REVIEW ARTICLE

Using the Threshold Technique to Elicit Patient Preferences: An Introduction to the Method and an Overview of Existing Empirical Applications

Brett Hauber[1] · Joshua Coulter[1]

Abstract

Patient preference information (PPI) is a topic of interest to regulators and industry. One of many known methods for eliciting PPI is the threshold technique (TT). However, empirical studies of the TT differ from each other in many ways and no effort to date has been made to summarize them or the evidence regarding the performance of the method. We sought to describe the TT and summarize the empirical applications of the method. Forty-three studies were reviewed. Most studies estimated the minimum level of benefit required to make a treatment worthwhile, and over half estimated the maximum level of risk patients would accept to achieve a treatment benefit. The evidence demonstrates that the TT can be used to elicit multiple types of thresholds and can be used to explore preference heterogeneity and preference non-linearity. Some evidence suggests that the method may be sensitive to anchoring and shift-framing effects; however, no evidence suggests that the method is more or less sensitive to these potential biases than other stated-preference methods. The TT may be a viable method for eliciting PPI to support regulatory decision-making; however, additional understanding of the performance of this method may be needed. Future research should focus on TT performance compared with other stated-preference methods, the extent to which results predict patient choice, and the ability of the TT to inform individual treatment decisions at the point of healthcare delivery.

1 Introduction

Recent initiatives by regulators and the pharmaceutical and medical device industries have resulted in an increase in demand for patient preference information (PPI). In 2016, the Center for Devices and Radiological Health (CDRH) of the US Food and Drug Administration (FDA) specifically cited a patient preference study [1] as being instrumental in informing the benefit–risk assessment and subsequent approval of a medical device [2]. The CDRH later issued guidance on the use of PPI in regulatory submissions [3]. Recently, the FDA began drafting a series of guidance documents in an effort to "systematically include the patients'… experience and preferences related to therapy, into drug development" [4] and has expressed its intention to systematically incorporate PPI into the assessment of trade-offs between benefits and risks [5]. At the same time, the

European Medicines Agency conducted a PPI pilot study and concluded that PPI may be useful in regulatory review of new drugs [6].

Multiple quantitative methods exist for eliciting PPI. In 2015, the Medical Device Innovation Consortium (MDIC) developed a catalog of patient preference methods [7]. In an overview of the empirical literature on welfare-theoretic approaches to quantifying benefit–risk preferences, Hauber et al. [8] described how quantitative patient preference methods can be used to estimate measures of risk tolerance for use in regulatory benefit–risk assessment. Recently, the Innovative Medicines Initiative's project Patient Preferences in Benefit and Risk Assessments during the Drug Lifecycle (IMI-PREFER) identified potential methods for eliciting PPI [9]. In the MDIC catalog [7] and the reviews by Hauber et al. [8] and IMI-PREFER [9], the threshold technique (TT) was identified as one potential method for eliciting patients' benefit–risk preferences. In addition, the CDRH used evidence from a TT study in its decision to expand the indications for use for home hemodialysis [10].

Because the TT has been identified as a method for quantifying the willingness of patients to trade-off between benefits and risks of treatments, it is important that researchers

✉ Brett Hauber
  abhauber@rti.org

[1] RTI Health Solutions, 3040 East Cornwallis Road,
    Research Triangle Park, NC 27709, USA

Key Points for Decision Makers
The demand for patient preference information to inform drug development and regulatory decision-making is growing.
The threshold technique (TT) is one of many potential methods for eliciting and quantifying patients' preferences.
A large number of empirical studies have demonstrated the use of the TT in eliciting the preferences of patients and other healthcare decision makers and the properties and reliability of the results of these preference estimates.
Future research on the TT should focus on comparing the performance of the TT to that of other stated-preference methods when applied to the same research question, measuring the ability of the results to predict patient choice, and the ability of the TT to inform individual treatment decisions at the point of healthcare delivery.

understand what is known about this method. Llewellyn-Thomas provides a description of the TT in the *Encyclopedia of Medical Decision Making* [11] and cites a number of empirical applications of the technique [12–20]. However, the TT has been used to estimate different types of thresholds in different disease areas more than the list of studies cited by Llewellyn-Thomas might suggest. In addition, the example described by Llewellyn-Thomas focused on the process of using an in-person interview to elicit an individual's preferences in a specific decision context rather than discussing the TT as a method of eliciting preferences for use in benefit–risk analysis.

To date, no overview of prior empirical applications of the TT or the implications of previous findings for using the TT in benefit–risk analysis exists. One reason such a study has not been conducted may be that the TT, as described by Llewellyn-Thomas [11], has been referred to using different names including, but not limited to, "modification of standard gamble" [21], "time trade-off" [22], "treatment trade-off" [23, 24], and "probability trade-off" [15, 18, 25–27]. Another reason may be that the technique has been applied as both a shared decision-making tool and a survey method. Our objective herein is to describe the TT, summarize the empirical literature employing this method (regardless of the name used to describe it), and discuss some of some the potential advantages and disadvantages of using this technique to elicit patients' preferences.

2 The Threshold Technique: An Introduction to the Method

In a TT exercise, a decision-maker—typically a patient or physician—is presented with a choice between two treatment or healthcare delivery options. One is the reference option that is the baseline against which an alternative is compared. It is often the option associated with the status quo or standard of care. The second is the target option and confers both an incremental benefit and an incremental burden relative to the status quo or standard of care. For cases in which both options represent potential alternatives to the status quo, which option is treated as the reference and which is treated as the target will depend on the researcher's objectives.

Once the reference and target options have been identified, the researcher must identify the key attribute of the target option (either a benefit or a burden) that will be used to estimate the strength of preference for the target relative to the reference option. The key attribute can be any attribute for which values can be expressed numerically. The most common key attributes are probability of benefit, risk of harm, waiting time, life expectancy, and cost. When the key attribute is a measure of burden (e.g., risk of harm, waiting time, or cost), the estimated threshold is a measure of the additional burden that exactly offsets the incremental benefit of the target option. If the key attribute is a benefit (e.g., probability of benefit or life expectancy), the estimated threshold is a measure of the minimum additional benefit that the target must provide to offset the incremental burden of that option.

After being presented with descriptions of the two options, respondents are asked to choose between them. In much of the empirical literature, the value of the key attribute is assumed to be the same in both the reference and target options in the initial question. This approach may be appropriate when one of the two options is unambiguously better than the other, when the clinically relevant levels of the key attribute associated with the reference and target options are not known with certainty, or when the reference and target options are purely hypothetical. However, when the reference and target options are real-world options with well-known attributes, using all available information in the initial question, including known differences between the options in the value of the key attribute, may provide a direct measure of decision makers' preference between the reference and target options in addition to providing a starting point for estimating the threshold value of the key attribute.

If the reference option is chosen in the initial question, the key attribute of the target is made better or more appealing and the question is repeated. If the target is chosen initially, the key attribute of the target is made worse or less appealing and the question is repeated. The process continues until

the researcher can identify the threshold level of the key attribute, i.e., the level at which a respondent is indifferent between the reference and target options. The difference between the threshold value of the key attribute and the level of the same attribute in the reference option is a measure of the strength of preference for the target option compared with the reference option. It is a measure of the change in the key attribute that exactly offsets the difference in benefit or burden between the reference and the target options.

The threshold can be a specific value or an interval within which the threshold lies. If the trade-off exercise yields a specific value, then that threshold for each respondent for each trade-off exercise is known. If, however, the trade-off exercise results in a threshold interval, then the researcher has options for how to utilize these data. The researcher can simply report the threshold interval [28] or the proportion of respondents choosing the target option at different threshold intervals [29]. Reporting the proportion of respondents who choose the target option or whose threshold lies within each interval can be informative; however, this does not provide a specific threshold estimate for the sample. If a specific threshold is necessary to answer the research question or to perform additional analyses but only interval data are provided, then the researcher must transform the interval into a specific threshold value. To derive a specific threshold value from a threshold interval, the researcher can assume that the threshold lies at the midpoint of the interval as is often done in swing weighting [30] or use an interval regression to estimate the mean value of the threshold for a given sample [31].

2.1 An Empirical Example of the Threshold Technique

To illustrate this method, consider an example from the literature. Devereaux et al. [27] used the TT to estimate the minimum required reduction in stroke risk that patients and physicians would require to accept the burdens of aspirin and warfarin in the treatment of atrial fibrillation. These researchers also elicited the maximum acceptable risk (MAR) of a severe bleed that would exactly offset the expected benefits of these antithrombotic therapies. Respondents were provided with descriptions of stroke and severe bleed and were informed that the 2-year risk of stroke and severe bleeding for a person without antithrombotic therapy would be 12% and 3%, respectively. Respondents were then informed of the cost and inconvenience of receiving aspirin or warfarin, including monitoring, hospital visits, and out-of-pocket costs of treatment. Each respondent was then presented with four TT exercises (Table 1).

In each trade-off exercise, patients were first asked to choose between having no antithrombotic therapy (the reference option) and either warfarin or aspirin (the target option). The order of the exercises was randomized for each respondent. In the first two exercises, respondents were presented with a choice between no antithrombotic therapy and either warfarin (Exercise 1) or aspirin (Exercise 2), each with a known 2-year risk of severe bleeding and the complete elimination of 2-year stroke risk (a 12 percentage point reduction). If the respondent chose treatment in the first question, then the risk of stroke was varied systematically until the minimum reduction in the risk of stroke required to offset the increase in the risk of severe bleeding was identified. Figure 1 presents this process for Exercise 1, which was used to estimate the minimum reduction in 2-year stroke risk that would make a 5% 2-year risk of severe bleeding worthwhile. In the third and fourth exercises, respondents were again presented with a choice between no antithrombotic therapy and either warfarin (Exercise 3) or aspirin (Exercise 4). In Exercise 3 (Exercise 4), the first choice was between no antithrombotic therapy and warfarin (aspirin) with a 4% (9%) 2-year risk of stroke and a 25% 2-year risk of severe bleeding. In each of these exercises, the 2-year risk of severe bleeding was varied systematically until the maximum level of risk that a patient would be willing to accept in exchange for the given reduction in stroke risk was determined. In each exercise, the difference between the threshold value of the key attribute and the value of that attribute in the reference condition was calculated. The mean risk differences, representing the minimum acceptable reduction in stroke risk and the MAR of severe bleed for both aspirin and warfarin, were calculated for both patients and physicians. These estimates are presented in the last column of Table 1.

A number of the results estimated by Devereaux et al. [27] are worth noting. First, physicians appear to have been more averse to bleeding risks than patients: they were not willing to accept the same increases in bleeding risk in exchange for a given reduction in stroke risks as were patients, and they required a larger benefit than did patients to accept a given increase in bleeding risk. Second, the ordering of the threshold estimates within each respondent group was consistent with expectations; i.e., the maximum acceptable level of risk is higher when treatment benefits are higher, and the minimum required benefit of treatment is higher when treatment risks are higher. Finally, distributions of estimated threshold values were wide in both the patient and physician samples across all exercises. Devereaux et al. [27] used univariate analysis to determine whether patients' or physicians' demographic characteristics explained this observed preference heterogeneity. None of the patient characteristics included in the analysis was a significant predictor of any threshold value; however, physicians' thresholds for bleeding risk were positively correlated with the number of their patients with atrial fibrillation; i.e., physicians who saw more patients had a higher tolerance for the risk of severe bleeding events.

Table 1 Summary of threshold technique exercises used by Devereaux et al. [27] to elicit preferences for antithrombotic therapies in atrial fibrillation

| Exercise	Reference option	Target option	Key attribute of target option	Threshold description	Percentage point change (threshold – reference rate) [mean (SD)]
1	No antithrombotic therapy	Warfarin resulting in a 5% 2-year risk of severe bleed (2 percentage point increase in risk of severe bleed)	2-year risk of stroke	Minimum required reduction in 2-year risk of stroke to accept warfarin therapy and a 2 percentage point increase in the risk of severe bleeding	Patients: 1.8 (1.9) Physicians: 2.5 (1.6)
2	No antithrombotic therapy	Aspirin with a 3.3% 2-year risk of severe bleed (0.3 percentage point increase in risk of severe bleed)	2-year risk of stroke	Minimum required reduction in 2-year risk of stroke to accept aspiring therapy and a 0.3 percentage point increase in the risk of severe bleeding	Patients: 1.3 (1.3) Physicians: 1.6 (1.5)
3	No antithrombotic therapy	Warfarin with a 4% risk of stroke (8 percentage point reduction in risk of stroke)	2-year risk of severe bleed	Maximum acceptable increase in the 2-year risk of serious bleed to achieve an 8 percentage point reduction in the risk of stroke with warfarin	Patients: 17.4 (7.1) Physicians: 10.3 (6.1)
4	No antithrombotic therapy	Aspirin with a 9% 2-year risk of stroke (3 percentage point reduction in risk of stroke)	2-year risk of severe bleed	Maximum acceptable increase in the 2-year risk of serious bleed to achieve a 3 percentage point reduction in the risk of stroke with aspirin	Patients: 14.7 (8.5) Physicians: 6.7 (6.2)

SD standard deviation
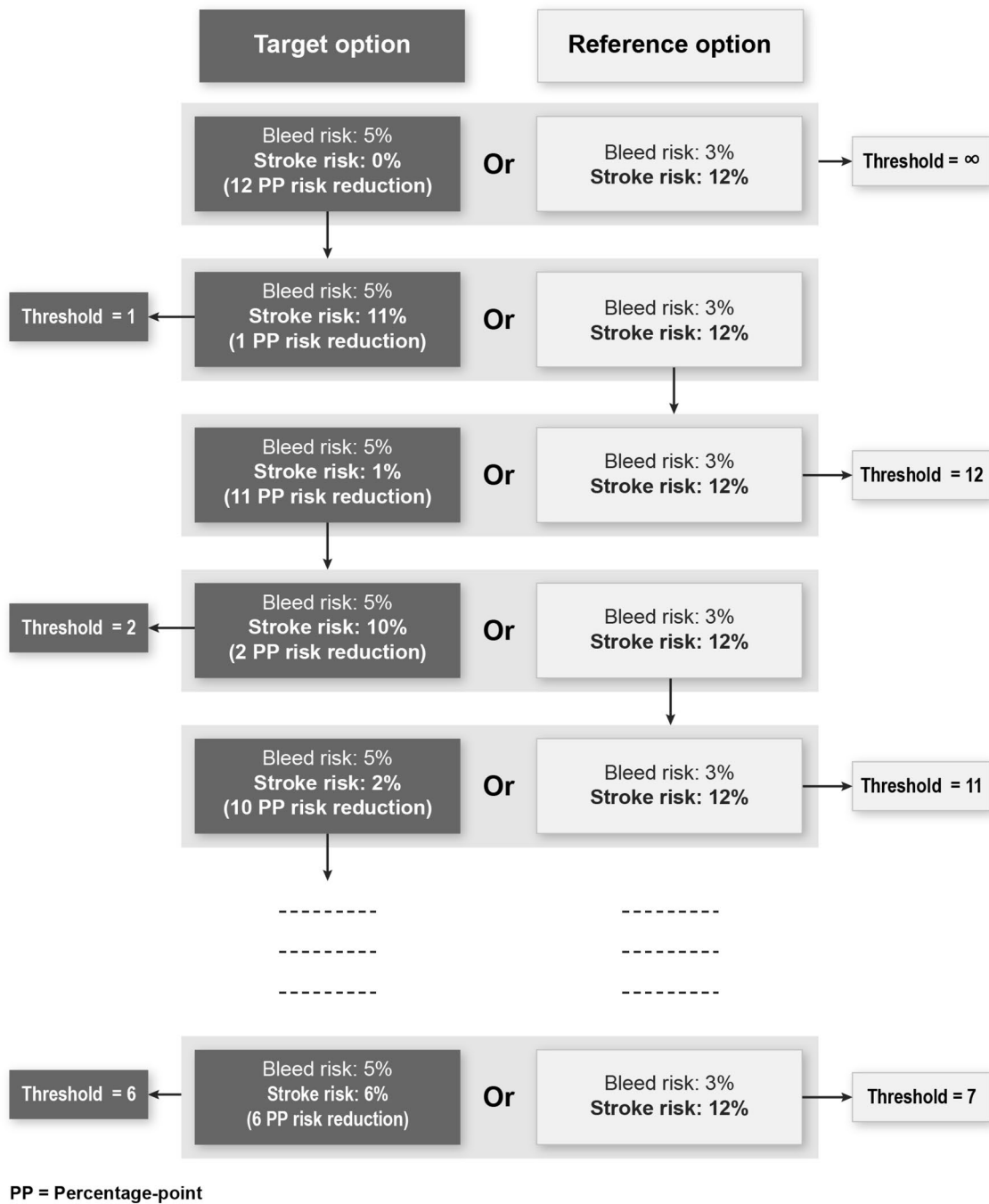
Fig. 1 Overview of series of threshold questions comparing warfarin with no treatment used by Devereaux et al. [27]. *PP* percentage point

3 Overview of the Empirical Literature

As noted in Sect. 1, many terms have been used to describe the method that Llewellyn-Thomas [11] presented as the TT. Therefore, conducting a prespecified systematic literature review was not feasible. The literature included in this overview was identified first by searching PubMed using the search terms *threshold technique* and *probability trade-off*. Additional studies were identified by reviewing the reference lists of the papers identified during the PubMed search. As additional names for the technique were identified, additional ad hoc searches were conducted until no additional papers were discovered. A final PubMed search was conducted in November 2018 to ensure that recently published papers were included in the review. Papers were included in this overview if they described a novel empirical

application of the TT to eliciting patients' or other decision makers' strength of preference for alternative healthcare options. Papers describing only the method or potential uses of results of the TT in decision-making were excluded, as were papers presenting the results of a study that was previously published. Studies that estimated only cost thresholds were also excluded because willingness-to-pay is typically not a consideration in regulatory decisions. Articles using the TT as described above were included regardless of the name the authors gave to the method. The final set of papers included 43 studies published between 1991 and 2016. The papers are listed by disease area in Table 2.

3.1 Type of Thresholds

In the example described in Sect. 2.1, Devereaux et al. [27] estimated patients' and physicians' willingness to accept an increase in the probability of bleeding risks for a given reduction in the rate of stroke and the minimum reduction in the probability of stroke that these decision makers would require to accept a given level of bleeding risk. Like Devereaux et al. [27], the majority of studies applied the TT to estimating probabilistic threshold values of minimum acceptable benefit (MAB) and MAR. These studies are summarized in Sect. 3.1.1.

3.1.1 Minimum Acceptable Benefit and Maximum Acceptable Risk

Twenty-eight TT studies (including Devereaux et al. [27]) estimated MAB. Among these, the majority quantified the minimum benefit required to accept the burden of cancer treatments [22, 32–40]. Numerous studies quantified the minimum required increase in the probability of survival resulting from cancer treatments from the perspective of patients and healthcare providers. The remaining studies estimating MAB in oncology are those that estimated the minimum reduction in the probability of local recurrence required to make treatments worthwhile to patients and physicians [20, 23, 25, 41]. Finally, two studies estimated the minimum probability of a cure that cancer patients would require to undergo chemotherapy [42, 43].

Twelve studies used the TT to estimate the minimum required probability of benefit in areas other than oncology. The types of MAB estimated in these studies include the minimum required reduction in the symptomatic recurrence of Crohn's disease [44], the minimum reduction in the probability of cardiovascular events in primary prevention [16, 45, 46], the minimum probability of symptom improvement in benign prostatic hyperplasia [18], the minimum required improvement in survival related to mechanical ventilation [47], the minimum acceptable probability of healthy survival required to justify resuscitation [48], the minimum reduction

in migraine headache frequency [49], and the minimum increase in the probability of pregnancy resulting from fertility treatment [29, 50, 51].

Fourteen studies used the TT to estimate the maximum acceptable probability of one or more risks associated with treatment. Half of these studies are related to cancer treatments [21, 24, 39, 52–55]. Other studies have used the TT to estimate MAR thresholds in treatments for cardiovascular disease [27, 28, 56], osteoarthritis [13], obstetrics [57], and fertility [50, 51].

3.1.2 Other Probabilistic Thresholds

Although the most common thresholds estimated using the TT are measures of risk tolerance—MAR and MAB—this technique also has been used to estimate probabilistic thresholds that may not be easily categorized using these terms. For example, Gupta et al. [55] estimated the minimum probability of treatment-related infertility that would induce parents, cancer survivors, and healthcare providers to agree that prepubertal patients should undergo testicular biopsy to potentially preserve future fertility, even though the technology to do this has yet to be developed. These researchers also estimated the influence that the probability that such a technology would be developed in the near future would have on the decision to undertake the procedure. In a study of elderly patients, Dales et al. [26] estimated the minimum probability of successful weaning that would be required to accept mechanical ventilation.

3.1.3 Non-Probabilistic Thresholds

Although the TT is well-suited to estimating probabilistic thresholds, it can also be used to estimate any threshold that can be expressed numerically. A few studies estimated a non-probabilistic numeric threshold, such as the maximum acceptable number of clinic visits [54] or the minimum improvement in quality of life [43]. However, the most common alternative to probability values in the TT is time. The two applications of time thresholds in the literature are survival time and wait time.

One group of studies used the same general technique to estimate the minimum increase in life expectancy that would be required for decision makers to accept the burdens of cancer treatments [22, 32, 34–38]. Additional studies estimated survival-time thresholds in end-of-life care [58] and pediatric oncology [43]. Four studies used the TT to estimate wait-time thresholds related to delaying chemotherapy to get the results of pharmacogenomic testing [42], delaying surgery in exchange for decreasing the risk of postoperative mortality [14, 17], and delaying treatment in order to have treatment closer to home [20].

Table 2 Summary of threshold technique studies by therapeutic area and threshold measure

Study, year	Thresholds estimated in the study						
	Probability thresholds			Time thresholds			Other thresholds
	Minimum acceptable benefit	Maximum acceptable risk	Other	Survival time	Wait time	Other	
Oncology							
Breast cancer							
Alvarado et al., 2014 [52]		X					
Corica et al., 2014 [53]		X					
Duric et al., 2005 [35]	X			X			
Duric et al., 2005 [36]	X			X			
Palda et al., 1997 [20]	X				X		
Simes and Coates, 2001 [37]	X			X			
Lung cancer							
Blinman et al., 2015 [32]	X			X			
Blinman et al., 2011 [22]	X			X			
Brundage et al., 1997 [33]	X						
Brundage et al., 2001 [34]	X			X			
Colon and rectal cancer							
Blinman et al., 2010 [38]	X			X			
Bossema et al., 2008 [24]		X					
Couture et al., 2005 [25]	X						
Kennedy et al., 2011 [41]	X						
Llewellyn-Thomas et al., 1991 [15]							
Pediatric oncology							
Gupta et al., 2016 [55]		X	X				
Sung et al., 2004 [54]		X					X
Tomlinson et al., 2011 [43]	X			X			X
Oncology—other							
Cuffe et al., 2014 [42]	X					X	
Finlayson et al., 1999 [21]		X					
Kunneman et al., 2014 [23]	X						
Molinari et al., 2014 [39]	X	X					
Wilke et al., 2010 [40]	X						
Cardiovascular							
Alonso-Coello et al., 2015 [56]		X					
Bates et al., 2016 [45]	X						
Devereaux et al., 2001 [27]	X	X					
Llewellyn-Thomas et al., 2002 [16]	X						
Llewellyn-Thomas et al., 1999 [17]					X		
McAlister et al., 2000 [46]	X						
Stafinski et al., 2015 [28]		X					
Reproductive health							
Bayram et al., 2005 [29]	X						
Kok et al., 2008 [57]		X					
Steures et al., 2005 [50]	X	X					
van Weert et al., 2007 [51]	X	X					
End-of-life and emergent care							
Crump and Llewellyn-Thomas, 2013 [59]				X			
Lloyd et al., 2004 [47]	X						

Table 2 (continued)

Study, year	Thresholds estimated in the study						
	Probability thresholds			Time thresholds			Other thresholds
	Minimum acceptable benefit	Maximum acceptable risk	Other	Survival time	Wait time	Other	
Percy and Llewellyn-Thomas, 1995 [48]	X						
Orthopedics							
Kopec et al., 2007 [13]		X					
Llewellyn-Thomas et al., 1998 [14]					X		
Respiratory							
Dales et al., 1999 [26]			X				
Gastrointestinal							
Kennedy et al., 2008 [44]	X						
Men's health							
Llewellyn-Thomas et al., 1996 [18]	X						
Migraine							
Turner et al., 2016 [49]	X						

3.2 Other Features of Threshold Technique Studies

In addition to the type of threshold estimated in each study, other features of the empirical applications of the TT are worth noting. The TT can be used to elicit multiple thresholds in a single study and understand heterogeneity in preferences by examining the distribution of threshold values and the relationship between this distribution and the characteristics of the respondents in the sample. These features of TT studies are summarized in Table 3 and described in Sects. 3.2.1 and 3.2.2.

3.2.1 Estimating Multiple Thresholds in a Single Study

Each TT exercise is designed to estimate the threshold value of a single key attribute that exactly offsets the difference in utility between the target and reference options. However, that does not mean that the trade-off exercise cannot be repeated for multiple key attributes. As shown in Table 3, approximately two-thirds of the studies estimated multiple benefit or burden thresholds. Some estimated multiple thresholds for the same key attributes by repeating the exercise for different magnitudes of difference between the reference and target options. For example, Kopec et al. [13] estimated MAR thresholds for two different levels of decrease in osteoarthritis pain. Some estimated thresholds for multiple benefits [28, 43, 47], and some estimated thresholds for multiple risks or burdens [13, 50, 54, 57]. Some estimated thresholds for multiple treatment options by repeating the exercise with multiple target options representing different approaches to treatment [27, 33, 44] or by evaluating multiple different treatment decisions [18, 59].

3.2.2 Preference Heterogeneity

Because the TT yields for each respondent a single threshold value or interval for each key attribute for each trade-off exercise, the results can be used to describe the distribution of preferences and, potentially, to explain preference heterogeneity. The distribution of threshold values or intervals for each key attribute for each trade-off exercise can be examined directly by plotting the results or summarizing the results in a table. In addition, researchers can evaluate the relationship between respondent characteristics and thresholds in multiple ways. These include correlation or univariate analysis within a sample, split- or stratified-sample analysis, or multivariate or multinomial regression analysis.

3.3 Rationality Assessments

Many TT studies have assessed the performance of the technique against some basic principles of rationality based on economic theory, psychometrics, and psychology. These include tests of monotonicity, anchoring and shift-framing effects, preference non-linearity, test–retest reliability, and preference stability. The empirical studies that employ each of these tests are shown in Table 4.

3.3.1 Monotonicity

Monotonicity implies that a lower level of benefit cannot be preferred to a higher level of benefit or that a higher level of risk or burden cannot be preferred to a lower level of risk. This concept can be operationalized by testing whether a larger benefit is only offset by a level of risk that is at least

Table 3 Summary of features of threshold technique studies

Study, year	Multiple thresholds	Heterogeneity				
		Distributions	Univariate analysis	Split sample	Multiple samples	Multivariate analysis
Alonso-Coello et al., 2015 [56]		X	X		X	
Alvarado et al., 2014 [52]		X	X			
Bates et al., 2016 [45]				X		X
Bayram et al., 2005 [29]	X			X		
Blinman et al., 2010 [38]	X	X	X	X		
Blinman et al., 2015 [32]	X	X	X		X	
Blinman et al., 2011 [22]	X	X	X	X		X
Bossema et al., 2008 [24]	X	X	X	X		
Brundage et al., 1997 [33]	X	X	X	X	X	
Brundage et al., 2001 [34]	X	X	X			
Corica et al., 2014 [53]		X	X	X		X
Couture et al., 2005 [25]		X		X		
Crump and Llewellyn-Thomas, 2013 [59]	X	X				
Cuffe et al., 2014 [42]				X		
Dales et al., 1999 [26]		X			X	
Devereaux et al., 2001 [27]	X	X	X		X	
Duric et al., 2005 [35]	X	X	X			X
Duric et al., 2005 [36]	X	X	X			X
Finlayson et al., 1999 [21]		X	X			X
Gupta et al., 2016 [55]	X				X	X
Kennedy et al., 2011 [41]	X	X	X			X
Kennedy et al., 2008 [44]	X	X				
Kok et al., 2008 [57]		X		X		
Kopec et al., 2007 [13]	X					X
Kunneman et al., 2014 [23]		X		X	X	
Llewellyn-Thomas et al., 1998 [14]			X	X		X
Llewellyn-Thomas et al., 1991 [15]				X		
Llewellyn-Thomas et al., 2002 [16]	X			X		
Llewellyn-Thomas et al., 1999 [17]						X
Llewellyn-Thomas et al., 1996 [18]	X			X		X
Lloyd et al., 2004 [47]	X			X		X
McAlister et al., 2000 [46]	X	X			X	X
Molinari et al., 2014 [39]	X	X	X			X
Palda et al., 1997 [20]	X	X		X		X
Percy and Llewellyn-Thomas, 1995 [48]	X		X	X		
Simes and Coates, 2001 [37]	X	X		X		X
Stafinski et al., 2015 [28]	X	X				
Steures et al., 2005 [50]	X	X	X	X		
Sung et al., 2004 [54]			X		X	
Tomlinson et al., 2011 [43]	X				X	X
Turner et al., 2016 [49]	X	X				
van Weert et al., 2007 [51]	X	X	X	X		
Wilke et al., 2010 [40]	X		X			X

as large as that required to offset a lower level of benefit or whether the benefit required to offset a higher level of risk is at least as large at the benefit required to offset a lower level of risk. Two TT studies include tests of monotonicity. Kopec et al. [13] elicited risk threshold values for two different levels of benefit and found that patients would accept higher levels of risk, on average, in exchange for greater pain relief. Lloyd et al. [47] elicited minimum survival benefit and minimum required quality of life under two different mechanical ventilation scenarios—acute and chronic—in which the chronic scenario was defined as unambiguously worse than the acute scenario. They found that the proportion accepting mechanical ventilation at each survival probability or probability of improved quality of life was lower for the chronic ventilation scenario than for the acute ventilation scenario, indicating that greater improvements in the probability of survival or quality of life were required to offset the additional burden of chronic ventilation.

3.3.2 Anchoring and Shift-Framing Effects

When conducting a TT study, the researcher must specify the starting levels of the key attribute. There are two potential impacts of this design choice on the resulting threshold estimates. The first is anchoring. Anchoring occurs when the threshold value for the key attribute is influenced by the numeric starting level of the attribute in the initial choice question because the respondent focuses on the initial level when answering subsequent questions. The second, shift-framing, occurs when the threshold value for the key attribute is influenced by the difference in the level of this attribute between the reference and target options in the initial question. Larger (smaller) initial differences in the starting values of the key attribute between the reference and target options in the initial choice question may result in larger (smaller) values of the threshold value of the key attribute. Anchoring and shift-framing are related concepts because both are artifacts of the level of the key attribute presented in the initial choice question in a TT exercise. However, to the extent that the starting point represents reality in that it is based on data or on a value that would be expected even if data do not exist, then the starting point reflects the true decision context and will reflect bias inherent in that decision context. Therefore, the possibility of anchoring and shift-framing effects may be related to the extent to which the choice scenario in the TT exercise is hypothetical.

Alvarado et al. [52] used different starting points for the risk of local recurrence associated with intraoperative radiotherapy and found that the effect of the starting point on the resulting threshold estimates was not statistically significant. Kopec et al. [13] directly addressed the question of anchoring in their study of osteoarthritis patients' willingness to accept risk to achieve treatment-related pain relief. The authors estimated risk-tolerance thresholds for five potential treatment-related risks. Half the sample were presented with hypothetical treatments in which all risks were set to zero in the initial treatment question and the other half were presented with hypothetical treatments in which all risks were initially set to non-zero levels reflecting the levels of these risks that could be expected with existing osteoarthritis treatments. When comparing the results between these two groups, Kopec et al. [13] found some indication that starting with higher levels of risks resulted in higher incremental MAR thresholds; however, the differences between the two arms were not statistically significant.

Duric et al. [36] and Simes and Coates [37] asked respondents to indicate the minimum increases in survival time and in survival probability that would be required for adjuvant therapy to be considered worthwhile for patients with breast cancer. Both studies tested the effect of starting points on threshold estimates by assigning half the sample to start with a larger difference in the levels of the key attribute and half the sample to start with a smaller difference in the levels of the key attribute in the initial choice question. Both studies found no statistically significant differences in the threshold estimates based on the starting point of the benefit in the target option.

Finally, Percy and Llewellyn-Thomas [48] directly addressed the issue of shift-framing. In their study, respondents were presented with a choice between having a "do not resuscitate" (DNR) order resulting in a certainty of immediate death or an order to be resuscitated with a 100% probability of living, split between the probability of surviving without brain damage and the probability of brain death. In one arm of their study (positive-to-negative), these authors first presented respondents with a choice between the DNR order and a 100% probability of surviving without brain death. The probability of surviving without brain death was then decreased and the probability of brain death was increased until the person chose the DNR order. In the second arm of the study (negative-to-positive), respondents were first presented with the choice between the DNR order and a 100% probability of brain death. In subsequent questions, the risk of brain death was reduced and the risk of survival without brain death was increased until the respondents who initially chose the DNR order no longer chose that option. The authors found that respondents in the positive-to-negative arm were willing to accept the DNR order at statistically significantly higher levels of resuscitation survival without brain death than were the respondents in the negative-to-positive arm of the study, thus providing evidence of shift-frame bias in this hypothetical scenario.

Table 4 Summary of rationality assessments in threshold technique studies

Study	Monotonicity	Anchoring and shift-framing effects	Preference non-linearity	Test–retest reliability and preference stability
Alonso-Coello et al., 2015 [56]				
Alvarado et al., 2014 [52]		X		
Bates et al., 2016 [45]				
Bayram et al., 2005 [29]				
Blinman et al., 2010 [38]			X	
Blinman et al., 2015 [32]			X	
Blinman et al., 2011 [22]			X	
Bossema et al., 2008 [24]				
Brundage et al., 1997 [33]				X
Brundage et al., 2001 [34]				
Corica et al., 2014 [53]				
Couture et al., 2005 [25]				
Crump and Llewellyn-Thomas, 2013 [59]				
Cuffe et al., 2014 [42]				
Dales et al., 1999 [26]				
Devereaux et al., 2001 [27]				
Duric et al., 2005 [35]		X	X	
Duric et al., 2005 [36]			X	X
Finlayson et al., 1999 [21]				
Gupta et al., 2016 [55]				
Kennedy et al., 2011 [41]				
Kennedy et al., 2008 [44]				
Kok et al., 2008 [57]				X
Kopec et al., 2007 [13]	X	X		
Kunneman et al., 2014 [23]				
Llewellyn-Thomas et al., 1998 [14]				
Llewellyn-Thomas et al., 1991 [15]				
Llewellyn-Thomas et al., 2002 [16]				
Llewellyn-Thomas et al., 1999 [17]				
Llewellyn-Thomas et al., 1996 [18]				
Lloyd et al., 2004 [47]	X			
McAlister et al., 2000 [46]				
Molinari et al., 2014 [39]				
Palda et al., 1997 [20]				
Percy and Llewellyn-Thomas, 1995 [48]		X		X
Simes and Coates, 2001 [37]		X	X	X
Stafinski et al., 2015 [28]				
Steures et al., 2005 [50]				
Sung et al., 2004 [54]				
Tomlinson et al., 2011 [43]				
Turner et al., 2016 [49]				
van Weert et al., 2007 [51]				
Wilke et al., 2010 [40]				

3.3.3 Preference Non-linearity

Preference non-linearity occurs when the incremental change in the value of the key attribute that exactly offsets the other differences between the reference and target option changes when the level of the key attribute in the reference option changes. If the incremental change remains constant, then preferences are linear over the range of potential values of the key attribute. For example, Simes and Coates [37] included two different baseline levels of survival time and survival probability in their study of the increases in expected survival that patients would require to accept the burdens associated with adjuvant chemotherapy for early breast cancer. In separate TT exercises, these authors elicited the minimum additional survival time required for adjuvant chemotherapy when the life expectancy in the reference option (no adjuvant chemotherapy) was 5 years and when the reference life expectancy was 15 years. These authors also elicited the minimum additional probability of 5-year survival required for adjuvant chemotherapy when the probability of 5-year survival in the reference option was 65% or 85%. These authors describe notable differences in incremental increases in required survival time based on the baseline levels, indicating that the incremental required increase in survival decreased as the expected survival in the reference option increased. On the basis of this result, the authors concluded that the women in their study were "discounting benefits of treatment that were appreciably delayed" (Simes and Coates [37], p. 148). However, they found that the proportion of patients accepting adjuvant chemotherapy appeared to be independent of the baseline 5-year survival probability. That is, these authors found that preferences were non-linear in survival time, but likely linear in 5-year survival probability.

A series of additional studies [22, 32, 35, 36, 38] used the same procedure for estimating minimum required increases in survival as that used by Simes and Coates [37]. In these studies, the incremental thresholds appear to differ based on the initial level in the reference option. In their study of physicians' preferences for adjuvant chemotherapy in non-small-cell lung cancer, Blinman et al. [22] found that statistically significantly smaller incremental increases in the probability of 5-year survival were required when the baseline survival probability was 65% than when the baseline probability was 50%. However, there were no statistically significant differences in the minimum required increase in survival time between the two baseline life expectancies. In contrast, Duric et al. [36] found no statistically significant differences in minimum required increases in survival based on different baseline levels of life expectancy or 5-year survival probability in their study of women's preferences for adjuvant endocrine therapy in early breast cancer.

3.3.4 Test–Retest Reliability and Preference Stability

Five studies administered TT exercises to the same subjects at different points in time to evaluate whether preferences were stable and reliable over time. Three studies repeated the trade-off exercise within 1–6 weeks [33, 44, 48]. Kennedy et al. [44] found fair to good test–retest reliability and noted that the mean difference in the threshold values between the first and second administration (2–4 weeks apart) was less than 5% of the highest possible change that could have been observed between the first and second administration of the exercise. Percy and Llewellyn-Thomas [48] found a statistically significant association between the results of the exercises administered 1 week apart. In contrast, Brundage et al. [33] noted that 85% of the threshold values in their study were lower in the retest interview (approximately 6 weeks later) than in the original interview. Two studies had longer intervals between the first and second times the trade-off exercises were administered. Duric et al. [35] found that agreement between the original test and the retest (17 months later, on average) was moderate to good and that differences in the results between the two times were not statistically significant. Simes and Coates [37] found no significant change in the minimum required increase in survival time; however, they did find that the minimum required increase in the 5-year survival probability was larger in the second interviews (3–6 months later).

4 Discussion

As the demand for greater patient input into medical and regulatory decision-making grows, PPI is becoming an increasingly important source of evidence. Many tools have been used to elicit PPI in numerous applications. The TT is one of many tools for eliciting patients' benefit–risk preferences. Although not as common in the patient preference literature as discrete-choice experiments (DCEs) [60–62], it has been used in numerous applications in multiple therapeutic areas over the past 25 years and has been used recently to provide evidence of patients' risk tolerance to support a change in labeling for home hemodialysis [7]. This is the first summary of the literature on empirical applications of the TT.

The TT asks respondents to make choices between two alternatives and systematically varies the features of the alternatives until the point at which a respondent is indifferent between alternatives is determined. Therefore, it is like many other preference elicitation techniques commonly used in health and medical decision-making. Early applications of the TT referred to the method as a modified standard gamble [21] because it offered decision makers a choice between two possible states of the world with uncertain outcomes and varied the probability of the outcomes until a person

was indifferent between the two states. Many early empirical applications of the TT also referred to the method as a time trade-off [22, 24] because respondents were offered a choice between a less burdensome alternative with shorter life expectancy and a more burdensome alternative with longer life expectancy. The TT is also similar to contingent valuation [63] because it can be used to estimate the marginal monetary value of incremental changes in treatment or health service options or to estimate the welfare gain of an alternative approach to treatment. Finally, the TT is not dissimilar from some forms of swing weighting in which two changes (i.e., swings) in outcomes are compared and the size of one swing is varied systematically until the respondent is indifferent between the two swings [64].

Empirical studies of the TT have demonstrated that it can be used to elicit multiple types of thresholds relevant to benefit–risk decision-making, including estimates of MAR and MAB [8]. In addition, multiple trade-offs and threshold values can be quantified in a single study. Because the TT yields a unique value or interval for each respondent for each threshold, it can also be used to characterize, quantify, and potentially explain preference heterogeneity, an important consideration in regulatory decision-making [3]. Finally, the TT can offer reliable estimates of PPI. It has been shown to have face validity in that MAR estimates are higher (lower) for higher (lower) levels of benefit and MAB estimates are higher (lower) for higher (lower) levels of risk. In addition, it can capture non-linearities in preferences over a range of benefits and risks.

Some evidence has shown that the TT may be subject to anchoring and shift-framing effects. The evidence regarding the sensitivity of threshold estimates to anchoring and shift-framing is mixed and may be directly related to the extent to which the decision in the TT exercise is hypothetical. However, as noted earlier, if the decision being addressed in the study is subject to anchoring and shift-framing because the actual choice decision has a natural starting point or decision frame, then any bias in the TT study may simply reflect the reality of the decision. The evidence regarding preference stability also appears to be mixed, and it appears that preference stability decreases as the interval between the first and second assessment of preferences increases. This result is not entirely surprising, as many events can occur between assessments that could impact how patients perceive the decision problem. However, the TT method is also likely subject to a number of potential limitations that are common to all stated-preference methods in which patients are asked to make hypothetical treatment decisions regarding risks and benefits. These include administering the exercise to respondents with potentially low levels of numeracy or health literacy, bias introduced by the mode (in-person interview or pencil-and-paper survey, or computer-administered

survey instrument) by which data are collected, sample selection bias due to the method of respondent recruitment, and non-trading or non-attendance responses and protest responses on the part of respondents. We are aware of no evidence that the TT is more or less susceptible to these potential sources of bias or preference instability than other hypothetical stated-preference techniques.

The TT differs from DCEs in that the level of only the key attribute is varied. In contrast, in a DCE, all attribute levels are varied simultaneously between options and across choice questions according to an experimental design. The DCE has the advantage of estimating the relationships among all attributes simultaneously; however, the DCE yields results for a sample rather than a single result or threshold for each respondent. As a result, the ability of the DCE to relate choice to the individual characteristics of the respondents is limited. In addition, the DCE is subject to scale heterogeneity [65], which makes pooling data from multiple sources or across time more difficult. Additional research is needed to determine the extent to which the TT and DCE yield similar results when applied to the same research question.

The TT may be a useful approach to eliciting patients' stated preferences to inform regulatory and other decision-making. Early studies employing this technique were designed to assess what benefits would be necessary to make more aggressive treatment worthwhile [22, 32, 35, 36, 38, 41, 44]. The TT may be most useful when the primary research question is less concerned with trade-offs among multiple attributes and more concerned with determining the minimal clinically important difference of a treatment for which the other characteristics are understood. In addition, anchoring and shift-framing effects may be a result of the decision context rather than an artifact of the methods when the decision context represents an actual or potential treatment decision.

5 Conclusion

The TT has been employed in a number of therapeutic areas and has been used to quantify treatment preferences of patients, caregivers, and healthcare professionals. The TT is similar to many other stated-preference methods (e.g., standard gamble, time trade-off, swing weighting, and contingent valuation) in many respects and has some characteristics that differ from other experimentally designed stated-preference methods (e.g., DCEs and best–worst scaling). As with all stated-preference techniques, the ability of the TT to predict treatment choice and the ability to inform an individual patient about the treatment that is most appropriate for her or him is not well-known. This review provides only a brief overview of the features and applications of the

TT. We recommend that additional empirical research focus on three primary areas: performance of the TT compared with other stated-preference methods when applied to the same research question, the extent to which results of a TT study are able to predict patient choice, and the ability of the TT to inform individual treatment decisions at the point of healthcare delivery.

Compliance with Ethical Standards

References

1. Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient preference evidence into regulatory decision making. Surg Endosc. 2015;29(10):2984–93.
2. Al-Faruque F. FDA weighs patients' risk tolerance in approving obesity device. The Pink Sheet; 2015 Jan 20. https://pink.pharmaintelligence.informa.com/PS077790/FDA-Weighs-Patients-Risk-Tolerance-in-Approving-Obesity-Device. Accessed 29 Jun 2019.
3. US Food and Drug Administration. Patient preference information—voluntary submission, review in premarket approval applications, humanitarian device exemption applications, and de novo requests, and inclusion in decision summaries and device labeling: guidance for industry, Food and Drug Administration staff, and other stakeholders. 2016. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM446680.pdf. Accessed 27 Dec 2018.
4. US Food and Drug Administration. Plan for issuance of patient-focused drug development guidance under 21st Century Cures Act Title III Section 3002. 2017. https://www.fda.gov/media/105979/download. Accessed 29 Jun 2019.
5. US Food and Drug Administration. Benefit-risk assessment throughout the drug lifecycle: FDA discussion document. 2019. https://healthpolicy.duke.edu/sites/default/files/atoms/files/discussion_guide_b-r_assessment_may16_0.pdf. Accessed 29 Jun 2019.
6. Postmus D, Mavris M, Hillege HL, et al. Incorporating patient preferences into drug development and regulatory decision making: results from a quantitative pilot study with cancer patients, carers, and regulators. Clin Pharmacol Ther. 2016;99(5):548–54.
7. Medical Device Innovation Consortium. Medical Device Innovation Consortium (MDIC) patient centered benefit-risk project report: a framework for incorporating information on patient preferences regarding benefit and risk in regulatory assessments of new medical technology. 2015. https://mdic.org/wp-content/uploads/2018/05/MDIC_PCBR_Framework_Web.pdf. Accessed 27 Dec 2018.
8. Hauber AB, Fairchild AO, Johnson FR. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy. 2013;11(4):319–29.
9. Soekhai V, Whichello C, Levitan B, et al. Methods for exploring and eliciting patient preference in the medical product lifecycle: a literature review. Drug Discov Today. 2019;24(7):1324–31.
10. Medical Device Innovation Consortium. A case study in how patient preference information contributed to regulatory decisions for medical devices. 2017. https://mdic.org/event/a-case-study-in-how-patient-preference-information-contributes-to-regulatory-decisions-for-medical-devices/. Accessed 27 Dec 2018.
11. Llewellyn-Thomas HA. Threshold technique. In: Kattan MW, editor. Encyclopedia of medical decision making. Thousand Oaks: Sage; 2009. p. 1134–7.
12. Brundage MD, Davidson JR, Mackillop WJ, Feldman-Stewart D, Groome P. Using a treatment-tradeoff method to elicit preferences for the treatment of locally advanced non-small-cell lung cancer. Med Decis Making. 1998;18(3):256–67.
13. Kopec JA, Richardson CG, Llewellyn-Thomas H, Klinkhoff A, Carswell A, Chalmers A. Probabilistic threshold technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied. J Clin Epidemiol. 2007;60(9):929–38.
14. Llewellyn-Thomas HA, Arshinoff R, Bell M, Williams JI, Naylor CD. In the queue for total joint replacement: patients' perspectives on waiting times. Ontario Hip and Knee Replacement Project Team. J Eval Clin Pract. 1998;4(1):63–74.
15. Llewellyn-Thomas H, McGreal MJ, Thiel EC, Fine S, Erlichman C. Patients' willingness to enter clinical trials: measuring the association with perceived benefit and preference for decision participation. Soc Sci Med. 1991;32:35–42.
16. Llewellyn-Thomas HA, Paterson JM, Carter JA, et al. Primary prevention drug therapy: can it meet patients' requirements for reduced risk? Med Decis Making. 2002;22:326–39.
17. Llewellyn-Thomas H, Thiel E, Paterson M, Naylor D. In the queue for coronary artery bypass grafting: patients' perceptions of risk and 'maximal acceptable waiting time'. J Health Serv Res Policy. 1999;4(2):65–72.
18. Llewellyn-Thomas HA, Williams JI, Levy L, Naylor CD. Using a trade-off technique to assess patients' treatment preferences for benign prostatic hyperplasia. Med Decis Making. 1996;16(3):262–82.
19. Naylor CD, Llewellyn-Thomas HA. Can there be a more patient-centred approach to determining clinically important effect sizes for randomized treatment trials? J Clin Epidemiol. 1994;47(7):787–95.
20. Palda VA, Llewellyn-Thomas HA, Mackenzie RG, Pritchard KI, Naylor CD. Breast cancer patients' attitudes about rationing post-lumpectomy radiation therapy: applicability of trade-off methods to policy-making. J Clin Oncol. 1997;15(10):3192–200.
21. Finlayson SR, Birkmeyer JD, Tosteson AN, Nease RF Jr. Patient preferences for location of care: implications for regionalization. Med Care. 1999;37(2):204–9.
22. Blinman P, McLachlan SA, Nowak AK, et al. Lung cancer clinicians' preferences for adjuvant chemotherapy in non-small-cell lung cancer: what makes it worthwhile? Lung Cancer. 2011;72(2):213–8. https://doi.org/10.1016/j.lungcan.2010.08.007.

23. Kunneman M, Pieterse AH, Stiggelbout AM, et al. Treatment preferences and involvement in treatment decision making of patients with endometrial cancer and clinicians. Br J Cancer. 2014;111:674–9.

24. Bossema E, Stiggelbout A, Baas-Thijssen M, van de Velde C, Marijnen C. Patients' preferences for low rectal cancer surgery. Eur J Surg Oncol. 2008;34(1):42–8.

25. Couture J, Chan R, Bouharaoui F. Patient's preferences for adjuvant postoperative chemoradiation therapy in rectal cancer. Dis Colon Rectum. 2005;48(11):2055–60.

26. Dales RE, O'Connor A, Hebert P, et al. Intubation and mechanical ventilation for COPD: development of an instrument to elicit patient preferences. Chest. 1999;116(3):792–800.

27. Devereaux PJ, Anderson DR, Gardner MJ, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ. 2001;323(7323):1218–22.

28. Stafinski T, Menon D, Nardelli A, et al. Incorporating patient preferences into clinical trial design: results of the Opinions of Patients on Treatment Implications of New Studies (OPTIONS) project. Am Heart J. 2015;169:122–31.

29. Bayram N, van Wely M, van der Veen F, Bossuyt PM, Nieuwkerk P. Treatment preferences and trade-offs for ovulation induction in clomiphene citrate-resistant patients with polycystic ovary syndrome. Fertil Steril. 2005;84(2):420–5.

30. Tervonen T, Naci H, van Valkenhoef G, et al. Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention. Med Decis Making. 2015;35(7):859–71.

31. Cameron AC, Trivedi PK. Microeconometrics using Stata. Rev. ed. College Station: Stata Press; 2010.

32. Blinman P, Hughes B, Crombie C, et al; Australasian Lung Cancer Trials Group (ALTG). Patients' and doctors' preferences for adjuvant chemotherapy in resected non-small-cell lung cancer: what makes it worthwhile? Eur J Cancer. 2015;51(12):1529–37. https://doi.org/10.1016/j.ejca.2015.05.022.

33. Brundage MD, Davidson JR, Mackillop WJ. Trading treatment toxicity for survival in locally advanced non-small cell lung cancer. J Clin Oncol. 1997;15(1):330–40.

34. Brundage MD, Feldman-Stewart D, Cosby R, et al. Cancer patients' attitudes toward treatment options for advanced non-small cell lung cancer: implications for patient education and decision support. Patient Educ Couns. 2001;45(2):149–57.

35. Duric VM, Stockler MR, Heritier S, et al. Patients' preferences for adjuvant chemotherapy in early breast cancer: what makes AC and CMF worthwhile now? Ann Oncol. 2005;16(11):1786–94.

36. Duric VM, Fallowfield LJ, Saunders C, Houghton J, Coates AS, Stockler MR. Patients' preferences for adjuvant endocrine therapy in early breast cancer: what makes it worthwhile? Br J Cancer. 2005;93(12):1319–23.

37. Simes RJ, Coates AS. Patient preferences for adjuvant chemotherapy of early breast cancer: how much benefit is needed? J Natl Cancer Inst Monogr. 2001;30:146–52.

38. Blinman P, Duric V, Nowak AK, et al. Adjuvant chemotherapy for early colon cancer: what survival benefits make it worthwhile? Eur J Cancer. 2010;46(10):1800–7. https://doi.org/10.1016/j.ejca.2009.12.032.

39. Molinari M, De Coutere S, Krahn M, Helton S, Urbach DR. Patients' preferences and trade-offs for the treatment of early stage hepatocellular carcinoma. J Surg Res. 2014;189:57–67.

40. Wilke DR, Krahn M, Tomlinson G, Bezjak A, Rutledge R, Warde P. Sex or survival: short-term versus long-term androgen deprivation in patients with locally advanced prostate cancer treated with radiotherapy. Cancer. 2010;116:1909–17.

41. Kennedy ED, Schmocker S, Victor C, et al. Do patients consider preoperative chemoradiation for primary rectal cancer worthwhile? Cancer. 2011;117(13):2853–62.

42. Cuffe S, Hon H, Qiu X, et al. Cancer patients' acceptance, understanding, and willingness-to-pay for pharmacogenomic testing. Pharmacogenet Genomics. 2014;24(7):348–55.

43. Tomlinson D, Bartels U, Gammon J, et al. Chemotherapy versus supportive care alone in pediatric palliative care for cancer: comparing the preferences of parents and health care professionals. CMAJ. 2011;183(17):E1252–8.

44. Kennedy ED, To T, Steinhart AH, Detsky A, Llewellyn-Thomas HA, McLeod RS. Do patients consider postoperative maintenance therapy for Crohn's disease worthwhile? Inflamm Bowel Dis. 2008;14(2):224–35.

45. Bates SM, Alonso-Coello P, Tikkinen KAO, et al. Women's values and preferences and health state valuations for thromboprophylaxis during pregnancy: a cross-sectional interview study. Thromb Res. 2016;140:22–9.

46. McAlister FA, O'Connor AM, Wells G, Grover SA, Laupacis A. When should hypertension be treated? The different perspectives of Canadian family physicians and patients. CMAJ. 2000;163(4):403–8.

47. Lloyd CB, Nietert PJ, Silvestri GA. Intensive care decision making in the seriously ill and elderly. Crit Care Med. 2004;32(3):649–54.

48. Percy ME, Llewellyn-Thomas H. Assessing preferences about the DNR order: does it depend on how you ask? Med Decis Making. 1995;15(3):209–16.

49. Turner DP, Golding AN, Houle TT. Using a graphical risk tool to examine willingness-to-take migraine prophylactic medications. Pain. 2016;157(10):2226–34.

50. Steures P, Berkhout JC, Hompes PG, et al. Patients' preferences in deciding between intrauterine insemination and expectant management. Hum Reprod. 2005;20(3):752–5.

51. van Weert JM, van den Broek J, van der Steeg JW, et al. Patients' preferences for intrauterine insemination or in vitro fertilization. Reprod Biomed Online. 2007;15(4):422–7.

52. Alvarado MD, Conolly J, Park C, et al. Patient preferences regarding intraoperative versus external beam radiotherapy following breast-conserving surgery. Breast Cancer Res Treat. 2014;143(1):135–40. https://doi.org/10.1007/s10549-013-2782-9.

53. Corica T, Joseph D, Saunders C, Bulsara M, Nowak AK. Intraoperative radiotherapy for early breast cancer: do health professionals choose convenience or risk? Radiat Oncol. 2014;25(9):33. https://doi.org/10.1186/1748-717X-9-33.

54. Sung L, Feldman BM, Schwamborn G, et al. Inpatient versus outpatient management of low-risk pediatric febrile neutropenia: measuring parents' and healthcare professionals' preferences. J Clin Oncol. 2004;22(19):3922–9.

55. Gupta AA, Donen RM, Sung L, et al. Testicular biopsy for fertility preservation in prepubertal boys with cancer: identifying preferences for procedure and reactions to disclosure practices. J Urol. 2016;196(1):219–24. https://doi.org/10.1016/j.juro.2016.02.2967.

56. Alonso-Coello P, Montori VM, Díaz MG, et al. Values and preferences for oral antithrombotic therapy in patients with atrial fibrillation: physician and patient perspectives. Health Expect. 2015;18(6):2318–27. https://doi.org/10.1111/hex.12201.

57. Kok M, Gravendeel L, Opmeer BC, van der Post JA, Mol BW. Expectant parents' preferences for mode of delivery and trade-offs of outcomes for breech presentation. Patient Educ Couns. 2008;72(2):305–10.

58. Crump T, Llewellyn-Thomas HA. Assessing Medicare beneficiaries' strength-of-preference scores for health care options: how engaging does the elicitation technique need to be? Health Expect. 2011;14(Suppl 1):33–45. https://doi.org/10.1111/j.1369-7625.2010.00632.x.

59. Crump RT, Llewellyn-Thomas H. Characterizing the public's preferential attitudes toward end-of-life care options: a role for the threshold technique? Health Serv Res. 2013;48(6 Pt 1):2101–24. https://doi.org/10.1111/1475-6773.12049.

60. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics. 2019;37(2):201–26. https://doi.org/10.1007/s40273-018-0734-2.

61. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883–902.

62. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.

63. Klose T. The contingent valuation method in health care. Health Policy. 1999;47:97–123.

64. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value trade-offs. Cambridge: Cambridge University Press; 1993.

65. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. J Mark Res. 1993;30:305–14.