Understanding the global measurement of willingness to pay in health

Jean A. McDougall\textsuperscript{a,b}, Wesley E. Furnback\textsuperscript{a}, Bruce C. M. Wang\textsuperscript{a} and Jörg Mahlich\textsuperscript{c,d}

\textsuperscript{a}Health Economics and Outcomes Research, Elysia Group, LLC, New York, NY, USA; \textsuperscript{b}Division of Epidemiology, Biostatistics, and Preventive Medicine, University of New Mexico, Albuquerque, NM, USA; \textsuperscript{c}Health Economics, Janssen Pharmaceutical KK, Tokyo, Japan; \textsuperscript{d}Düsseldorf Institute for Competition Economics (DICE), University of Düsseldorf, Düsseldorf, Germany

\textbf{ABSTRACT}

\textbf{Objective}: To understand the different methodologies used to elicit willingness to pay for health and the value of a statistical life year through surveys.

\textbf{Methodology}: A systematic review of the literature was undertaken to identify studies using surveys to estimate either willingness to pay for health or the value of a statistical life year. Each study was reviewed and the study setting, sample size, sample description, survey administration (online or face to face), survey methodology, and results were extracted. The results of the studies were then compared to any published national guidelines of cost-effectiveness thresholds to determine their accuracy.

\textbf{Results}: Eighteen studies were included in the review with 15 classified as willingness to pay and 3 value of a statistical life. The included studies covered Asia (n = 6), Europe (n = 4), the Middle East (n = 1), and North America (n = 5), with one study taking a global perspective. There were substantial differences in both the methodologies and the estimates of both willingness to pay and value of a statistical life between the different studies.

\textbf{Conclusion}: Different methods used to elicit willingness to pay and the value of a statistical life year resulted in a wide range of estimates.

\textbf{Introduction}

The rapid pace at which discoveries in basic and clinical research are being translated into medical practice has resulted in tremendous advances in how disease is prevented, detected, and treated. To keep pace with emerging health technologies, healthcare systems and decision makers require timely information on the value of new interventions. Cost-effectiveness analyses (CEA) compare the costs and effects of two or more interventions, with the result expressed in terms of the incremental cost-effectiveness ratio (ICER). For decision making about what constitutes a cost-effective health technology, the ICER must then be compared to a pre-specified referent value, or cost-effectiveness (CE) threshold. As health technology assessment (HTA) has proliferated to meet the needs of global healthcare systems, several countries use explicit or implicit CE thresholds in healthcare decision making.

The CE threshold is defined as the maximum cost per health outcome that a health system is willing to pay \cite{1}. Whether and how to set a CE threshold has been a source of considerable debate. Central to this debate is the issue of defining an appropriate value for the CE-threshold based on opportunity costs based on either the health forgone through displacement of other-health generating interventions or the consumption value of health \cite{2}. Three general approaches to set CE thresholds are commonly used, including per capita income based thresholds, benchmarking interventions, and league tables \cite{3}. The World Health Organization’ Choosing Interventions that are Cost-Effective (WHO-CHOICE) project recommends a threshold of less than three times the national annual gross domestic product (GDP) per capita, with interventions that cost less than once the national annual GDP per capita considered highly cost-effective \cite{4}. In the USA, the commonly cited $50,000 CE threshold is thought to be benchmarked to the cost of treated end-stage renal disease when it became a criterion for Medicare-enrollment under the Nixon administration in the 1970s \cite{5,6}. Both of these approaches take a consumption value of health perspective, rather than basing the threshold on the amount of healthcare resource required to improve health \cite{2}. On the other hand, the league table approach attempts to maximize the health impact given a set healthcare budget with interventions ranked according to their ICERS. The cost-effectiveness analysis registry of Tufts Medical Center and the WHO-CHOICE league table are examples of this approach \cite{4,7}.
Estimates of individuals’ willingness to pay (WTP) to improve their own health are often used to define the consumption value of health [2]. The measurement of WTP seeks to define social value based on revealed and stated preference approaches [5,8]. In addition to estimates of WTP, the value of a statistical life (VSL) is a commonly used metric that examines the risk/reward trade-offs that individuals make with regard to their health. The study of the WTP and VSL is premised on the idea that health system decisions about how to allocate resources should represent societal preferences. Although the goal of WTP, VSL and CE threshold studies is often to characterize societal preferences to arrive at an acceptable value threshold, the methods for surveying individuals, including the survey instruments and the population sampled varies widely.

The VSL approach measures the WTP for an incremental change in the risk of death. In economic terms it equals the marginal rate of substitution between wealth and survival probability [9]. For example, if the WTP for risk reduction of 1/10,000 in a given period is 100 USD, the associated VSL would be 100/0.0001 = 1 million USD. One implication of this approach is the “dead anyway” effect that asserts that VSL increases with baseline risk [10]. The WTP approach on the other hand directly relates to the value of a quality adjusted life year (QALY), by asking respondents how much they are willing to spend for a certain health gain in terms of quality and lengths of life [11]. Quality of life is usually measured by utility values that are in turn elicited by either standard gamble or time trade off methods.

The objective of this systematic literature review is to comprehensively examine the survey methods used to estimate the VSL, WTP, or CE threshold in a global context. While several recent reviews have summarized the literature on estimating a CE threshold [12,13], we present a critical examination of the strengths and limitations of survey methods that incorporates an epidemiologic approach to discuss issues of sample selection and potential biases that may impact the resulting CE threshold estimate. Addressing this gap in our current understanding of methods for estimating the CE threshold is an important step towards proposing rigorous thresholds that take the diverse perspectives of individuals within a society into account.

Methods

Data sources and search strategy

Two independent searches using the PubMed resource, a service of the National Library of Medicine®, were conducted to systematically identify studies for this literature review. The first search used terms (VSL[title/abstract] OR Value of a Statistical Life[title/abstract] OR Value of Statistical Life[title/abstract] OR Value of Life[title/abstract] OR Value-of-Life[title/abstract] OR Preventing a Statistical Fatality[title/abstract]) AND (Survey[title/abstract] to identify VSL studies. To identify WTP per QALY studies, the second search used the terms (Willingness to Pay[title] OR WTP[title]) AND (QALY[title] OR Quality-Adjusted[title] OR Life-Year[title] OR Life Year[title]). All results were filtered to return only articles with full text in the English language.

Study selection

Studies identified from the initial searches were first screened by the title of the manuscript and the contents of the abstract. The full text was then reviewed for studies with titles and abstracts that indicated that they met the inclusion criteria. Studies were excluded if during this two-step sequential screening process, it was revealed that the full text was not in English and if they did not use a survey to estimate either WTP or VSL.

Comparison of available estimates from reviewed studies to published guidance

In addition to summarizing the estimates of WTP and VSL from the studies meeting the inclusion criteria for this review, these estimates were compared to published guidance for CE thresholds. To do this, we compared the estimates from our literature review to those estimated by Nanavaty et al. in their review of global CE thresholds [14]. Specifically, the reviewed study estimate was compared to the explicit or implicit country-specific cost per QALY threshold identified by Nanavaty et al. and ranked as being above or below the published threshold. Finally, a comparison of reviewed studies that were conducted in Asian countries to the WHO-CHOICE recommended CE threshold of 1–3 times GDP per capita was conducted to show the regional variation in the consumption value of health.

In addition, all estimates in from reviewed studies were converted to US dollar values to enhance the comparability of estimates across countries and currencies. To convert currencies, the online Google Finance currency conversion tool was accessed on 7 December 2017 [15].

Results

The initial search produced a total of 67 studies (Figure 1). A total of 18 studies were included in the final review after screening for content and duplicates. Of the 18 studies, 15 WTP studies and 3 VSL studies were included in the review. Studies were conducted in diverse geographical...
regions including Asia [16–21], Europe [11,22–24], the Middle East [25], and North America [26–30]. In addition, one study by Shiroiwa et al. took a global perspective, including countries across multiple regions [31].

Willingness to pay (WTP) studies

Table 1 outlines the results of the WTP literature search by region. Shiroiwa et al. conducted an international survey with the goal of measuring and comparing CE thresholds across countries in East Asia and the West [31]. Specifically, individuals aged 20 to 59 years were randomly sampled from internet panels in Taiwan (n = 504), Japan (n = 1,114), the Republic of Korea (n = 1,000), Australia (n = 1,000), the UK (n = 1,002), and the USA (n = 1,000). A contingent valuation survey was administered online to all participants. A bidding game was presented to elicit an individual’s WTP for each question. The authors found that the cost per QALY estimated using the nonparametric Turnbull method, to obtain the nonparametric maximum likelihood estimators of the probability that the respondent will reject the bid values [32], was JPY 5.0 million in Japan (USD $41,000 per QALY), KNW 608 million in the Republic of Korea (USD $74,000 per QALY), NT 2.1 million in Taiwan (USD $77,000 per QALY), £23,000 in the UK (USD $36,000 per QALY), AU$64,000 in Australia (USD $47,000 per QALY), and USD $ 62,000 per QALY in the USA. In terms of US$ adjusted by comparative price levels, the Republic of Korea and Taiwan had the highest WTP and no relationship between WTP per QALY and GDP per capita was observed in this international study.

Of the four WTP studies conducted in Asia, two were conducted in Southern Thailand [17,18,20], one in Penang, Malaysia, and Shiroiwa conducted a study in Japan [19], distinct from the international study discussed previously. Using face-to-face interviews and a visual analogue scale and time trade-off exercise, Thavorncharoensap et al., reported a mean WTP per QALY of 59,000–285,000 baht in Thailand (USD $1,810–$8,742) for treatment and a mean WTP per QALY of 26,000–137,000 baht (USD $798–$4,202) for preventive interventions [20]. Nimdet et al., reported a mean WTP per QALY of 244,720 baht (USD $7,505) and 243,120 baht (USD $7,458) using face-to-face interviews and a visual analogue scale and a EQ-5D scale of measurement, respectively [17]. Shafie et al., conducted a cross-sectional contingent valuation face-to-face interview study in Malaysia and found a mean WTP per QALY of MYR 29,080 (equivalent to US$ 9,000) [18]. Finally, Shiroiwa et al., randomly sampled 2,400 participants from an internet panel in Japan and found a significant gradient between WTP and the severity of health state, ranging from JPY 2 million (US$ 20,000) to JPY 8 million (US$ 80,000) [19].
Table 1. Summary of willingness to pay (WTP) studies by region, N = 15.

| First Author | Year | Setting | Age (Mean) | Gender (% Male) | n   | Sample Description      | Administration Method |
|--------------|------|---------|------------|-----------------|-----|------------------------|-----------------------|
| Shiroiwa     | 2010 | Japan, Korea, Taiwan, Australia, UK, and US | N/R | 50.36% | 5500 | Randomly Selected | Online Double-Bounded Dichotomous -> Bidding Game |
| Shiroiwa     | 2013 | Japan | N/R | 50.80% | 2283 | Randomly Selected | Online Double-Bounded Dichotomous -> Bidding Game |
| Thavorncharoensap | 2013 | Thailand (Southern) | 40.1 | 42.91% | 1191 | Randomly Selected | Face-to-Face Bidding Game |
| Shafie       | 2014 | Penang, Malaysia | N/R | 47.60% | 347 | Randomly Selected | Face-to-Face Double-Bounded Dichotomous -> Bidding Game |
| Nimdet       | 2015 | Thailand (Southern) | 41.4 | 46.20% | 554 | Randomly Selected | Face-to-Face Single Question |
| Gyrd-Hansen  | 2003 | Denmark | 46.8 | 56% | 3201 | Randomly Selected | Face-to-Face Closed-Ended Question Payment Scale (Bidding Game) -> Bounded Open-Ended (OE) Question |
| Bobinac      | 2010 | Netherlands | 42.1 | 47% | 1091 | Randomly Selected | Face-to-Face Online |
| Robinson     | 2013 | Netherlands, UK, France, Spain, Sweden, Norway, Denmark, Poland, and Hungary | N/R | N/R | 21896 | Randomly Selected | Online Risk Variant Questions -> Card Sort; Time Variant Questions -> Card Sort |
| Martin-Fernandez | 2014 | Spain | 65.4 | 60.7% | 662 | Randomly Selected | Face-to-Face Bidding Game |
| Ahlert       | 2016 | Germany | N/R | N/R | 5000 | Representative Sample | Face-to-Face or Online Decision Scenario |
| Moradi       | 2015 | Tehran City, Iran | 61.1 | 73% | 149 | Diabetes Patients | Face-to-Face Double-Bounded Dichotomous Choice |
| Byrne        | 2005 | Harris County, Texas, USA | 48.4 | 47.1% | 193 | Randomly Selected | Face-to-Face Open Ended Question |
| Franic       | 2005 | USA | 29 | 0% | 146 | Convenience Sample | Face-to-Face Bidding Game |
| Lieu         | 2009 | USA | N/R | 60.9% | 478 | Representative Sample | Online Bidding Game -> Open Ended Question |
| Haninger     | 2011 | USA | 46 | 48% | 2858 | Randomly Selected | Online Double-Bounded Dichotomous Choice |

Results:

- JPY 5 million (Japan)
- KWN 68 million (ROK) NTS 2.1 million (Taiwan)
- £23,000 (UK) AUS 64,000 (Australia) US$ 62,000 (US)
- JPY 1M per 0.1 Utility Increase
- 0.4 to 2x GDP Per Capita
- 29,080 MYR (9,000 USD)
- VAS = 244,720 Baht EQ-5D-3L = 243,120 Baht
- 88,000 DKK VAS = 12,900 Euros EuroQol = 24,500 Euros
- 18,247–34,097 USD
- 11757–23033 USD for treatment toxicity
- 11757–23033 USD for breast cancer
- 26,000–45,000 USD Could Not be Accurately Estimated
Five European WTP studies were included in our review. Conducting face-to-face interviews using an EQ-5D health state and a closed ended questions, Gyrd-Hansen found a WTP per QALY of DKK 88,000 (USD $13,968) [24]. Bobinac et al. used an online questionnaire with a representative sample of participants in the Netherlands (n = 1,091) and estimated a WTP per QALY of 12,900 Euros (USD $15,219) using a visual analogue scale and 24,500 Euros (USD $28,905) based on the EQ-5D valuation of health states [23]. In a study across several European countries, Robinson et al. used a standard gamble and time trade-off, followed by a chained approach of presenting a monetary value to avoid some risk/duration. They estimated a WTP per QALY ranging from US$18,247 to US$ 34,097 [33].

Martin-Fernandez conducted a contingent valuation exercise with a convenience sample of 662 participants aged 18 years or older in Madrid Spain using the EQ-5D to elicit WTP per QALY [11]. Mean WTP per QALY was estimated to be 10,119 Euros (USD $11,938) when subjects estimated WTP from personal money versus 28,187 (USD $33,255) when WTP per QALY was estimated from taxes. Finally, Ahlert et al., conducted a comparative methodological study of WTP in Germany with the goal of understanding how the method of administration and wording of the questionnaire, including the emphasis placed on the hypothetical health state, impacted the WTP estimate [22]. Based on various framing techniques, the WTP per QALY ranged from 8,580 Euros to 18,420 Euros (USD $10,123 – $21,732).

One study was conducted in the Middle East. Using a double-bounded dichotomous face-to-face interview survey amongst 149 diabetic patients in Tehran, Iran, Moradi et al. estimated a mean WTP per QALY of US$ 2,107 – US$ 4,453 [25].

Four WTP studies were conducted in North America, all in the USA. Byrne et al. randomly selected 193 adults in Harris County, Texas and conducted a telephone survey followed by a face-to-face interview to estimate WTP per QALY [26]. Using an open-ended question to elicit preferences, they found a mean WTP ranging from USD $1,221-$5,690. Franic et al. used a bidding game conducted in face-to-face interviews in convenience sample of 146 women and reported WTP ranging from USD $187 – $408 for post-chemotherapy treatment toxicity and USD $11,757 – $23,033 for avoidance of various breast cancer scenarios [27]. Ultimately, Franic et al. concluded that the QALYs were a poor predictor of WTP in this study. Lieu et al. conducted an internet based survey using time trade-off and WTP questions among a nationally representative survey research panel (n = 478). They estimate a mean WTP per QALY USD $26,000 – $45,000 [29]. Finally, Haninger sampled 2,858 randomly selected US adults and attempted to elicit WTP using a double bounded dichotomous choice method [28]. Ultimately, the authors concluded that participants did not have a constant rate of WTP per QALY when valuing acute foodborne illness.

**Value of a statistical life (VSL) studies**

Only three VSL studies with survey methods were uncovered during our review (Table 2) [16,21,30]. Of these, two were conducted in Asia and one in the USA. Lee et al. used online survey methods, employing a double-bounded dichotomous choice experiment among 1,434 people living in large Korean cities [16]. The authors estimated a VSL of 796 million KRW per person (USD $728,271). Yang et al. conducted a face-to-face stated preference experiment among 1,277 motorists and non-motorists in Nanjing, China to investigate travelers’ willingness to pay for traffic risk reduction and found that VSL was estimated at mean value of CNY 3,729,493 (USD $563,367) and CNY 2,181,592 (USD $329,536) for motorists and non-motorists, respectively [21].

### Table 2. Summary of value of a statistical life (VSL) studies by region, N = 3.

| First Author | Year | Setting | n     | Sample Description | Administration | Method                                      | Results                      |
|--------------|------|---------|-------|--------------------|----------------|---------------------------------------------|------------------------------|
| **Asia**     |      |         |       |                    |                |                                             |                              |
| Lee          | 2015 | Korea   | 1,434 | Randomly Sampled   | Online         | Double-Bounded Dichotomous Choice            | 796,000,000 KRW              |
| Yang         | 2016 | Nanjing, China | 1,277 | Randomly Sampled motorists and non motorists | Face-to-Face Stated Preference | 3,729,493 RMB for motorists and 2,181,592 for non-motorists | 7,184,406 RMB estimated across income groups from tax system social fund contributions |
| **North America** | | USA     | 77    | Randomly Sampled (Undergraduate Students) | Mailed         | Stated Preference                           | 5,100,00 USD                 |
In the USA, Muller et al., conducted a stated preference experiment among 77 undergraduate students [30]. They estimated a value of US$ 5,100,000 when extending willingness to pay of $180 per year for an implied level of fatality risk reduction across the expected life course.

Comparison of reviewed studies to published guidance

Table 3 compares the results from our reviewed studies to available published thresholds identified by Nanavaty et al. Of the studies identified in our review that were conducted in countries with established CE thresholds identified by Nanavaty et al., the WTP estimates were all under the published thresholds [11,23,26,29,31,33].

Using the WHO-CHOICE recommendation that interventions with a cost per QALY of less than 3-times GDP be considered cost-effective, Table 4 explores the results from WTP and VSL studies conducted in Asian countries that were included in our review in comparison to the WHO-CHOICE recommendation for GDP in 2016 [44].

Discussion

Out of a total of 18 studies reviewed across five global regions, we found enormous variation in WTP and the estimated VSL both within countries and across regions. In the US alone, estimates ranged from US$ 1,221 when and open ended question about osteoarthritis health states in the small study by Byrne et al. [26] to US$ 62,000 in the international study by Shiroiwa et al. [31]. Within Asia, we also find substantial variation in WTP, with Japanese estimates ranging from JPY 8 million to JPY 2 million, again based on the severity of the health state [19]. Overall, from this review, it is clear that regional CE thresholds are difficult to determine from available evidence. Instead, we focus on how the choice of method impacts the WTP and VSL estimates and provide some context of how the estimates in studies reviewed here compare to the guidance of the WHO that the CE threshold be 2-3x GDP and other published thresholds.

Methodologic issues in estimating the cost effectiveness threshold from willingness to pay and value of a statistical life studies

There are several methodological issues that should be considered in estimating the CE threshold from available evidence. From an epidemiologic perspective, the strength of the evidence can be assessed using several key parameters.

Sample selection and study population

The sampling frame, sample size, and characteristics of the final sample are all important considerations. Mean age and the percentage of respondents that were male are shown in Table 1. A common method of identifying participants for WTP and VSL studies is to randomly

| Country   | ICER Threshold (Cost per QALY) | Review Study Result   | Relationship between WTP/VSL and Established Threshold |
|-----------|--------------------------------|-----------------------|--------------------------------------------------------|
| Australia | 69,000 AUD [34]                 | 64,000 AUD [19]       | Below                                                  |
| Canada    | 42,000–78,000 AUD [35]          | No results (NR)       | NR                                                     |
| Ireland   | 32,000–$137,000 CAD [37]        | €45,000 [38]          | NR                                                     |
| The Netherlands | €10,000–€80,000 [39, 40] | €12,900–€24,500 [23,33] | Below                                                  |
| New Zealand | 20,000 NZD [34]                | €30,000 [41]          | NR                                                     |
| Spain     | £20,000–£30,000 [34,42,43]   | £10,119–£28,187 [11] | Below                                                  |
| UK        | £50,000 [6, 37]                 | £23,000 [31]          | Below upper bound                                      |
| USA       | $50,000 [6, 37]                 | $1,221–$45,000 [26, 29] | Below                                                  |

| Country   | Review Study Result | Review Study(ies) | GDP per capita (2016) | 3-times GDP per capita (2016) | Relationship between WTP/VSL and WHO-CHOICE Recommendation |
|-----------|---------------------|-------------------|-----------------------|-------------------------------|-------------------------------------------------------------|
| Japan     | 1,000,000–5,000,000 JPY [19,31] | 2010; 2013       | 3,906,787 JPY (2010) | 11,720,361 JPY (2010)–12,179,250 JPY (2013) | Below or Within Range                                      |
| Korea     | 68,000,000 KRW [31]  | 2010              | 25,533,865 KRW (2010) | 76,601,595 KRW                | Within range                                               |
| Thailand  | 244,720–243,120 THB [12] | 2015              | 138,565 THB (2015)   | 415,695 THB                   | Within range                                               |
| Malaysia  | 29,080 MYR [18]     | 2014              | 33,899 MYR (2014)    | 101,697 (MYR)                 | Below                                                      |
select individuals from established internet panels. This method offers the convenience of reaching potentially large numbers of individuals, but it is important to consider how representative the sample obtained is compared to the general population. Often online surveys can skew towards a younger population than the general population. While the majority of studies reported basic demographic information for the respondents, verification of this information in online-based studies is limited. This can be accomplished by comparing the demographics of the sample population to the demographics of the entire internet panel and to the demographics of the geographic region or country over which the results of the study are to be applied. This is especially important in the context of income. Across studies included in this review, variation in WTP and VSL estimates occurred by participants’ income level, with higher income participants putting higher values on smaller incremental differences in health states. It is also important to consider the age of the sample. Results from younger participants tend to estimate higher CE thresholds, at least in the USA, as seen by the study of undergraduate students conducted by Muller et al. [30].

Randomly sampling a representative sample of participants from the general population is likely to return very different results from sampling patients with a specific health condition. While sampling patients with a specific health condition may be useful for health care decision makers wanting to understand how their decisions will impact those most affected, setting a general CE threshold using this method is not recommended for several reasons. In this review, Moradi et al. was the only study included that used this sampling method, enrolling and surveying diabetic patients. They found a WTP value of US$ 4,453, substantially lower than the WHO recommendation of 2-3x GDP. It is possible that lower threshold values will be obtained from disease patients because of a biased perspective when estimating utility from improvements in health states [25].

Mode of survey administration

Of the included studies with surveys, 43% (n = 6) were administered online, 50% (n = 7) were administered using face-to-face interviews, and 7% (n = 1) included both online and face-to-face interviews. Interestingly, Ahlert et al. reported that the survey setting (online versus face-to-face) changed the WTP estimate substantially, with participants receiving the face-to-face survey reporting a WTP per QALY of 18,420 Euro, compared to 10,892 Euro WTP per QALY when the same survey was administered online [22]. While it is possible that the human interaction of face-to-face interviews evokes a higher WTP, other studies using face-to-face methods still found WTP ratios that were considerably lower than the 2-3x GDP threshold [11,18,20,26]. This result was consistent across regions. Given the higher cost and resource burden of face-to-face compared to online surveys, trade-offs between the potential to obtain a larger, potentially more representative sample of participants in online surveys and the suggestion that face-to-face interviews may increase WTP should be considered.

Survey methods

Methods to elicit WTP and VSL also varied across studies and regions. Contingent valuation methods, including bidding games and discrete choice experiments (double-bounded dichotomous choice) were used alone or paired in several studies. The results of these studies may be quite sensitive to the valuation of the health states based on the specific scenarios presented. For example, Haninger concluded that WTP could not accurately be estimated because of large differences when the health states presented were about chemotherapy toxicity versus breast cancer cures [28].

Health states

Within the studies there were different strategies used in regards to the health states presented to survey respondents. The differences in health states were mainly driven by study design and methodology. Shiroiwa et al. determined that involving specific health states and descriptions of the health states confused respondents [31]. Without specific health states respondents were simply asked how much they were willing to pay for one additional year of perfect health [31]. In other studies, such as the later study by Shiroiwa et al. in Japan, there were health states provided to respondents based on EQ5D scores. This mix of the inclusion and exclusion of health states in studies, as well as the variability in the specific health states from study to study makes comparing across studies difficult [19].

With the growing cost of new health technology, health care decision makers around the world must make difficult decisions about what constitutes reasonable value for money. The considerable variation in each countries’ available resources, as well as the relative cultural value placed on health, warrants country-specific CE thresholds. However, few countries explicitly state a CE threshold. In accordance with WHO-CHOICE recommendations, the most commonly cited CE thresholds are those based upon a country’s per-capita GDP and the estimate of the economic value of a year of
healthy life from the Commission on Macroeconomics and Health [45]. Specifically, economists from the WHO-CHOICE project recommend a CE threshold <3xGDP [4]. These thresholds, explicit or implicit, are overwhelmingly elicited to reflect the consumption value of health, rather than to reflect the opportunity costs of healthcare spending. Woods et al. argue that setting a CE-threshold in which the consumption value of health does not equal the amount of health care resource required to improve health is a sign that the healthcare system may not be meeting individual preferences [2]. Our results suggest that using WTP to estimate the consumption value of health may often result in lower valuation than the currently available thresholds, but that this is sensitive to the sociodemographic characteristics of the surveyed population.

In addition to the WHO-CHOICE threshold, several countries have published ICER thresholds. The review by Nanavaty et al. conducted in 2015 identified published thresholds in Australia, Canada, Ireland, the Netherlands, New Zealand, Spain, the UK, and the USA [14]. Although many of the studies identified in our review were conducted in Asia, no published CE thresholds exist specific to Asian countries. However, the results from the WTP studies in Asia, presented in Table 1, show that the methods used to elicit the societal WTP come in below the WHO-CHOICE threshold for all countries except for Korea. Shiroiwa et al. discuss their findings from the Korean study of WTP and comment that this may be a true representation of social and cultural values in Korean society [31].

This review contributes important information about the global measurement of WTP and VSL studies. Unlike other recent reviews, the focus of this project was to critically evaluate how the sample, mode of administration, and survey methods can influence CE estimates. The results of this review should be interpreted in the context of a few important limitations. First, we restricted the review to English language articles with full text available in PubMed. In addition, the published estimates vary across time and place, making a direct comparison of published estimates difficult. Another limitation of this study is that we converted each study to USD based on an exchange rate at a given time even though the studies were completed at different time points.

We suggest future researchers build on the results and shortcomings of this study. While we acknowledge that the number of VSL studies (3) is fewer than WTP studies (15), both of these concepts play an important role in policy decisions, so we chose to include both in this review. Additional VSL studies are needed to help policy-makers with their decisions. Also, we believe the methodological differences can only explain part of variation of results. Other issues, such as heterogenous sampling populations, could also impact the results. As such, we believe users of these studies should perform sensitivity and scenario analyses around the thresholds implied.

In summary, this review provides important evidence that different methods used to elicit WTP and VSL lead to a wide range of different values. In addition, the values from the reviewed studies also highlight the substantial global heterogeneity in the consumption value of health, suggesting that the individual preferences elicited incorporate additional costs that fall on consumption opportunities outside of health. As noted by Woods et al., this valuation of the opportunity cost of funding additional interventions is fundamentally different from estimating the opportunity costs in terms of health forgone when all costs fall on healthcare budgets [2].

This study was funded by Janssen Pharmaceutical KK. JM is an employee of Janssen Pharmaceutical KK.

Disclosure statement
JAM, WEF, and BCMW are paid consultants to Janssen Pharmaceutical KK. JM is an employee of Janssen Pharmaceutical KK.

Funding
This study was funded by Janssen Pharmaceutical KK.

References
[1] Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev Pharmacoecon Outcomes Res. 2008;8(2):165–178.
[2] Woods B, Revill P, Sculpher M, et al. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. Value Health. 2016;19(8):929–935.
Rocchi A, Menon D, Verma S, et al. The role of economic resilience of the $50,000-per-QALY threshold. N Engl J Med. 2014;371(9):796–797.

Tufts Medical Center. Center for the evaluation of value and risk in health. Cost-effectiveness analysis registry. [cited 2017 Oct 22]. Available from: https://cevr.tuftsmc.com/databases/cea-registry

King JT Jr., Tsevat J, Lave JR, et al. Willingness to pay for a quality-adjusted life year: implications for societal health care resource allocation. Med Decis Making. 2005;25(6):667–677.

Jones-Lee M. The value of changes in the probability of death or injury. J Political Econ. 1974;82(4):835–849.

Pratt J, Zeckhauser R. Willingness to pay and the distribution of risk and wealth. J Political Econ. 1996;104(4):747–763.

Martin-Fernandez J, Polentinos-Castro E, Del Cura-Gonzalez M, et al. Willingness to pay for a quality-adjusted life year: an evaluation of attitudes towards risk and preferences. BMC Health Serv Res. 2014;14:287.

Nimdet K, Chaiyakunapruk N, Vichansavakul K, et al. A systematic review of studies eliciting willingness-to-pay per quality-adjusted life-year: does it justify CE threshold? PLoS One. 2015;10(4):e0122760.

Vallejo-Torres L, Garcia-Lorenzo B, Castilla I, et al. On the estimation of the cost-effectiveness threshold: why, what, how? Value Health. 2016;19(5):558–566.

Nanavaty M, Kaura S, Mwamburi M, et al. The use of incremental cost-effectiveness ratio thresholds in health technology assessment decisions. J Clin Pathways. 2015;1(1):29–36

Google Finance. [cited 2017 Oct 22]. Available from: https://finance.google.com/finance?q=USDEU

Lee G, Lee Y, Lee H, et al. Value of a statistical life estimation of carcinogenic chemicals for socioeconomic analysis in Korea. Environ Health Toxicol. 2015;30(Suppl):s2015005.

Nimdet K, Ngorsuraches S. Willingness to pay per quality-adjusted life year for life-saving treatments in the state of Penang, Malaysia. Clinicoecon Outcomes Res. 2014;6:473–481.

Shiroiwa T, Sung YK, Fukuda T, et al. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ. 2010;19(4):422–437.

Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc. 1976;38(3):290–295.

Robinson A, Gyrd-Hansen D, Bacon P, et al. Estimating a WTP-based value of a QALY: the ‘chained’ approach. Soc Sci Med. 2013;92:92–104.

Cleemput I, Neyt M, Thierry N, et al. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. Int J Technol Assess Health Care. 2011;27(1):71–76.

George B, Harris A, Mitchell A. Cost-effectiveness analysis and the consistency of decision making: evidence from pharmaceutical reimbursement in australia (1991 to 1996). Pharmacoconomics. 2001;19 (11):1103–1109.

Laupacis A, Feeny D, Detsky AS, et al. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. CMAJ. 1992;146(4):473–481.

Rocchi A, Menon D, Verma S, et al. The role of economic evidence in Canadian oncology reimbursement decision-making: to lambda and beyond. Value Health. 2008;11(4):771–783.
[38] Barry M, Tilson L. Recent developments in pricing and reimbursement of medicines in Ireland. Expert Rev Pharmacoecon Outcomes Res. 2007;7(6):605–611.

[39] Franken M, Koopmanschap M, Steenhoek A. Health economic evaluations in reimbursement decision making in the Netherlands: time to take it seriously? Z Evid Fortbild Qual Gesundhwes. 2014;108(7):383–389.

[40] Pauwels K, Huys I, Casteels M, et al. Market access of cancer drugs in European countries: improving resource allocation. Target Oncol. 2014;9(2):95–110.

[41] Rodriguez JM, Paz S, Lizan L, et al. The use of quality-adjusted life-years in the economic evaluation of health technologies in Spain: a review of the 1990–2009 literature. Value Health. 2011;14(4):458–464.

[42] Buxton MJ. Economic evaluation and decision making in the UK. Pharmacoconomics. 2006;24(11):1133–1142.

[43] Simoens S. How to assess the value of medicines? Front Pharmacol. 2010;1:115.

[44] Trading Economics. GDP per capita. 2016 [cited 2017 Oct 22]. https://tradingeconomics.com/country-list/gdp-per-capita-ppp

[45] Bertram MY, Lauer JA, De Joncheere K, et al. Cost-effectiveness thresholds: pros and cons. Bull World Health Organ. 2016;94(12):925–930.