ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Health Policy Analysis

Ranking the Criteria Used in the Appraisal of Drugs for Reimbursement: A Stated Preferences Elicitation With Health Technology Assessment Stakeholders Across Jurisdictional Contexts

Wiesława Dominika Wranik, PhD,[1,2,3,*] Michał Jakubczyk, PhD,[2] Krzysztof Drachal, MA[4]

[1]School of Public Administration, Dalhousie University, Halifax, Canada; [2]College of Economic Analysis, SGH Warsaw School of Economics, Warszawa, Poland; [3]Department of Community Health and Epidemiology, Dalhousie University, Halifax, Canada; [4]Faculty of Economic Sciences, University of Warsaw, Warszawa, Poland.

ABSTRACT

Objectives: Our goal was to estimate the relative importance assigned to health technology assessment (HTA) criteria by stakeholders involved in the HTA process. HTA is an increasingly common framework used in the appraisal of drugs for public reimbursement. It identifies clinical, economic, social, and organizational criteria to be considered. The criteria can vary across jurisdictions and are typically appraised by multidisciplinary expert committees. Guidance on the relative weighing of criteria is often absent.

Methods: We elicited stakeholders' preferences using a single-scenario discrete choice experiment and a best-worst scaling model with conviction scores to assess the weights assigned to selected criteria by HTA stakeholders. We recruited 111 HTA stakeholders across multiple jurisdictions, including members of expert committees, clinical and economic experts, patients, and public payer representatives. Each judged twelve hypothetical cancer drug profiles for suitability for public funding and identified which characteristics were best and worst. In addition to standard discrete choice experiment and best-worst scaling models, we estimated a hybrid model to obtain a ranking of criteria by importance they played in the appraisal.

Results: A strong clinical benefit proved the most important criterion, followed by cost considerations, presence of adverse events, and availability of other treatments. The importance of clinical benefit was moderated by unmet need, adverse events, and number of patients.

Conclusion: Policymakers might want to consider providing an explicit weighing scheme, or moving to a 2-stage selection process with an assessment of the quality of clinical evidence as a gatekeeping step for a full HTA review.

Keywords: drug appraisal committees, health technologies assessment, selection criteria, stated preference elicitation.

VALUE HEALTH. 2020; 23(4):471–480

Introduction

Background

Pharmaceutical expenditures have grown as a percentage of healthcare budgets across Organisation for Economic Co-operation and Development countries. By 2018, 12% of health expenditures in the UK, 14% in Germany, 15% in Australia, 17% in Canada, and 20% in Poland were on pharmaceuticals.[1] Drug costs are expected to grow faster than other health expenditures.[2] Healthcare budget allocation must be approached with a systematic and rigorous consideration of supporting evidence. HTA provides a framework for incorporating evidence review into the budget allocation, and therefore HTA has become an important element of numerous healthcare systems.[3,4] Its role is to guide the use of evidence to support decisions about the funding of new drugs, therapies, or healthcare programs.

Health technology assessment is a multidisciplinary process to assess clinical, economic, social, and organizational implications of new health technologies. Assessment and subsequent appraisal

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Dalhousie University Research Ethics Board (REB File # 2017-4238), which is compliant with the Canadian Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

* Address correspondence to: Wiesława Dominika Wranik, PhD, School of Public Administration, Faculty of Management, Dalhousie University, 6100 University Avenue, Halifax, Nova Scotia, B3H 2R2, Canada. Email: dwl@dal.ca

https://doi.org/10.1016/j.jval.2019.10.012

are typically conducted by multidisciplinary committees of experts. Examples of such committees include the Technology Appraisal Committees in the National Institute for Health and Care Excellence (NICE) in the United Kingdom,[5] the Common Drug Review and pan-Canadian Oncology Drug Review (pCODR) expert review committees in the Canadian Agency for Drugs and Technologies in Health,[6] the Transparency Council in the Agency for Health Technology Assessment and Tariff System in Poland,[7] and the Pharmaceutical Benefits Advisory Committee in Australia.[8]

The HTA assessment phase consists of a review and quality rating of evidence that is guided by well-developed scholarly standards. The appraisal phase consists of the collective judgment by committee members about the clinical benefit and value for money of the therapy based on the considered evidence package. The collective judgement is arrived at via a deliberative process that aims to reconcile the individual members' positions. The committee members judgments' are rooted in their preferences, which vary by discipline, as well as members' individual differences in training, experience, and perspective. Additionally, committees are often requested to consider a variety of perspectives (eg, public, patients, and clinicians). Although guidelines emphasize the importance of a well-balanced HTA review, further instructions on how criteria are to be weighed relative to one another are typically missing.

The goal of this study was to estimate the relative importance assigned by stakeholders involved in the HTA process to selected characteristics of cancer drugs that are being considered for reimbursement.

Specifically, we addressed three related questions:

1. What is the relative importance of drug characteristics in the reimbursement recommendation?
2. How confident are HTA stakeholders in their choices when faced with different combinations of drug characteristics?
3. Do stakeholders' preferences vary across professional backgrounds and jurisdictions?

HTA stakeholders included members of HTA committees, decision makers involved in drug reimbursement decisions, clinicians, and patients from several jurisdictions. We focused on a specific disease context (cancer) to increase the feasibility of the experiment. Cancer drugs are often considered unique and meriting special consideration, for example, separate funding or appraisal.[9–12] This study builds on a pilot study that was conducted in Canada.[13]

Literature

Two strands of literature assess the preferences of HTA stakeholders, as described in a recent review.[14] The first relies on revealed preference methods and uses the outcomes of committees' deliberations (ie, their recommendations) as data sources. This approach focuses on actual recommendations made, but it treats committees as one whole, in other words, assesses group preferences post-deliberation. The second relies on stated preference methods and uses data from experiments with committee members. This approach relies on the measurement of individuals' preferences, which allows unpacking of differences by individuals' characteristics. Our study fits into the second stream.

Revealed preferences of committees have been estimated in several jurisdictions based on available decision documents. Typically, variables describing clinical, economic, social, or organizational characteristics are extracted, and their contribution to the recommendation (yes, no, or an interim option) is estimated using regression analysis. For example, drivers of recommendations have been estimated in Australia,[15] Belgium,[16] Canada,[17,18] Poland,[19,20] Scotland,[21] South Korea,[22] Sweden,[23] the UK,[24–27] and Wales.[28]

The stated preference approach is less common, likely owing to the practical challenges of reaching respondents. We identified 5 studies that estimated the preferences of HTA stakeholders using this approach (see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.10.012).[29–33] Their samples included between 11 and 41 members of appraisal committees. Each used a different combination of attributes. All except Tappenden et al (2007) included clinical gain, and all but Skedgel (2016) included some measure of cost or cost-effectiveness. Koopmanschap et al (2010) conducted a pairwise comparison experiment with Dutch healthcare professionals,[29] Linley and Hughes (2013) with members of the All Wales Medicines Strategy Group,[31] and Whitty et al (2011) with members of the Pharmaceutical Benefits Advisory Committee.[32] Skedgel (2016) conducted a pairwise choice and constant-sum comparison with members of the pCODR.[33] Tappenden et al (2007) performed a binary choice experiment with members of NICE appraisal committees.[30] Stated preference methods have also been used to measure general public preferences across characteristics of new therapies in Australia[32,34] and Canada.[33]

We contribute to the literature in method and content—we obtain information from more respondents and more information from each respondent compared to previous studies. First, our experiment is the first to allow for comparison across jurisdictions and across respondent types. Previous HTA preference studies focused on committee members, some excluded other HTA stakeholders and focused on specific jurisdictions. Our study included multiple jurisdictions and stakeholder groups, allowing for the analysis of heterogeneity between groups and inter-group comparisons. Second, we applied a novel estimation technique that combines a single-scenario discrete choice experiment (DCE)[35] with a profile case best-worst scaling (BWS).[36,37] The combining of DCE with BWS has 2 advantages: (1) typically the BWS task is considered to be cognitively less strenuous,[38] and (2) a hybrid DCE/BWS model allows for the elicitation of more preference information from a given sample while limiting the respondents' cognitive burden.[39] This is appropriate in our context given the limited pool of HTA stakeholders. Third, we introduce a conviction score to the experiment that aims to measure respondents' degrees of conviction versus hesitation in their choice. This adds insight into the situations where respondents waver in what is ultimately a forced selection.

Methods

Pilot Study

The study was piloted on Canadian HTA stakeholders (n = 36) between September 2016 and February 2017.[13] We built on the pilot by (1) expanding the study population and (2) clarifying attributes and wording (see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.10.012). The experiment was originally designed in English and professionally translated into Polish and German. Details of model specifications and estimation are provided in Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.10.012.

Selection of Attributes and Levels

Attributes for the DCE and BWS (Table 1) were selected using a mix of the three approaches advocated or used in the literature[40–43]: (1) synthesis of attributes described in policy documents and other

Table 1. Attributes and levels for the discrete choice experiment and best-worst scaling.

Attribute	Description	Levels
Survival benefit	Information about the survival benefit in comparison to the comparator drug. The measure of survival (overall, progression free, etc.) was not specified. We expected clinically superior drugs to be most often approved, and uncertain to be most often rejected.	Clinically superior to comparator Near identical to comparator Uncertain
Added cost per patient	The additional cost of the new drug to the healthcare system on a per-patient basis. Specific values were not assigned given differential pricing between countries. We expected cost-saving drugs to be most often approved and high cost to be most often rejected.	Cost-saving Low High
Number of patients	The number of patients is expressed as high or low as per the respondents' own interpretation. We had no prior expectation of directionality, since a higher number of patients suggests a higher budget impact, but also suggests that more individuals benefit from funding.	Low High
Other treatment options	The availability of other treatment options for the same indication was expected to decrease the probability of approval, expecting that absence of other options created a greater concern for the patient group.	Not available Available
Adverse events	The severity of adverse events was indicated as 1 or 2 and 3 or 4. We expected that more severe adverse events (grades 3 and 4) would reduce the probability of approval.	Grades 1 and 2 Grades 3 and 4

studies, (2) consultation with experts, and (3) results from past pilot studies (see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.10.012). We excluded cost-effectiveness measures (eg, incremental cost-utility ratio) but included information about the relative clinical benefit, added cost, and patient numbers.

Respondents to the pilot study inquired about the use of qualitative levels. We chose not to quantify levels to increase the applicability of the questions to several healthcare contexts. For example, the costs per patient for the same drug vary between jurisdictions given differential (and confidential) pricing. Similarly, an overall survival advantage of 4 months can be meaningful in one clinical context and not meaningful in another. Qualitative levels have been used in other health preference literature.[39,44]

Experimental Design for the Single-Scenario DCE and BWS

The experimental design mirrored that of the pilot study[13] and was consistent with other published conjoint analyses in healthcare[45,46] and International Society for Pharmacoeconomics and Outcomes Research recommendations.[47] Two attributes were assigned three levels, and three attributes were assigned two levels ($3^2 2^3$). A full factorial design involving all 72 scenarios was not feasible given our estimated pool of potential respondents (~450), the response rate during the pilot experiment (~21%), and the need to present each respondent with a manageable number of choice tasks. We eliminated implausible scenarios from the full factorial. A binary experimental design was developed in R statistical software using the AlgDesign package. A randomized selection process (optBlock function) was used to identify a D-efficient set of 36 scenarios from the full factorial design, scenarios were randomly allocated to three blocks of twelve scenarios, and each respondent was randomly allocated to one of the three blocks.

Respondents were presented with hypothetical scenarios describing attributes of cancer drugs submitted for appraisal (Fig. 1) and were given three tasks: (1) indicate whether they would approve the drug for funding or not (DCE), (2) indicate on a 4-point scale the level of difficulty of making the selection (conviction score), and (3) identify which attribute they considered best (influenced their appraisal most strongly in the positive direction) and which they considered worst (influenced their appraisal most strongly in the negative direction) (BWS).

Our experiment simulates the policy context of HTA appraisal. Respondents are forced to make a selection in a single-scenario binary choice setting[30,37] rather than the pairwise comparison or multiple-scenario approach.[31,47,48] We add the profile BWS approach[36] for three reasons: (1) to differentiate between levels of attributes (unlike the BWS object case), (2) to obtain new information when combined with the DCE, and (3) to mimic consensus building, during which individuals might focus on the best or the worst attribute level in any scenario to justify their appraisal to the group.

Study Population

Our study population consisted of HTA stakeholders primarily from Australia, Canada, Germany, Poland, and the United Kingdom. Respondents included members of HTA committees and others engaged in the appraisal process who self-identified as clinical and economic experts, patients, or public payer representatives (see Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.10.012).

Data Collection

Data were collected between February 1 and May 18, 2018, via online anonymous questionnaires using the Opinio software.[48] We used a purposive and snowball sampling approach[49,50] aiming to reach current, past, and potential stakeholders in the HTA process. We did not do a power calculation for required sample size, as we had no a priori hypotheses to test. Invitations were sent directly to HTA committee members, HTA scholars, and select patient organizations. Invitations were sent using publicly available email addresses, the LinkedIn platform, and the UK Health Economics Distribution Superlist. Invitees were asked to distribute the invitation to their professional networks. Additionally, we approached HTA organizations with the request to distribute the invitation (Transparency Council in the Agency for Health Technology Assessment and Tariff System, Central and Eastern European Society of Technology Assessment in Health Care, Deutsche Agentur

Figure 1. A sample discrete choice experiment and best-worst scaling choice task.

SCENARIO 1

In your assessment, should this drug qualify for public reimbursement?

Criterion	Value			
Survival Benefit	*Uncertain*			
Added cost (per patient)	*Low*		Yes	☐
Number of patients	*Low*		No	☐
Other treatment options	*Not Available*			
Adverse events	*Grades 3 and 4*			

On a scale of 1 to 4, with 1 being very easy and 4 being very challenging, rate the difficulty of making this decision.

	1	2	3	4	
Very easy	○	○	○	○	Very challenging

			The best attribute of this drug is:	The worst attribute of this drug is:
Criterion	**Value**		(Pick one)	(Pick one)
Survival Benefit	*Uncertain*		☐	☐
Added cost (per patient)	*Low*		☐	☐
Number of patients	*Low*		☐	☐
Other treatment options	*Not Available*		☐	☐
Adverse events	*Grades 3 and 4*		☐	☐

für Health Technology Assessment, European Health Economics Association, European Network for Health Technology Assessment, Health Technology Assessment International, International Society for Pharmacoeconomics and Outcomes Research, pCODR) and selected patient organizations. Organizations did not report back regarding distribution. All data collection procedures performed in this study were in accordance with the ethical standards of the Dalhousie University Research Ethics Board, which is compliant with the Canadian Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans and with the 1964 Helsinki Declaration and its later amendments or comprable ethical standards.

Data Analysis

The specification of the model is presented in Appendix 4 in Supplemental Materials (found at https://doi.org/10.1016/j.jval.2019.10.012). For the estimation involving the DCE data only, we relied on the standard binomial logistic regression using R to estimate the impact of attribute levels on the probability of a positive recommendation. For the hybrid DCE/BWS modeling, we relied the flexible Bayesian approach using Just Another Gibbs Sampler. Non-informative priors made the result depend solely on the data. We estimated the impacts of attribute levels on three observables (the binary choice, the best characteristics, and the worst characteristic). Given that in the BWS we look into individual attributes, we did not remove any variables based on insignificance. The median values of posterior distributions were used as point estimates, and 95% credible interval was calculated using the percentile method.

Results

Respondents

A total of 111 respondents completed the DCE/BWS for a total of 1258 observations (on average respondents completed 11.3 scenarios). When respondents completed fewer than 12 scenarios (n = 11), we retained their responses to maximize the number of observations. Table 2 provides detailed information about respondents and observations: 60% of respondents had clinical or economic expertise, and 8% identified the patient or public payer

Table 2. Respondent and observation characteristics.

Respondents			Observations		
	n	%		n	%
Total number of respondents	111	100	Total number of observations	1258	100
Canada	32	29	Observations from Canada	373	30
Poland	38	34	Observations from Poland	426	34
United Kingdom	20	18	Observations from the United Kingdom	235	19
Germany	7	6	Observations from Germany	67	5
Australia	3	3	Observations from Australia	26	2
Other countries	11	10	Observations from other countries	131	10
Clinical experts	30	27	Observations from clinical experts	333	27
Economic experts	37	33	Observations from economic experts	422	34
Patient perspective	7	4	Observations from the patient perspective	24	2
Public payer perspective	7	4	Observations from the payer perspective	72	6
Member of appraisal committee	26	23	Observations from members of appraisal committees	306	24

Note. Respondents are the individuals who participated in the experiment. An observation is a completed choice task (up to 12 per respondent). Four respondents did not identify their expertise precisely.

perspective as their primary expertise. Twenty-three percent of respondents served on HTA committees.

Results of the DCE

Results of the conventional DCE model are presented in Table 3. Other specifications were tested and yielded similar results. The coefficient of each attribute level is positively associated with the probability of a "yes" recommendation (YES from here on) compared to its base level, whereas the odds ratio shows the increase in the odds of a YES. The odds of a YES increased by multiples of ten-folds (in each specification) when the survival benefit was superior compared to uncertain, and this was the most important criterion. The odds of a YES also increased with a decrease in costs from high to low or cost-saving, with lower adverse events, and with the lack of alternative treatment options for patients. The number of patients was not statistically significant. The probability of a YES spanned almost the entire 0%-100% range with the attributes and levels we used.

We tested for 2-way interactions between all the variables using a general-to-specific approach, starting with all interactions and removing the insignificant ones in succession. This process yielded one specification showing four statistically significant interactions (Table 3): a superior survival benefit interacted negatively with low adverse events, the lack of other treatment options, and a high number of patients, meaning that when superior survival benefit is observed, the 3 other criteria became less important than they otherwise would be. The lack of other treatment options interacted negatively with the cost-savings level.

Conviction in the DCE and BWS Votes

Descriptive statistics of the outcomes of the DCE recommendation by conviction score are presented in Table 4. In addition, a visual presentation of the BWS results is provided in Figure 2.

Table 4 reports the percentage of YES and NO that were made with varying levels of conviction. For example, the row labeled "YES + best: survival benefit" shows that of 739 positive choices, in 411 survival benefit was the considered the best characteristic, and that this selection was made with strong conviction in 42.1% of cases. Figure 2 shows that a superior survival benefit was selected as worst in most scenarios in which it appeared. In general, most choices were made with a strong degree of conviction (4 or 3). In the subset of choice tasks for which survival benefit was rated as the best or the worst characteristic, the conviction score increased for all recommendations. When added cost was selected as the best characteristic, the conviction score for the YES increased.

DCE-BWS Hybrid Model

The hybrid model allowed for the estimation of the contribution of each specific attribute level in reference to one common base level, thereby permitting a ranking of all attribute levels. The hybrid model controlled for respondent characteristics (primary expertise, expert committee membership, and country). For the results of the uncontrolled model, see Appendix 5 in Supplemental Materials (found at https://doi.org/10.1016/j.jval.2019.10.012).

In our case, "uncertain clinical benefit" was selected as the base attribute level because it had the lowest contribution to a YES. Table 5 shows that a superior survival benefit had the greatest contribution to a YES followed by the cost-saving characteristic, the absence of other treatment options for patients, and the low-cost characteristic. Relative to the base level, the odds of approval were 49.70 times greater when survival benefit was superior, 10.18 times greater when the drug was cost-saving, 9.87 times greater when no alternatives were available for patients, and 5.58 times greater when the drug had a low cost.

The multiplier effects on the (fixed) utility contribution of each attribute level in the multivariate model measures the impact of respondents' type of expertise, country, or membership on an appraisal committee on the odds that a particular attribute level contributes to an approval (Table 5). Economists attached more importance to added cost (multiplier is 1.18) and less importance to survival benefit (multiplier is 0.90) compared to all respondents. Respondents from Poland and the United Kingdom attached more importance to the availability of other treatments than did respondents from other countries. Polish respondents also placed greater stress on the number of patients. Members of appraisal committees were less concerned with adverse events than other respondents.

Table 3. Discrete choice experiment: logit regression results.

		Full model without interactions			Parsimonious model without interactions		Parsimonious model with interactions		
Attribute	**Level**	**Coefficient (95% CI)**	**OR**	***P* value**	**Coefficient† (95% CI)**	**OR**	**Coefficient (95% CI)**	**OR**	***P* value**
Probability of a "yes" recommendation									
- For base levels		3.70%			3.25%		0.95%		
- For highest levels		99.24%			99.30%		99.01%		
Constant		−3.26 (−3.81 to −2.74)	NA	<.001	−3.39 (−3.91 to −2.91)	NA	−3.33 (−4 to −2.72)	NA	<.001
Survival benefit (base: uncertain)	Near identical	−0.13 (−0.66 to 0.4)	0.88 (0.52-1.49)	.624	-	-	-	-	-
	Superior	3.77 (3.34-4.23)	43.30 (28.09-68.78)	<.001*	3.81 (3.38-4.27)	45.11 (29.38-71.40)	5.99 (5.06-7.01)	398.90 (157.03-1110.98)	<.001
Added cost per patient (base: high)	Low	1.33 (0.71-1.94)	3.78 (2.03-6.93)	<.001	1.39 (0.78-1.99)	4.02 (2.18-7.32)	1.80 (1.14-2.47)	6.06 (3.13-11.77)	<.001
	Cost-saving	2.29 (1.91-2.69)	9.84 (6.73-14.69)	<.001*	2.29 (1.91-2.69)	9.90 (6.77-14.78)	1.826 (1.25-2.45)	26.14 (12.89-58.24)	<.001
Number of patients (base: low)	High	–0.21 (–0.54 to 0.11)	0.81 (0.58-1.12)	.202	-	-	-	-	-
Other treatment options (base: available)	Not available	1.29 (0.95-1.63)	3.63 (2.59-5.13)	<.001	1.26 (0.94-1.59)	3.52 (2.55-4.89)	2.31 (1.63-3.04)	10.03 (5.12-20.97)	<.001
Adverse events (base: grades 3 & 4)	Grades 1 & 2	1.00 (0.69-1.33)	2.73 (1.98-3.79)	<.001	0.98 (0.67-1.31)	2.67 (1.95-3.70)	1.58 (1.55-1.97)	4.75 (3.17-7.20)	.006
Superior survival benefit X Adverse events grades 1 & 2		-	-	-	-	-	−1.64 (−2.35 to −0.94)	0.19 (0.10-0.39)	<.001
Superior survival benefit X Other treatment options not available							−1.10 (−1.93 to −0.33)	0.33 (0.15-0.72)	.006
Superior survival benefit X High number of patients		-	-	-	-	-	−0.70 (−1.30 to −0.17)	0.50 (0.27-0.88)	.018
Cost-saving X Other treatment options not available		-	-	-	-	-	−1.10 (−1.93 to −0.33)	0.33 (0.15-0.72)	.006

OR indicates odds ratio.
**P* value calculated as incremental dummy (ie, level 3 vs level 2).
†All significant with *P* value < .001 (level 3 tested vs level 2, ie, incremental dummy).

Discussion

The key results of our study are that in the simulated experiment as presented, clinical benefit was the most important relative to other considerations. Other important characteristics included the cost and the availability of alternative treatments to patients and selected interactions between them. These results were consistent across the approaches we used: the DCE, the conviction, and the DCE/BWS hybrid. In this section we discuss how these results compare to those of other studies, what the implications of these results are for HTA policy, and the limitations of our work.

Table 4. Percentage of discrete choice results made with varying conviction, by driving best-worst attributes.

DCE/ BSW response	n	Conviction score*			
		4	3	2	1
Yes in the DCE					
All yes recommendations	739	38.2	31.0	20.6	10.3
Yes + best: survival benefit	411	42.1	31.1	15.6	11.2
Yes + best: added cost	166	42.2	31.3	20.5	6.0
Yes + best: other therapies	140	25.0	30.7	33.6	10.7
No in the DCE					
All no recommendations	519	31.0	31.2	25.2	12.5
No + worst: survival benefit	260	38.1	30.8	21.9	9.2
No + worst: added cost	109	25.7	35.8	23.9	14.7
No + worst: adverse events	116	22.4	28.4	31.0	18.1

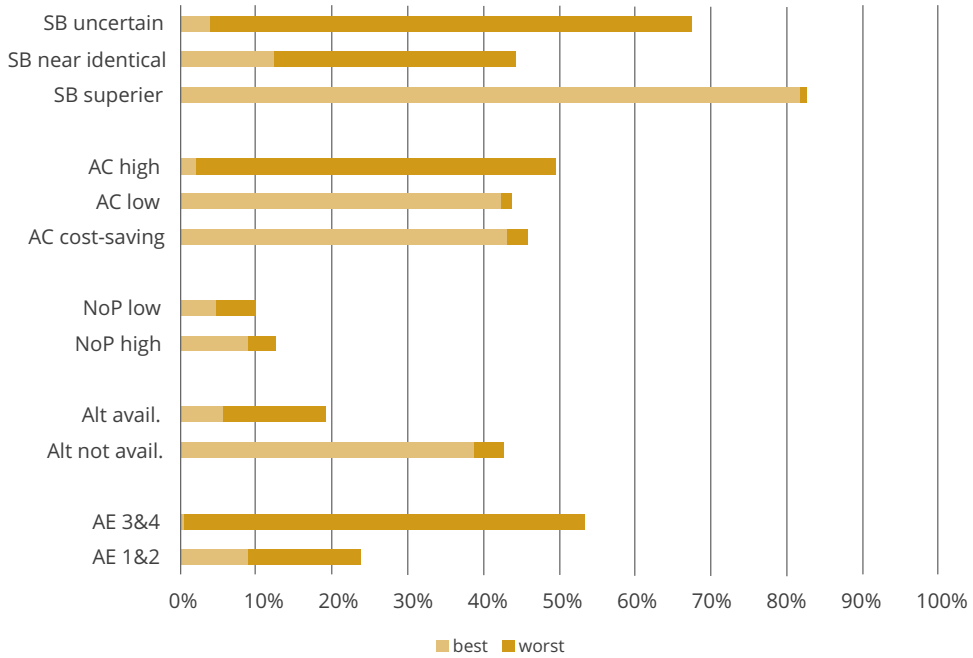
Note. Rows with too few data are omitted.
BWS indicates best-worst scaling; DCE, discrete choice experiment;
*The conviction score is based on the respondents' rating of the difficulty of the task (from 1 easy task to 4 difficult task). The conviction score is expressed as (5-difficulty rating). A score of 4 is strong conviction and 1 is weak conviction or hesitation.

How Do Our Results Compare to Previous Studies' Results?

Three of the five previous preference elicitation studies captured the clinical benefit as the number of quality-adjusted life years (QALYs gained and respondents made concrete tradeoffs between QALYs and costs or other characteristics.[29,31,33] The clinical benefit was found to be a significant driver of votes by both Koopmanschap et al (2010)[31] and Liney and Hughes (2013),[29] but the cost-effectiveness was a stronger driver in both studies. Their respondents likely considered the cost-effectiveness ratio to be a function of both costs and clinical benefit, thereby identifying a combined effect, whereas our study isolates respondents' perceptions about costs and clinical benefits. Tappenden et al (2009) did not include the QALY nor cost as isolated characteristics and their respondents also considered the cost-effectiveness ratio to be a strong driver of assessment.[30] Skedgel (2016) showed that QALY maximization was the most common choice.[33]

Studies of revealed preferences show similar results. At the committee level, clinical characteristics were found to be the strongest, if not exclusive, drivers of funding recommendations in Belgium, Canada, Poland, the United Kingdom, and Wales,[15,17-20,23,24,28] with clinical uncertainty being significant in Belgium and Wales[17,28] and not significant in Scotland.[21] Cost-effectiveness was a significant influence in all countries except Wales.[28] A few other factors found to have significant impact on appraisal were disease type,[16,21,23] process factors,[24] submissions from patient groups,[26] and the recommendations made elsewhere.[51]

Figure 2. Best-worst scaling responses. A visual presentation of BWS results provides an overview of the characteristics most often indicated as either best (gray) or worst (black) in the given scenarios.



AC indicates added cost (per patient); AE, adverse event; BWS indicates best-worst scaling; SB, survival benefit.

Table 5. Contribution of attributes and levels by respondent characteristics (hybrid discrete choice experiment/best-worst scaling with heterogeneity analysis).

Attribute	Level	Coefficient (95% CrI)	Odds ratio	Respondent has clinical expertise	Respondent has economic expertise	Respondent is from Poland	Respondent is from Canada	Respondents is from the United Kingdom	Respondent is a member of an appraisal committee
Constant		−6.673 (−7.516 to −5.792)	NA	NA	NA	NA	NA	NA	NA
Survival benefit	Uncertain	0.00	0	1.07 (0.96-1.19)	0.90 (0.81-0.98)	0.95 (0.850-1.048)	1.01 (0.906-1.130)	0.96 (0.85-1.08)	1.00 (0.90-1.10)
	Near identical	0.84 (0.60-1.11)	2.32 (1.81-3.03)						
	Superior	3.91 (3.36-4.48)	49.70 (28.62-88.44)						
Added cost per	High	0.185 (0.077-0.293)	1.20 (1.08-1.34)	1.11 (0.99-1.23)	1.18 (1.07-1.29)	1.09 (0.97-1.21)	1.07 (0.95-1.19)	1.12 (0.99-1.26)	0.96 (0.87-1.06)
	Low	1.72 (1.39-2.09)	5.58 (4.01-8.07)						
	Cost-saving	2.32 (1.96-2.69)	10.18 (7.12-14.73)						
Number of patients	Low	1.34 (1.07-1.62)	3.80 (2.90-5.03)	1.03 (0.89-1.19)	1.02 (0.89-1.16)	1.18 (1.01-1.38)	1.12 (0.96-1.30)	1.01 (0.84-1.21)	0.92 (0.79-1.05)
	High	1.38 (1.12-1.65)	3.96 (3.05-5.23)						
Other treatment options	Available	0.94 (0.752-1.137)	2.56 (2.121-3.118)	0.91 (0.81-1.02)	0.94 (0.85-1.04)	1.36 (1.21-1.52)	1.06 (0.94-1.20)	1.16 (1.01-1.32)	0.91 (0.81-1.00)
	Not available	2.29 (1.953-2.675)	9.87 (7.049-14.511)						
Adverse events	Grades 3 & 4	0.11 (−0.01 to 0.24)	1.12 (0.99-1.27)	1.00 (0.81-1.21)	0.93 (0.77-1.12)	1.05 (0.840-1.29)	1.08 (0.86-1.31)	1.00 (0.76-1.27)	0.82 (0.67-1.00)
	Grades 1 & 2	1.38 (1.08-1.70)	3.98 (2.95-5.48)						

CrI indicates credible interval; OR, odds ratio.
NA indicates not applicable.

Our experimental results further corroborate revealed preference results in terms of the non-influence of budget impact considerations.[15] The interaction between cost and number of patients was not significant. This could be explained by the policy reality across jurisdictions. In Canada, a recommendation of the HTA committee is implemented at the level of the provincial governments. Although a budget impact analysis is considered, the funding decision is not in the purview of the committee. Similarly, until recently, NICE committees were guided not to consider budgets. The recent introduction of the budget impact test is designed primarily to streamline the price negotiation process.

Economic considerations appeared to play a role in differentiating between unconditional and conditional positive recommendations.[15] Economic criteria including cost-effectiveness estimates and managed entry agreements were shown to be statistically significant drivers of reimbursement decisions in 6 European Union countries in a study that did not include the clinical benefit as an independent variable in its model.[50]

What Do Our Results Mean for Policy?

Our study shows that clinical benefit is the strongest predictor and superior clinical benefit is the most important characteristic from the perspective of HTA stakeholders. This is interesting since clinical evidence available across jurisdictions is largely the same, yet many reimbursement decisions differ across countries.[52,53]

In part, this could be explained given our interaction results, which suggest that other factors potentially contextualize the interpretation of clinical benefits. Respondents were more willing to tolerate higher adverse events in the presence of superior survival benefit, which is consistent with how patients might experience this tradeoff. When alternative treatment options were not available, respondents were more willing to forgo superior survival benefit and cost savings. This suggests that these considerations may be substitutes in the appraisal and respondents take the moral view that greater unmet need can justify lower value-for-money options. This perspective is also consistent with what is often discussed as important to patients, and our results corroborate that patients' perspectives can frame the interpretation of HTA criteria.[54] In contrast, when the number of patients was high, respondents were more willing to forgo superior survival benefit. We interpret this finding with caution given that in isolation, the number of patients was not statistically significant. It may suggest that respondents took a summative view in which small health gains for many people are treated as equivalent to large health gains to few people, a utilitarian perspective. Or perhaps a high number of patients was interpreted as a greater need in society.

If clinical evidence is the driving force, there may be value in considering a multistep HTA process. In the first step, the clinical criterion would be assessed and either qualify or disqualify the drug from a full HTA review. This is similar to the initial rapid

review used, for example, in Ireland,[55] although the rapid review includes economic criteria. Our suggestion is in line with the recent policy trends to collaborate on review processes across jurisdictions and reduce the multiplication of activities.[4,56,57] In addition, explication of the relative importance of the various criteria would support consistency across HTA stakeholders, and our results did suggest that weighing of criteria varied by stakeholder characteristics. Suggestions to add more explicit recommendation frameworks to the HTA process have been made by others as well, for instance, Thokala et al (2010) suggested applying a multi-criteria decision-making framework to HTA.[58] Explicit weighing schemes, of course, would be reflective of the preferences of some, not all stakeholder groups. A full discussion of the normative question of whose preferences we should consider and how is not in the scope of the current study but has been considered elsewhere.[59–61]

Limitations

Our study had limitations, which threaten its external validity and reduce the confidence we can place in policy recommendations drawn from it. First, any DCE necessarily simplifies reality into a manageable set of variables. Our selection process for attributes and criteria followed rigorous standards, nonetheless, a concern remains that relevant attributes or levels may have been omitted, in which case the weighing of the selected attributes would be biased. In addition, our experiment was situational in the context of cancer, which limits the validity of generalizing the results outside of this disease context. Second, the study sample is not random and we cannot be sure of representativeness. Distribution of the invitation to participate relied on a purposive selection of initial seed respondents or organizations and on subsequent snowballing, therefore we do not know who was invited. Furthermore, respondents self-selected to participate in the study. Of note is that all previous preference elicitation studies with HTA stakeholders suffer from sampling limitations owing to practical challenges given the nature of the study population. Additionally, some respondents terminated the experiment prematurely. If there is a correlation between respondents' preference patterns and their likelihood to complete the experiment (we see no reason to suspect such a correlation), the results of our study may be biased.

Conclusion

Our results add to the understanding of how HTA principles are applied in the practical policy setting. We have shown that HTA stakeholders consistently focus on a strong clinical benefit as the most relevant characteristic of a cancer drug. Some concerns have been voiced by the public that HTA agencies focus on cost-effectiveness alone,[62] but our respondents suggested otherwise. Cost attributes ranked below a strong clinical benefit, and recommendations based on cost attributes were made with less conviction.

Increasingly we are learning that the weights placed on HTA criteria are not balanced and that clinical criteria play the largest role. In pursuit of process efficiency, policymakers could consider relying on an initial screening of clinical evidence to qualify new drugs for a full HTA review. Furthermore, an explicit weighing framework to guide individual expert committee members could improve consistency of recommendations across groups and contexts.

Lastly, a stated preference experiment relies on simplifications of real-world situations. The preferences around complex decisions, and especially the interactions between drug characteristics, could be explored further with qualitative techniques that may be well suited to unpack the cognitive mechanisms applied by individuals when making such decisions.

Acknowledgments

The authors thank the following individuals for their thoughtful feedback and discussion during the design phases of the stated preference questionnaires: Dr Jeffrey Hoch (University of California Davis), Dr Stirling Bryan (University of British Columbia), Dr Mark Dobrow (University of Toronto), and Dr Tallal Younis (Dalhousie University). We would like to thank Dr Alexandra Lepucka for the professional translation and language services provided in support of this study.

This study was funded by the European Union Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 665778, as administered by the National Centre for Science POLONEZ Fellowship number UMO-2016/23/P/hs5/04134. It is based on a preceding pilot study that was funded by the Canadian Institutes for Health Research, Partnerships for Health Systems Improvement number PHE 129912.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.10.012.

REFERENCES

1. Organisation for Economic Co-operation and Development. Pharmaceutical spending (indicator) 2019. https://data.oecd.org/healthres/pharmaceutical-spending.htm. Accessed September 25, 2019.
2. Canadian Institute for Health Information. Where is the most money being spent in health care in 2017. https://www.cihi.ca/en/where-is-most-of-the-money-being-spent-in-health-care-in-2017. Accessed April 2, 2019.
3. Pantelli D, Busse R. Health technology assessment at age 25 – squaring the circle of strong methodology and context-dependency? *Health Policy*. 2019;123:115–117.
4. Barnieh L, Manns B, Harris A, et al. A synthesis of drug reimbursement decision-making processes in Organisation for Economic Co-Operation and Development countries. *Value Health*. 2014;17(1):98–108.
5. National Institute for Health and Care Excellence. Technology Appraisal Committee. https://www.nice.org.uk/get-involved/meetings-in-public/technology-appraisal-committee. Accessed March 3, 2019.
6. Canadian Agency for Drugs and Technologies in Health. https://www.cadth.ca/. Accessed March 3, 2019.
7. Agencja Oceny Technologii Medycznych i Taryfikacji. http://www.aotm.gov.pl/www/. Accessed March 3, 2019.
8. Australian Government Department of Health. Pharmaceutical Benefits Scheme. http://www.pbs.gov.au/info/industry/listing/participants/pbac. Accessed March 3, 2019.
9. Hoch J, Sabharwal M. Informing Canada's cancer drug funding decisions with scientific evidence and patient perspectives: the Pan-Canadian Oncology Drug Review. *Curr Oncol*. 2013;20(2):121–124.
10. Trudeau ME, Chambers A, Christiansen K, Mai H. Pan-Canadian Oncology Drug Review (pCODR): a unique model to support harmonization of cancer drug funding decisions in Canada. *J Clin Oncol*. 2018;36(30):s41.
11. Krzakowski M. The algorithm assessing the value of new anticancer drugs – the proposal of the Polish Society of Clinical Oncology and Polish Oncological Society (Algorytm oceny wartości nowych leków przeciwnowotworowych – propozycje Polskiego Towarzystwa Onkologii Klinicznej I Polskiego Towarzystwa Onkologicznego). *Onkol Prak Klin (Oncol Clin Pract)*. 2015;11(1):9–15.
12. National Institute for Health and Care Excellence. Cancer Drug Fund. https://www.england.nhs.uk/cancer/cdf/. Accessed March 3, 2019.
13. Wranik WD, Skedgel C, Hu M. Drug attributes associated with the selection of drugs for reimbursement: a pilot stated preferences experiment with Canadian stakeholders. *Expert Rev Pharmacoecon Outcomes Res*. 2018;19(1):59–69.
14. Ghijben P, Gu Y, Lanscar E, Zavarsek S. Revealed and stated preferences of decision makers for priority setting in health technology assessment: a systematic review. *Pharmacoeconomics*. 2018;36:323–340.
15. Harris A, Hill SH, Chin G, et al. The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994-2004. *Med Decis Making*. 2008;28(5):713–722.
16. Pauwels K, Huys I, De Nys K, Casteels M, Simoens S. Predictors for reimbursement of oncology drugs in Belgium between 2002 and 2013. *Expert Rev Pharmacoecon Outcomes Res*. 2015;15(5):859–868.

17. Skedgel C, Wranik D, Hu M. The relative importance of clinical, economic, patient values and feasibility criteria in cancer drug reimbursement in Canada: a revealed preferences analysis of recommendations of the pan-Canadian Oncology Drug Review 2011-2017. *Pharmacoeconomics*. 2018;36(4):467–475.
18. Rocchi A, Miller E, Hopkins RB, Goeree R. Common drug review recommendations: an evidence base for expectations. *Pharmacoeconomics*. 2012;30(3):229–246.
19. Malinowski KP, Kawalec P, Trabka W. Impact of patient outcomes and cost aspects on reimbursement recommendations in Poland in 2012-2014. *Health Policy*. 2016;120:1249–1255.
20. Niewada M, Polkowska M, Jakubczyk M, Golicki D. What influences recommendations issued by the Agency for Health Technology Assessment in Poland? A glimpse into decision makers' preferences. *Value Health Regional Issues*. 2013;2(2):267–272.
21. Charokopou M, Majer IM, de Raad J, et al. Which factors enhance positive drug reimbursement recommendation in Scotland? A retrospective analysis 2006-2013. *Value Health*. 2015;18(2):284–291.
22. Bae EY, Hong JM, Kwon HY, et al. Eight-year experience of using HTA in drug reimbursement: South Korea. *Health Policy*. 2016;120:612–320.
23. Svensson M, Nilsson FOL, Arnberg K. Reimbursement decision for pharmaceuticals in Sweden: the impact of disease severity and cost effectiveness. *Pharmacoeconomics*. 2015;33:1229–1236.
24. Cerri KH, Knapp M, Fernandez JL. Decision making by NICE: examining the influence of evidence, process and context. *Health Econ Policy Law*. 2014;9:119–141.
25. Dakin H, Devlin N, Feng Y, et al. The influence of cost-effectiveness and other factors on NICE decisions. *Health Econ*. 2015;24:1256–1271.
26. Dakin HA, Devlin NJ, Odeyemi IA. "Yes", "no" or "yes, but"? Multinomial modelling of NICE decision-making. *Health Policy*. 2006;77(3):352–367.
27. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ*. 2004;13:437–452.
28. Linley WG, Hughes DA. Reimbursement decisions of the All Wales Medicines Strategy Group: influence of policy and clinical and economic factors. *Pharmacoeconomics*. 2012;30(9):779–794.
29. Koopmanschap MA, Stolk EA, Koolman X. Dear policy maker: have you made up your mind? A discrete choice experiment among policy makers and other health professionals. *Int J Health Technol Assess Health Care*. 2010;26(2):192–204.
30. Tappenden P, Brazier J, Ratcliffe J, Chilcott J. A stated preference binary choice experiment to explore NICE decision making. *Pharmacoeconomics*. 2007;25(8):685–693.
31. Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity. *Pharmacoeconomics*. 2013;31:345–355.
32. Whitty JA, Scuffham PA, Rundle-Thiele SR. Public and decision maker stated preferences for pharmaceutical subsidy decisions: a pilot study. *Appl Health Econ Health Policy*. 2011;9(2):73–79.
33. Skedgel C. The prioritization of preferences on pan-Canadian Oncology Drug Review members and the Canadian public: a stated-preferences comparison. *Current Oncol*. 2016;23(5):322–328.
34. Whitty JA, Ratcliffe J, Chen G, Scuffham PA. Australian public preferences for the funding of new health technologies: a comparison of discrete choice and profile case best-worst scaling methods. *Med Decis Making*. 2014;34(5):638–654.
35. Muller K, Tao R, Goring S, Lane S. Use of discrete choice experiments designed with a single scenario and two or more choices: a systematic review. *Value Health*. 2016;19(3):A92.
36. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Econ Rev*. 2016;6(1):2.
37. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Phamacoecon Outcomes Res*. 2010;10(3):259–267.
38. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: what it can do for health care research and how to do it. *J Health Econ*. 2007;26:171–189.
39. Zhang J, Johnson FR, Mohamed AF, Hauber AB. Too many attributes: a test of the validity of combining discrete-choice and best-worst scaling data. *J Choice Modelling*. 2015;15:1–13.
40. Coast J, Al-Janabi H, Sutton EJ, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ*. 2011;21(6):730–741.
41. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. *J Health Services Res Policy*. 2007;12(1):25–30.
42. Bech M, Kjaer T, Lauridsen J. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Econ*. 2011;20(3):273–286.
43. Pedersen LB, Hess S, Kjaer T. Asymmetric information and user orientation in general practice: exploring the agency relationship in a best-worst scaling study. *J Health Econ*. 2016;50:115–130.
44. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008;26(8):661–677.
45. Green C, Gerard K. Exploring the social value of health-care interventions: a stated preferences discrete choice experiment. *Health Econ*. 2009;18:951–976.
46. Marshall D, Bridges JF, Hauber B, et al. Conjoint analysis applications in health – how are studies being designed and reported?: an update on current practices in published literature between 2005 and 2008. *Patient*. 2010;3(4):249–256.
47. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health*. 2013;16(1):3–13.
48. *Opinio Survey Software. Version 1*. Oslo, Norway: ObjectPlanet Inc.; 2018.
49. Babbie E, Roberts LW. The logic of sampling. In: Babbie E, Roberts LW, eds. *Fundamentals of Social Research*. 4th ed. Nelson Education Ltd.; 2018.
50. Etikan I, Bala K. Sampling and sampling methods. *Biom Biostat Int J*. 2017;5(6):215–217.
51. Pujoras LM, Cairns J. Why do some countries approve a cancer drug and others don't? *J Cancer Policy*. 2015;4:21–25.
52. Maynou L, Cairns J. What is driving HTA decision-making? Evidence from cancer drug reimbursement decisions from 6 European countries. *Health Policy*. 2019;123:130–139.
53. Nicod E, Kanavos P. Scientific and social value judgements for orphan drugs in health technology assessment. *Int J Technol Assess Health Care*. 2016;32(4):218–232.
54. Berglas S, Jutai L, MacKean G, Weeks L. Patients' perspectives can be integrated in health technology assessments: an exploratory analysis of CADTH Common Drug Review. *Res Involv Engagem*. 2016;2(21):1–13.
55. Murphy A, Redmond S. Rapid reviews with health-technology assessments in reimbursement systems – an examination of Ireland as a case study. *Global Reg Health Technol Assess*. 2016;4(1):e34–e40.
56. European Network for Health Technology Assessment. What are joint and collaborative assessments?. https://www.eunethta.eu/assessments/. Accessed April 2, 2019.
57. European Medicines Agency and European Network for Health Technology Assessment. EMA-EUnetHTA three-year work plan. https://www.ema.europa.eu/en/documents/other/ema-eunethta-three-year-work-plan-2017-2020_en.pdf. Accessed April 2, 2019.
58. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Health*. 2012;15:1172–1181.
59. Coast J. Is economic evaluation in touch with society's health values? *BMJ*. 2004;329:1233.
60. Buxton MJ, Chambers JD. What values do the public want their health care systems to use in evaluating technologies. *Eur J Health Econ*. 2011;12(4):285–288.
61. Kolasa K, Lewandowski T. Does it matter whose opinion we seek regarding the allocation of healthcare resources? – a case study. *BMC Health Services Res*. 2015;15(1):564.
62. Roman Topór-Mądry: Nie tylko efektywność kosztowa się liczy. http://www.medexpress.pl/roman-topor-madry-nie-tylko-efektywnosc-kosztowa-sie-liczy/71352. Accessed August 23, 2018.