 750)	IG4
10	(0.35, 950; 0.65, 750)	IG5
Mixed		
11	(0.5, IM; 0.5, 825)	750

Table 3
Medians and interquartile ranges (IQR) of parameter estimates for both tasks.

	α	β	γ	δ	λ
Complications					
Median	0.953	**1.321**	**0.400**	**0.254**	**1.356**
IQR	0.671–1.477	0.885–2.052	0.237–0.686	0.032–0.778	0.476–2.788
N	199	193	199	193	186
Information					
Median	1.012	**1.341**	**0.239**	**0.533**	**1.898**
IQR	0.673–1.538	0.912–2.027	0.059–0.738	0.192–0.896	0.167–122.6
N	147	147	147	147	146

*Bold numbers reflect a significant difference from 1 (Wilcoxon signed ranks test).

Table 4
Median parameter estimates per subject group (interquartile ranges in parentheses).

	α	β	γ	δ	λ
Complications					
Ophthalmologist (n = 77)	0.973 (0.764–1.541)	1.189 (0.901–1.534)	0.383 (0.153–0.740)	0.524 (0.156–0.909)	1.117 (0.427–1.717)
Purchaser (n = 37)	0.988 (0.831–1.199)	1.320 (1.018–1.611)	0.483 (0.321–0.720)	0.592 (0.245–0.943)	1.327 (0.774–2.058)
Patient (n = 81)	0.891 (0.313–1.733)	1.765 (0.765–16.417	0.297 (0.237–0.686)	0.146 (0–0.433)	2.348 (0.475–45,513)
Information					
Ophthalmologist (n = 59)	1.071 (0.654–1.559)	1.320 (0.946–2.011)	0.239 (0–0.663)	0.414 (0.120–0.829)	1.450 (0.158–136.4)
Purchaser (n = 34)	0.965 (0.853–1.150)	1.351 (1.143–1.662)	0.399 (0.092–0.709)	0.448 (0.135–0.760)	2.923 (0.679–16.16)
Patient (n = 52)	0.961 (0.346–4.922)	1.341 (0.637–6.587)	0.230 (0.071–2.043)	0.673 (0.503–0.898)	1.777 (0-inf.)

Table 5
Regression estimates (standard errors in parentheses).

Explanatory variable	α Comp	γ Comp	β Comp	δ Comp	λ Comp	α Info	γ Info	β Info	δ Info	λ Info
Dummy Male	0.46 (0.39)	0.14 (0.14)	−0.77 (0.75)	0.12 (0.10)	−2912 (4548)	−0.41 (0.42)	0.22 (0.23)	0.09 (0.76)	−0.19 (0.17)	Inf. (inf.)
Dummy Patient	−0.16 (0.54)	0.26 (0.19)	3.41 (1.06)***	−0.07 (0.14)	15,968 (6397)**	1.07 (0.57)[a]	0.22 (0.31)	1.05 (1.04)	0.21 (0.23)	Inf. (inf.)
Dummy Purchaser	−0.76 (0.54)	−0.13 (0.19)	−0.01 (1.03)	−0.12 (0.13)	−1217 (6165)	−0.47 (0.53)	−0.07 (0.29)	−1.22 (0.96)	0.01 (0.21)	Inf. (inf.)
Age	0.03 (0.02)[a]	−0.01 (0.01)	0.04 (0.03)	−0.01 (0.004)[a]	229 (193)	−0.01 (0.02)	0.01 (0.01)	0.01 (0.03)	0.01 (0.01)	Inf. (inf.)

Dependent variable.
[a] Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

(Kahneman and Tversky, 1979). Figs. 1 and 2 plot the average risk premiums against the probabilities of the best outcomes in the lotteries. Figs. 3 and 4 does the same for losses, where the average risk premiums are shown as a function of the probability of the worst outcome of the lotteries.

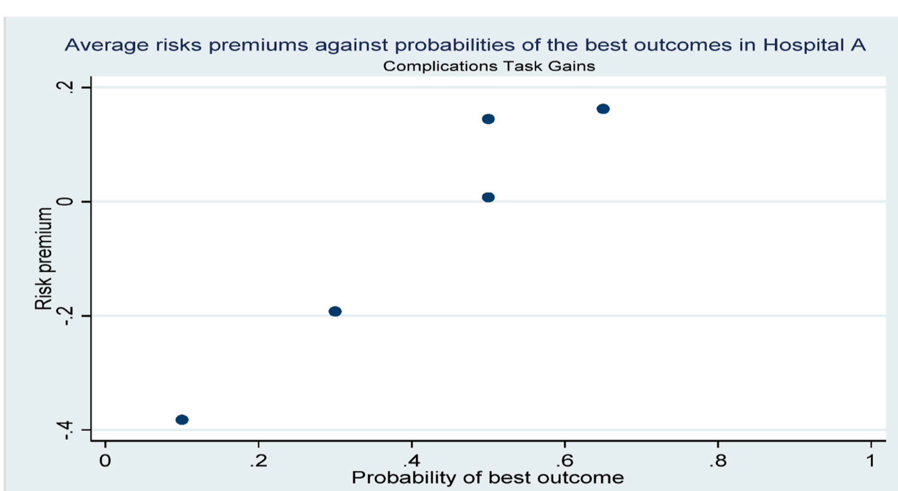

Fig. 1. Average risks premiums against probabilities of the best outcomes in Hospital A for Complications.

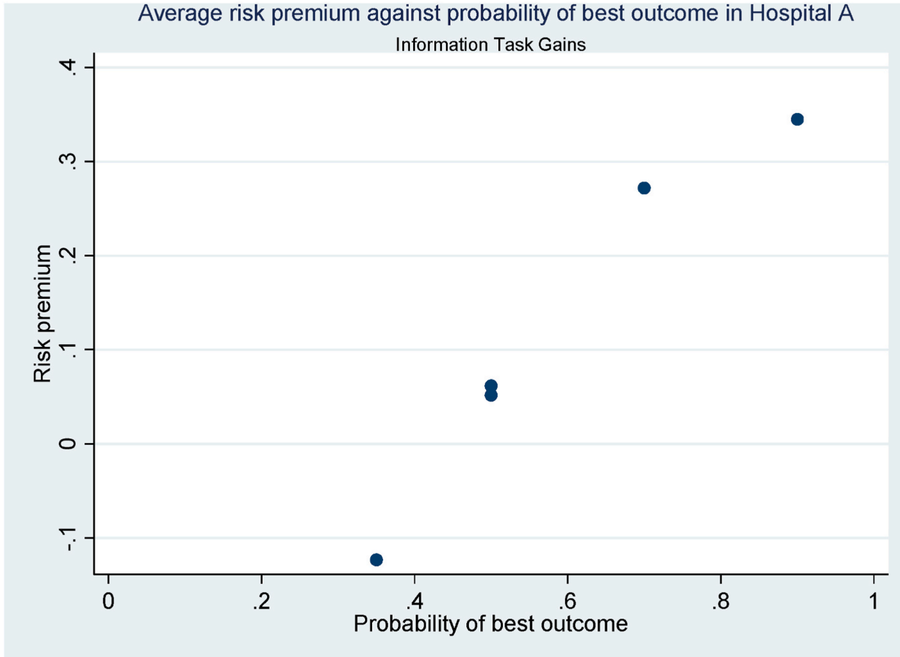

Fig. 2. Average risks premiums against probabilities of the best outcomes in Hospital A for Information Provisioning.

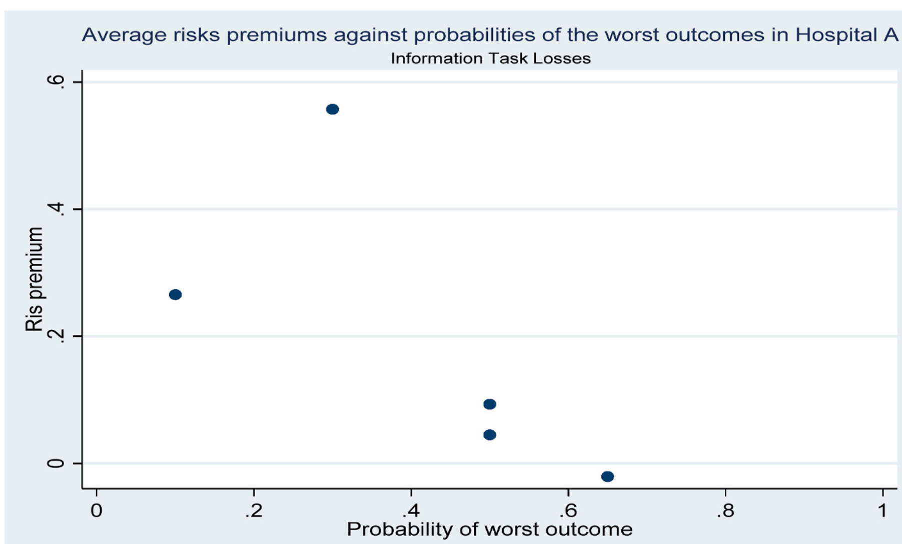

Fig. 3. Average risks premiums against probabilities of the worst outcomes in Hospital A for Complications.

4. Discussion

This study provides the first quantitative estimation of valuation of health service quality by patients and other stakeholders using PT. The stated patient preferences differed significantly from the preferences of ophthalmologists and healthcare purchasers, and partially followed PT. Our results are not the first evidence partly supporting PT in the healthcare domain (Winter and Parker, 2007; Lipman et al., 2019; Attema et al., 2013; Rouyard et al., 2018). However, the study is the first to include actual patients and other stakeholders as respondents to evaluate quality indicator scores rather than health outcomes. Moreover, our study is explicit about risk framing and advances beyond additive linear expected utility-based risk modelling as recently called for (Harrison et al., 2014).

4.1. A prospect theory perspective

In conformance with PT, we found significant loss aversion and an inverse S-shaped function for probability weighting. In contrast to PT however, we found no significant deviations from expected utility for gains and a concave utility function for losses. Moreover, the value functions of the patients differed significantly from those of the ophthalmologists and purchasers. Taken together, our results therefore invalidate existing practical frameworks and expected utility-based models valuing healthcare quality as a weighted sum of indicator scores. Such frameworks tend to disregard nonlinear utility for losses, los aversion, and probability weighting, all of which especially applied to patients.

Our finding of concave utility for losses confirms previous studies in the health domain (Attema et al., 2013, 2016) and provides further

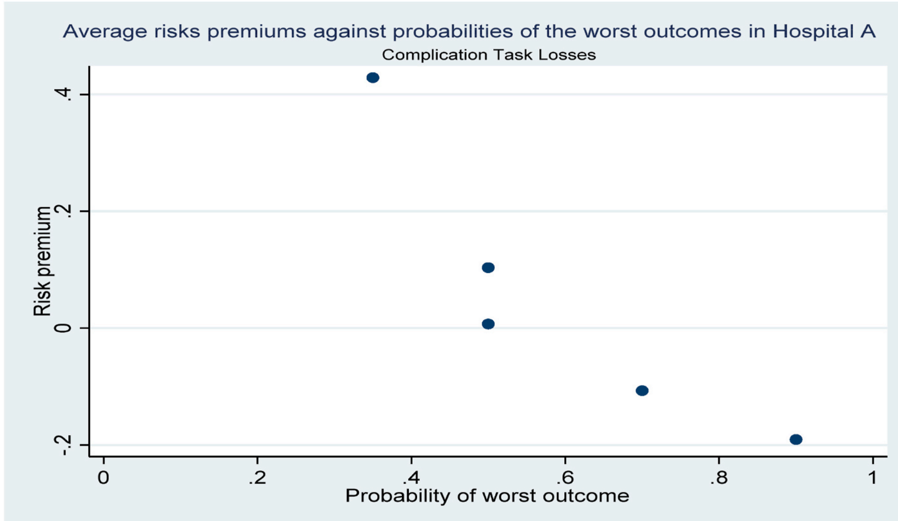

Fig. 4. Average risks premiums against probabilities of the worst outcomes in Hospital A for Complications.

evidence that valuation within the health domain is different from valuation in the monetary domain (Attema et al., 2018; Galizzi et al., 2016; Wakker and Deneffe, 1996). Together with the large loss aversion values found, the concavity reveals that especially patients weigh quality losses increasingly heavily. This is further exacerbated by probability weighting in case variation in quality increases. This risk aversion regarding the quality indicators on the highly standardized treatment cataract surgery contrasts with the risk seeking behaviors found for the progressive disease MS for which no effective standardized cure is presently known (Murino et al., 2021).

Patients' valuations deviate more from expected utility than the valuations of ophthalmologists and purchasers. Patients gave more weight to losses and were more risk averse for losses, especially regarding Complications. By contrast, ophthalmologists and healthcare purchasers were more loss averse for Information Provision than for Complications. Further research is needed to understand why the valuation of quality differs between stakeholders and indicators, e.g., is it different for clinical outcomes than for PROMs and PREMs (Groene et al., 2015)?

4.2. Empirical findings on perspectives of patients, ophthalmologists and purchasers

The differences found between the risk and quality preferences of patients and other stakeholders emphasize the importance of including the patient perspective in quality assessment and shared decision making (Quentin et al., 2019; Vaughn et al., 2019). A patient-centered approach requires ophthalmologists and purchasers to make the patient's valuation leading and not to follow their own valuations of quality measures and risks. If, however, one perceives patient valuations to deviate too much from expected utility, then purchasers and physicians need to better inform patients or correct for these biases after learning the patient preferences (Winter and Parker, 2007).

4.3. Limitations

The strength of including actual stakeholders in our study may in turn bring along some limitations. The complexity of the task may impact the quality of response and lead to bias in respondent groups, e.g., excluding older patients and time-pressed professionals. Another limitation of our method might be that all patients are recruited by one physician at one hospital. A third limitation might be the relatively small number of purchasers included, even though respondents cover almost all Dutch healthcare insurers. To strengthen validity and reliability, we encourage future studies to include patients from multiple ophthalmologists and hospitals and to be conducted in larger and/or multiple countries.

5. Conclusions

The identified heterogeneity in the valuation of quality indicator scores for cataract care invalidates commonly adopted quality assessment frameworks and therefore has implications for the construction of such frameworks and for cataract care provisioning, purchasing and policy. To be representative of stakeholder quality valuation, and specifically of quality valuations by patients, frameworks need to adopt nonlinear valuations of quality scores which express the loss aversion of patients, concave utility for losses, and the probability weighting of variation in outcomes instead of being solely based on average scores. For decision making on the services provided to individual patients, the results give new forms of support to the importance of communication and shared decision making when aiming for patient-centered care and for the practice to incorporate communication and shared decision making in treatment guidelines, purchasing practices, and regulatory policy.

Credit author statement

AA, AS, JK: Conceptualization, Methodology **AS, MM**: Investigation, Resources **AS**: Project administration **AA, AS**: Software, Writing -Original draft, Visualisation **AA**: Formal analysis, Data curation **JK**: Supervision **JK, MM**: Writing - Review & Editing.

Acknowledgement

This project could not have been completed without the cooperation of stakeholder representatives. We thank the patients, ophthalmologist, and employees at health insurances companies for their participation in this study. We are also thankful to Jan van Busschbach who provided expertise during the project and comments to the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2022.114730.

References

Abdellaoui, M., Bleichrodt, H., l'Haridon, O., 2008. A tractable method to measure utility and loss aversion under prospect theory. J. Risk Uncertain. 36 (3), 245.

Attema, A.E., Brouwer, W.B.F., l'Haridon, O., 2013. Prospect theory in the health domain: a quantitative assessment. J. Health Econ. 32 (6), 1057–1065.

Attema, A.E., Brouwer, W.B.F., L'Haridon, O., Pinto, J.L., 2016. An elicitation of utility for quality of life under prospect theory. J. Health Econ. 48, 121–134.

Attema, A.E., Bleichrodt, H., L'Haridon, O., Peretti-Watel, P., Seror, V., 2018. Discounting health and money: new evidence using a more robust method. J. Risk Uncertain. 56 (2), 117–140.

Bostic, R., Herrnstein, R.J., Luce, R.D., 1990. The effect on the preference-reversal phenomenon of using choice indifferences. J. Econ. Behav. Organ. 13 (2), 193–212.

Caroff, D.A., Wang, R., Zhang, Z., Wolf, R., Septimus, E., Harris, A.D., Jackson, S.S., Poland, R.E., Hickok, J., Huang, S.S., Platt, R., 2020. The limited utility of ranking hospitals based on their colon surgery infection rates. Clin. Infect. Dis. 9 ciaa012.

Cua, S., Moffatt-Bruce, S., White, S., 2017. Reputation and the best hospital rankings: what does it really mean? Am. J. Med. Qual. 32 (6), 632–637.

EIT Health, 2020. Implementing Value-Based Health Care in Europe: Handbook for Pioneers. Gregory Katz, Director.

Galizzi, M.M., Miraldo, M., Stavropoulou, C., 2016. In sickness but not in wealth: field evidence on patients' risk preferences in financial and health domains. Med. Decis. Making 36 (4), 503–517.

Groene, O., Arah, O.A., Klazinga, N.S., et al., 2015. Patient experience shows little relationship with hospital quality management strategies. PLoS One 10 (7), e0131805.

Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J., Payne, K., 2014. Risk as an attribute in discrete choice experiments: a systematic review of the literature. Patient 7 (2), 151–170.

Hofstede, S.N., Ceyisakar, I.E., Lingsma, H.F., Kringos, D.S., Marang-van de Mheen, P.J., 2019. Ranking hospitals: do we gain reliability by using composite rather than individual indicators? BMJ Qual. Saf. 28 (2), 94–102.

Hota, B., Webb, T.A., Stein, B.D., Gupta, R., Ansell, D., Lateef, O., 2016. Consumer rankings and health care: toward validation and transparency. Joint Comm. J. Qual. Patient Saf. 42 (10), 439–446.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–291.

Klasa, K., Greer, S.L., van Ginneken, E., 2018. Strategic purchasing in practice: comparing ten European countries. Health Pol. 122 (5), 457–472.

Lichtman, J.H., Leifheit, E.C., Wang, Y., Goldstein, L.B., 2019 Feb. Hospital quality metrics: "America's best hospitals" and outcomes after ischemic stroke. J. Stroke Cerebrovasc. Dis. 28 (2), 430–434.

Lingsma, H.F., Steyerberg, E.W., Eijkemans, M.J., Dippel, D.W., Scholte Op Reimer, W.J., Van Houwelingen, H.C., Netherlands Stroke Survey Investigators, 2010. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. QJM 103 (2), 99–108.

Lipman, S.A., Brouwer, W.B.F., Attema, A.E., 2019. The corrective approach: policy implications of recent developments in QALY measurement based on prospect theory. Value Health 22 (7), 816–821.

Liu, Y.C., Wilkins, M., Kim, T., Malyugin, B., Mehta, J.S., 2017. Cataracts. Lancet 390 (10094), 600–612.

Mahmud, I., Kelley, T., Stowell, C., et al., 2015. A proposed minimum standard set of outcome measures for cataract surgery. JAMA Ophthalmol. 133 (11), 1247–1252.

Mehta, R., Paredes, A.Z., Pawlik, T.M., 2020. Redefining the "Honor Roll:" do hospital rankings predict surgical outcomes or receipt of quality surgical care? Am. J. Surg. 220 (2), 438–440.

Murino, J., Sotoca, J., Sempere, Á.P., Brieva, L., López de Silanes, C., Caminero, A.B., Terzaghi, M., Gracia-Gil, J., Saposnik, G., 2021 Mar. High-efficacy disease-modifying therapies in people with relapsing-remitting multiple sclerosis: the role of risk attitude in treatment decisions. Patient 14 (2), 241–248.

Neumann, L., Morgenstern, O., 1947. Theory of Games and Economic Behavior. Princeton University Press, Princeton.

Prelec, D., 1998. The probability weighting function. Econometrica 66 (3), 497–527.

Quentin, W., Partanen, V.M., Brownwood, I., et al., 2019. Measuring healthcare quality. No. 53.. In: Busse, R., Klazinga, N., Panteli, D., et al. (Eds.), Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies [Internet]. Copenhagen: European Observatory on Health Systems and Policies. Health Policy Series, 3. Available at: https://www.ncbi.nlm.nih.gov/books/NBK549260.

Roessler, M., Schmitt, J., Schoffer, O., 2019. Ranking hospitals when performance and risk factors are correlated: a simulation-based comparison of risk adjustment approaches for binary outcomes. PLoS One 14 (12), e0225844.

Rouyard, T., Attema, A.E., Baskerville, R., Leal, J., Gray, A., 2018. Risk attitudes of people with 'manageable' chronic disease: an analysis under prospect theory. Soc. Sci. Med. 214, 144–153.

Steenhuis, S., Struijs, J., Koolman, X., Ket, J., van der Hijden, E., 2020. Unraveling the complexity in the design and implementation of bundled payments: a scoping review of key elements from a payer's perspective. Milbank Q. 98, 197–222.

Stolk-Vos, A.C., van de Klundert, J.J., Maijers, N., Zijlmans, B.L.M., Busschbach, J.J.V., 2017. Multi-stakeholder perspectives in defining health-services quality in cataract care. Int. J. Qual. Health Care 29 (4), 470–476.

Stolk-Vos, A.C., de Korne, D.F., Lamoureux, E., Wai, C., van Busschbach, J.J., van de Klundert, J.J., 2021. Multi-stakeholder perspectives in defining health services quality indicators and dimensions: a concept mapping based comparison for cataract care between Singapore and The Netherlands. BMJ Open 7;11 (4), e046226.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. 5, 297–323.

Vaughn, V.M., Saint, S., Krein, S.L., et al., 2019. Characteristics of healthcare organisations struggling to improve quality: results from a systematic review of qualitative studies. BMJ Qual. Saf. 28 (1), 74–84.

Wakker, P., 1994. Separating marginal utility and probabilistic risk aversion. Theor. Decis. 36 (1), 1–44.

Wakker, P., Deneffe, D., 1996. Eliciting von Neumann-morgenstern utilities when probabilities are distorted or unknown. Manag. Sci. 42 (8), 1131–1150.

Winter, L., Parker, B., 2007. Current health and preferences for life-prolonging treatments: an application of prospect theory to end-of-life decision making. Soc. Sci. Med. 65 (8), 1695–1707.