Article:

Shiroiwa, T., Morii, Y., Hoshino, E. et al. (2026) Departures from health universalism? A value set of AP-7D in Japan as an attempt to develop a “culture-specific” preference-based measure. Social Science and Medicine, 393. 119022. ISSN: 0277-9536

https://doi.org/10.1016/j.socscimed.2026.119022

Social Science & Medicine 393 (2026) 119022

Contents lists available at ScienceDirect

Social Science & Medicine

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

Departures from health universalism? A value set of AP-7D in Japan as an attempt to develop a "culture-specific" preference-based measure

Takeru Shiroiwa [a,*], Yasuhiro Morii [a], Eri Hoshino [b], Tatsunori Murata [c], Richard Norman [d], Brendan Mulhern [e], Nan Luo [f], Donna Rowen [g], Takashi Fukuda [a]

[a] *Center for Outcomes Research and Economic Evaluation for Health (C2H), National Institute of Public Health, Wako, Japan*
[b] *Division of Policy Evaluation, Department of Health Policy, Research Institute, National Center for Child Health and Development, Tokyo, Japan*
[c] *CRECON Medical Assessment Inc., Tokyo, Japan*
[d] *School of Population Health, Curtin University, Perth, WA, Australia*
[e] *Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW, Australia*
[f] *Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore*
[g] *Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, UK*

ARTICLE INFO

Keywords:
Health universalism
Health pluralism
Preference based-measure
AP-7D
QALY
Cultural-specific measure

ABSTRACT

Health universalists believe that preference-based measure (PBM) instruments can be applied across cultures because they share similar health concepts in the world. This is the prevailing policy in PBM development. However, health pluralists refute this idea, as they argue that the concept and components of health differ depending on culture. To incorporate the pluralist view, we developed the Asian Preference-Based Measure-7 Dimensions (AP-7D), a "culture-specific" PBM for Asian countries. This survey aimed to address cultural differences in utility measurement by developing an AP-7D value set in Japan, as part of a series of AP-7D developments. This study used a web-based survey to evaluate AP-7D health states with a triplet discrete choice experiment (DCE). The design followed an established international protocol. We conducted the web-based survey and data collection in October 2024. Respondents aged 20–79 were recruited via quota sampling based on sex and age. A total of 2681 individuals were included in the DCE analysis. We applied a mixed logit model to the DCE data and obtained decrements for each level in each domain. The worst health state had an AP-7D value of −0.448. Pain/discomfort, mobility, and burden to others were the most influential items on AP-7D values. Only one non-monotonicity were observed. Our survey successfully created the first AP-7D culture-specific PBM value set, and we can convert responses to AP-7D value for QALY calculation. We believe that our pluralistic approach is a novel and important attempt to reconsider health universalism and inform the future development of PBMs.

1. Introduction

1.1. Health universalism and pluralism

The concept of "morality" in multicultural societies is frequently discussed in relation to bioethics. Some argue for a common (or universal) morality that applies to everyone, regardless of region, culture, or belief (Beauchamp and Childress, 2019). They insist that this notion is supported by common sense, which is shared by people across regions and historical periods. However, some opponents of universalism argue that the concept cannot sufficiently capture the voices of minorities because "patients and their families bring many different cultural models of morality, health, illness, healing, and kinship to clinical encounters" (Turner, 2003).

Using these terminologies, we can discuss the policy of developing health-related quality of life (HRQoL) instruments. As will be discussed later, most of these instruments have been mainly developed by Westerners, and they constructed the concepts, domains, and items primarily by using data mainly collected from Western populations. Health universalists believe these instruments can also be applied to people in non-Western environments because they assume a shared global concept of health. However, health pluralists may find it difficult to accept such an

* Corresponding author. Center for Outcomes Research and Economic Evaluation for Health (C2H), National Institute of Public Health, 2-3-6 Minami, Wako, Saitama, 351-0197, Japan.
E-mail address: t.shiroiwa@gmail.com (T. Shiroiwa).

https://doi.org/10.1016/j.socscimed.2026.119022
Received 20 August 2025; Received in revised form 19 January 2026; Accepted 22 January 2026
Available online 29 January 2026

idea, because they believe the concept and components of health differ depending on culture. Even if pluralists accept that the structure of human bodies is similar and that physical and mental health are universally important, other aspects of health concepts appear to be culture-dependent.

In addition to the difference in health components, people in other regions have difficulty even understanding some health-related concepts derived from Western culture. For example, recently, the development of instruments for measuring well-being has progressed (Brazier et al., 2022). However, many Japanese people do not fully understand the meaning of "well-being," a recent keyword in health science that originates with Aristotle (Kraut, 2022), one of the fathers of Western philosophy. The Japanese language lacks a direct equivalent to the word (some authors of this paper are native Japanese speakers). Consequently, well-being is often conflated with terms such as "(good) health," " "happiness," or "(good) quality of life." Therefore, it is natural to assume cultural differences in how well-being is conceptualized, although empirical evidence remains limited.

As a standard process to develop instruments, surveys have examined the psychometric properties of EQ-5D for people in different cultures (for example, aboriginal Australians (Ribeiro Santiago et al., 2021) and Māori (Perkins et al., 2004)). Value sets of PBMs are normally developed in each country to reflect people's preferences for health states because of differences in health preferences among countries. However, these practices are based on the belief that health preferences vary among different cultures, but the health concepts states do not change significantly so the components and descriptions do not need to be changed, even when PBMs are used by substantially different cultural groups. From a pluralistic viewpoint, this seems a strong assumption. Few empirical studies have examined whether universal PBMs can overcome cultural and regional differences.

1.2. Consideration of cultural differences by using "culture-specific" PBMs

This study focuses on preference-based measures (PBMs), which are used to calculate quality-adjusted life years (QALYs). Many health technology assessment (HTA) agencies officially use QALYs for economic evaluations. Since the 1990s, several generic PBMs have been developed that are currently used as standardized instruments, such as EQ-5D (EUROQOL Group, 1990), the Health Utilities Index (HUI) (Torrance et al., 1995; Feeny et al., 2002), and the Short-form-6 dimensions (SF-6D) (Brazier et al., 1998, 2020). More recently, many disease- or condition-specific PBMs (Rowen et al., 2017) have been established, for instance, to measure the health of patients with cancer (King et al., 2016, 2021). Besides these PBMs, the Adult Social Care Outcomes Toolkit (ASCOT) (Netten et al., 2012) can be used to calculate social care–related quality of life, EQ-5D-Y to measure the HRQOL of younger people (Wille et al., 2010), and the ICEpop CAPability measure (ICECAP) (Coast et al., 2008) to measure capabilities as advocated by Sen. However, most of these widely used PBMs were developed in Western countries: EQ-5D (Kind et al., 2005) in Western Europe; HUI (Torrance et al., 1995; Horsman et al., 2003) in Canada; 15D (Sintonen, 2001) in Northern Europe; AQoL (Hawthorne et al., 1999) in Australia; and SF-6D (Brazier et al., 1998) (SF-36, which is origin of SF-6D in the US (Ware and Sherbourne, 1992; McHorney et al., 1993)), ASCOT (Forder and Caiels, 2011), and CHU-9D (Stevens, 2009) in the United Kingdom.

To clarify, we do not deny the role of universal PBM instruments. However, if we introduce some concepts of health pluralism to current universal PBMs, revising existing instruments may be necessary, but this scenario is often unrealistic because patent holders are not "allowed to make drastic changes to established questionnaires" even when important cultural adaptations are required (Cheung and Thumboo, 2006). Consequently, health pluralists must either accept minor changes to existing instruments or develop new ones that more appropriately reflect cultural differences.

To address this issue, we began to develop the Asian Preference-Based Measure-7 Dimensions (AP-7D). We describe it as one of the "culture-specific" PBMs for Asian countries. This is an attempt to consider cultural differences when measuring utilities from a pluralistic perspective, although there may be other ways to avoid such a time-consuming task (Fayers and Machin, 2016).[1] To establish the AP-7D classification system, we conducted an interview survey and qualitative analysis in nine Asian countries (Shiroiwa et al., 2022). Another PBM, the China Health-Related Outcomes Measures (CHROME), has also been developed for a specific country (Wu et al., 2022).

This empirical survey is intended to address cultural differences in utility measurement by developing a new PBM and establishing an AP-7D value set in Japan, with the goal of putting a culture-specific PBM into practical use. A value set is needed to transform responses into utility scores based on the preferences of the general population. This survey was performed using conventional methods for valuation surveys based on our constructed international standard protocol. First, we present the methods and results of our valuation survey in Japan. Secondly, we re-analyze the data from the study by Suzuki et al. to compare the responses between the EQ-5D and the AP-7D, taking into account the established value set. This contributes to clarifying the psychometric properties of the AP-7D (Suzuki et al., 2025). Then, we return to the discussion of health universalism and pluralism, as well as the interpretation of the empirical survey results.

2. Methods

2.1. Asia preference-based measure-7 dimensions

AP-7D was codeveloped by HTAsiaLink and the Center for Outcomes Research and Economic Evaluation for Health (C2H) as a culture- or region-specific PBM for general health states (Shiroiwa et al., 2022). HTAsiaLink is composed of official HTA bodies and universities in the Asia-Pacific region, and its members from the aforementioned nine countries participated in the development process of AP-7D. Some of these countries implement HTA systems for drug reimbursement and pricing, others for improvement of their healthcare systems. The nine-language version is available from the C2H site [https://c2h.niph.go.jp/ap7d/], which requires no registration and no fee to use.

AP-7D contains seven items (pain/discomfort (PD), mental health (MH), energy (EN), mobility (MO), work/school (WS), interpersonal interactions (II), and burden to others (BO)) with four levels for each item (not at all, a little, quite a bit, and very much). The instrument is simple enough to allow patients to respond quickly. Similar to EQ-HWB (Brazier et al., 2022) and one version of SF-6D (Brazier et al., 2020), the recall period is one week, as opposed to one day for EQ-5D; the questionnaire does not specify the recall period for HUI.

Some items, such as PD, MH, and MO, are also included in conventional PBMs such as EQ-5D, HUI, and SF-6D. These items seem universally relevant, as human bodies have a common structure and do not depend on culture. In contrast, items such as BO are not adopted by other generic PBMs because of the higher psychological barrier to receiving or seeking help from others (including anonymous individuals) among some Asian cultures, and the refusal to be a bother to other people may even be a virtue. Work may also fall into a different category compared to Western countries. While some PBMs consider work to be a component of daily activities (e.g., EQ-5D), certain Asian countries have a particular understanding of work as transcending

[1] "Developing new instruments is a time-consuming task. In summary, our advice is: don't develop your own instrument - unless you have to. Whenever possible consider using or building upon existing instruments. If you must develop a new instrument, be prepared for much hard work over a period of years."

everyday tasks. Some people view work as indicative of respectability, a capacity to fully contribute to one's family and society beyond earning money. Such assigned importance is clearly different from "leisure" as described in EQ-5D. In addition, energy constitutes one item of the vitality domain in the SF-36, while SF-6D includes only an item on fatigue. For some Asian countries, energy is a broader and more positive concept that does not necessarily pertain to a lack of fatigue. We mention these differences because the description of each item is important, as is knowing which concepts are included in the instrument.

2.2. Discrete choice experiment design

This study assessed AP-7D health states through a discrete choice experiment (DCE), which has been frequently applied to PBM valuation surveys (SF-6Dv2 (Mulhern et al., 2020; Shiroiwa et al., 2025), EORTC QLU-C10D (King et al., 2018; Shiroiwa et al., 2024a) etc.). A triplet (or "ternary") DCE was adopted, in which three health states (states A and B and "immediate death") were shown to respondents for each set of DCE questions. Health states A and B combine seven AP-7D items with expected life years (1, 2, 5, and 10 years). First, the participants chose their most preferred health state from the three, followed by their (second most) preferred state from the two left.

Appling triplet DCE to this survey was determined through expert discussion based on the pilot survey (Shiroiwa et al., 2024b), since the range of utilities measured by the AP-7D was too large when a simple DCE without immediate death was employed. This pilot survey showed that triplet DCE had intermediate characteristics between cTTO and paired DCE with duration. We understand that this methodology, which includes the immediate death card, is controversial (Flynn et al., 2008). However, it is also important to measure the preference of each health state to immediate death through a direct comparison.

Because each AP-7D item has four levels, there were $4^7 \times 4$ (levels of duration) = 65,536 possible health states, from which 300 were selected to construct 150 pairs. Combined with the immediate death card, these pairs were randomly allocated to 10 blocks, with each block containing 15 unique health state triplets. To reduce respondents' cognitive burden, we simplified the choice task by limiting the number of attributes that differed in each pair of health states to five of the seven AP-7D attributes. This was based on D-Optimal design methods in NGene. These triplet health states in the block were presented in a random order, and health state card positions (left or right) were also randomized to avoid position effects (Shiroiwa and Fukuda, 2024). Instead of color coding (Jonker et al., 2019), we used an underlined degree adverb for each item to reduce the complexity of the DCE task (Fig. 1).

2.3. Data collection and sampling

A cross-sectional valuation survey was conducted to collect preferences for health states from a representative sample of the Japanese general population. This study design was based on an international protocol constructed by our research group. The procedures were approved by the ethics committee of the National Institute of Public Health (NIPH-IBRA #24010).

Triplet DCE data were collected using a web-based survey. Respondents aged 20–79 were recruited through a research company panel (INTAGE Healthcare Inc.) using quota sampling by sex and age. Because some Asian countries legally define adults as individuals aged 20 or older, those aged 18 and 19 were excluded. In addition, the survey included only Japanese speakers (English and other languages were not used). The target sample size was 3,000, as recommended by Lancsar and Louviere (2008), and was larger than other valuation surveys (SF-6Dv2, EORTC QLU C-10D etc.)(Bahrampour et al., 2020). Potential respondents belonging to the survey company's web panel were e-mailed invitations to the online survey, and their informed consent was obtained on the first screen. Only those who agreed to our conditions were able to proceed with the survey. After providing informed consent and answering basic demographic questions (sex and age) to confirm consistency with their registered data and self-reports of AP-7D status, the respondents answered 15 triplet DCE questions (for a total of 30 choice tasks). Demographic information and survey feedback were collected. Respondents who dropped out by not answering all the questions including demographics were treated as having withdrawn their consent. The research company did not provide us with their data. Only those who had completed all questions were included in the analysis. In addition, the time spent on each and all DCE tasks was recorded.

2.4. Comparison EQ-5D-5L and AP-7D responses

To consider the relationship between EQ-5D-5L and AP-7D responses, we recalculated the Spearman correlation coefficients for each domain of both instruments using data from Suzuki et al. (2025). EQ-5D and AP-7D responses were collected from 500 Japanese people through face-to-face surveys. The detailed study design has been described by Suzuki et al. (2025). Participants were recruited based on non-random quota sampling by a research company considering sex and age categories. The respondents were sampled from the general population; therefore, most people were in good health and some outliers overestimated the correlation coefficients. To avoid this, outliers were deleted, defined as utilities less than Q3-(Q1-Q3)*1.5. The utilities were calculated using the Japanese EQ-5D and AP-7D values estimated in this study.

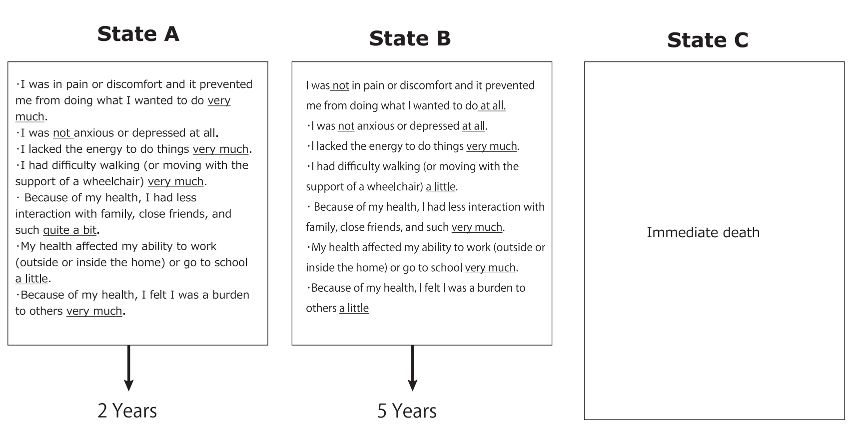

Fig. 1. Screenshot of a triplet DCE task.

2.5. Statistical analysis

Demographics and time to complete DCE tasks were summarized using descriptive statistics. The demographic data characteristics confirmed whether quota sampling was successful.

We reviewed the sequence of DCE responses and excluded respondents whose responses followed these patterns from the analysis set: (1) all first responses to the triplet DCEs were the same (A or B or immediate death), (2) responses followed certain patterns (e.g., A-B-A-B-A-B-A…), and (3) all second responses were the same (A or B) (exclusion criteria 1). Moreover, if the time to complete all tasks was shorter than 187.5 s (averaging 12.5 s per triplet DCE task based on Roudijk et al. (2024)), these respondents were also excluded (exclusion criterion 2). Sensitivity analysis of the time was performed (no exclusion, 5 s, 8 s, 15 s).

Respondents' choices were analyzed based on random utility theory, in which U_{ij} (the utility respondent i derives from choosing health state j) is divided into an explainable component (V_{ij}) and a random error (ε_{ij}).

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

$$V_{ij} = \alpha \text{ x time} + \left(\beta_{12}X_{12j} + \beta_{13}X_{13j} + \beta_{14}X_{14j} + \beta_{22}X_{22j} + \ldots + \beta_{74}X_{74j}\right) \times \text{time} \quad (1)$$

where α is the utility per year obtained in proportion to the expected life year, and β_{pq} is the disutility of the qth ($2 \leq q \leq 4$) level of the pth item (p = 1 PD, 2 MH, 3 EN, 4 MO, 5 WS, 6 II, 7 BD) referring to the first level of the same item. Norman et al. (2016) suggested some methods to anchor utility scores. We applied one of them, which uses the duration of death as 0 to be consistent with the definition of utility (dead = 0) based on the standard cTTO (Robinson and Spencer, 2006; Devlin et al., 2011).

If the error term (ε_{ij}) follows a Gumbel distribution, the probability of choosing health states k by respondent i in the first triplet task is

$$P_{ik} = \frac{\exp\left(V_{ik}\right)}{\sum_{j=1}^{3} \exp\left(V_{ij}\right)}$$

Our primary analysis was based on the mixed logit model that can consider the heterogeneity of respondents' preference, where β_p^m and β_p^s are the mean and scale parameters, respectively, for the random coefficient β_p:

$$\beta_p = \beta_p^m + \beta_p^s . \eta,$$

where η is a stochastic compartment, which assumed a multivariate normal distribution in our analysis. We also used fixed normal and panel logit models with no random effects. We considered interactions with any level-4 responses by adding the N4 term (N4 = 1, if the health states included any level-4 responses) to the conditional logit model. We also defined the N34 term similarly (N34 = 1, if the health states included any level-3 or level-4 responses). If we observed non-monotonicity (i.e., the estimated coefficient of the lower level was larger than that of the higher level), we combined the two levels and re-estimated the model.

The utility measured by PBMs must be anchored to 1 = full health and 0 = dead to calculate QALY (Bansback et al., 2012). We assumed that immediate death has a utility of zero. To anchor the estimated coefficients ($\widehat{\beta}_{pq}$) to the QALY scale (β_{pq_QALY}), the following equation was used for the conversion. $\widehat{\alpha}$ indicated the estimated coefficient of duration in equation (1).

$$\widehat{\beta}_{pq\_QALY} = \widehat{\beta}_{pq} \Big/ \widehat{\alpha}$$

Parameters based on the mixed logit model were estimated using the Markov chain Monte Carlo method via BCHOICE procedure in SAS 9.4 (SAS Institute Inc, 2024), a simulation-based calculation in which estimators depend on the seed value. To remove seed value effects, we repeated the estimation 10 times using different, randomly generated seed values and applied the averaged coefficients $\left(\widehat{\beta}_{pq\_QALY}\right)$ to an AP-7D scoring algorithm. The SAS PHREG procedure was also applied to the DCE data to fit the fixed effect models.

3. Results

3.1. Demographic factors

We conducted the web-based survey and data collection in October 2024. Fig. 2 shows the respondent flowchart. A total of 5458 candidates were invited to take this survey via e-mail, of whom 4973 started the DCE tasks, and a final group of 3401 respondents completed all tasks and background data input. Table 1 presents the demographic factors of the 3401 respondents. As of October 2023, according to Japanese general population norms, the male–female ratio was 1:1, with individuals aged 20–29 accounting for 13.8 % of the population, individuals aged 30–39 accounting for 14.6 %, individuals aged 40–49 accounting for 18.3 %, individuals aged 50–59 accounting for 19.5 %, individuals aged 60–69 accounting for 16.1 %, and individuals aged 70–79 accounting for 17.7 % (Ministry of Health Labour and Welfare, 2023). These statistics indicated that quota sampling was successful. Because individuals older

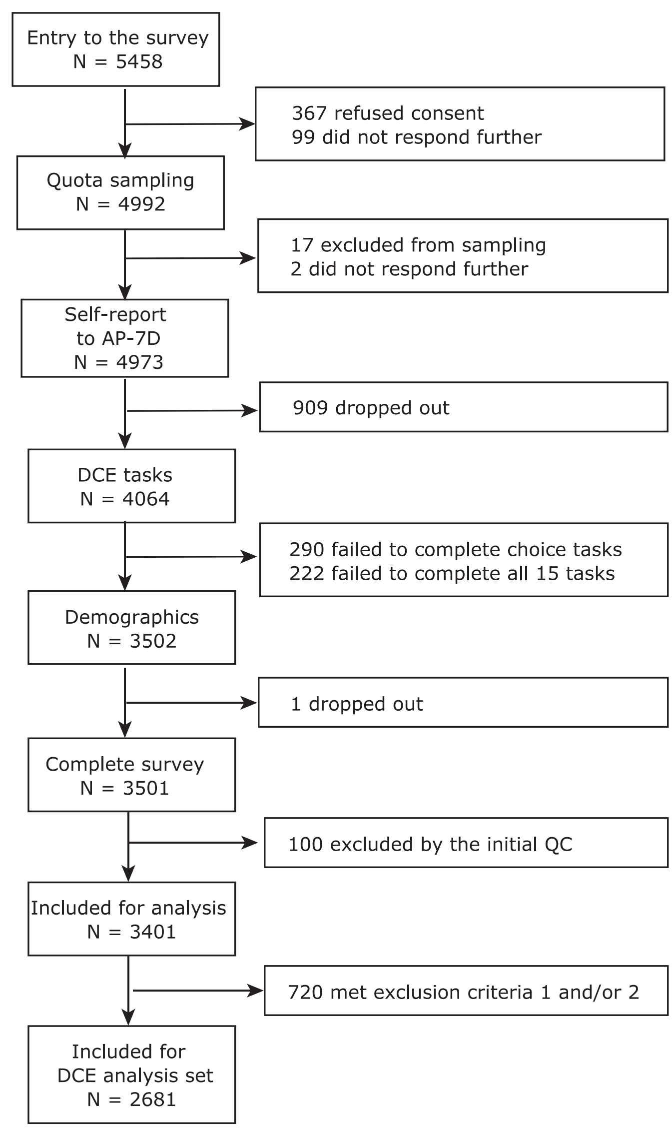

Fig. 2. Respondent flowchart.

Table 1
Demographic factors of respondents.

	All population (N = 3401)		DCE analysis set (N = 2681)	
	Number	Percentage	Number	Percentage
Sex				
Male	1695	49.7 %	1261	47.0 %
Female	1715	50.3 %	1420	53.0 %
Age				
20-29	435	12.8 %	267	10.0 %
30-39	481	14.1 %	335	12.5 %
40-49	626	18.4 %	453	16.9 %
50-59	682	20.0 %	549	20.5 %
60-69	571	16.7 %	498	18.6 %
70-79	615	18.0 %	579	21.6 %
Employment				
Full-time worker	1471	43.1 %	1034	38.6 %
Part-time worker	546	16.0 %	442	16.5 %
Self employed	172	5.0 %	153	5.7 %
Housemaker	617	18.1 %	536	20.0 %
Retired	481	14.1 %	425	15.9 %
Student	69	2.0 %	46	1.7 %
Others	54	1.6 %	45	1.7 %
Education				
Elementary or Junior high school	89	2.6 %	62	2.3 %
High school	1076	31.6 %	870	32.5 %
College	749	22.0 %	593	22.1 %
University	1353	39.7 %	1046	39.0 %
Postgraduate	139	4.1 %	106	4.0 %
Others	4	0.1 %	4	0.2 %
Marital status				
Unmarried	1098	32.2 %	770	28.7 %
Married	1982	58.1 %	1633	60.9 %
Divorced/Bereaved	330	9.7 %	278	10.4 %
Household income (JPY 1mil)				
<1	142	4.2 %	97	3.6 %
1≤ <3	554	16.3 %	441	16.5 %
3≤ <5	757	22.2 %	616	23.0 %
5≤ <7	544	16.0 %	437	16.3 %
7≤ <10	427	12.5 %	345	12.9 %
10≤ <15	229	6.7 %	184	6.9 %
15≤ <20	50	1.5 %	41	1.5 %
20≥	36	1.1 %	21	0.8 %
Unknown	671	19.7 %	499	18.6 %

than age 79 were not included in our sample, a mere comparison of other demographics with Japanese norms, which include data on this population, was difficult. Nevertheless, the median household income was JPY 5.4 million, which was consistent with that of our respondents.

Among the respondents, 346 met exclusion criteria 1, while 485 met exclusion criteria 2, for a total of 720 excluded respondents, and 2681 were included in the DCE analysis. Exclusion rates for men and younger respondents tended to be higher than for other participants. The median time to complete 15 DCE triplet tasks for the 2681 respondents was 606 s (Q1–Q3: 382–1084 s); that is, participants with that median time took 40.4 s on average to complete one triplet DCE task. The median time for the first DCE task was 80 s, with the time monotonically decreasing as DCE tasks were repeated, and 25 s for the final DCE tasks (Fig. 3). Table 2 shows respondents' feedback on the DCE tasks. They reported subjectively experiencing greater difficulty in answering this survey than others (most of which are simple marketing surveys).

3.2. Mixed and fixed conditional logit model results

Table 3 shows the results for the mixed and fixed conditional logit model. The base-case mixed logit model (model 1) had one non-monotonicity (i.e., levels 3 and 4 of AR are not logically consistent, whereas AR worsens as utility increases), and the results for the constraint mixed logit model combining these two levels were also shown (model 2). The results of sensitivity analysis of exclusion time are shown in Online Table 4. If the N4 and N34 terms were also considered

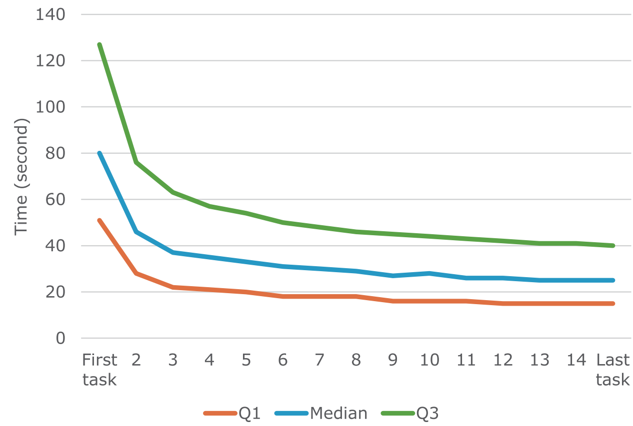

Fig. 3. Time to complete each DCE task.

Table 2
Respondents' feedback on discrete choice experiment (DCE) tasks.

(a) Difficulty compared with previously experienced surveys		
	Number	Percentage
Easier	175	6.5 %
The same	524	19.5 %
Harder	1688	63.0 %
Do not know	294	11.0 %
(b) Difficulty understanding health states		
	Number	Percentage
Very difficult	178	6.6 %
Difficult	1065	39.7 %
Neutral	873	32.6 %
Not very difficult	517	19.3 %
Not difficult at all	48	1.8 %
(c) Difficulty choosing one health state by DCE tasks		
	Number	Percentage
Very difficult	293	10.9 %
Difficult	1394	52.0 %
Neutral	689	25.7 %
Not very difficult	265	9.9 %
Not difficult at all	40	1.5 %

(models 3 and 4), the N4 and N34 coefficients were positive and too large to interpret meaningfully. The results for the normal and panel fixed effect models were also presented (models 5 and 6), which had more non-monotonicities than the mixed effects models.

3.3. AP-7D scoring algorithm

Table 5 shows the scoring algorithm based on model 2. AP-7D values can be calculated by 1+ (adding seven negative coefficients) in the table. The AP-7D value of the worst health state [4444444] was −0.448 (although the values cannot be simply compared, EQ-5D-5L: −0.025 (Shiroiwa et al., 2016) and SF-6Dv2: −0.772 (Shiroiwa et al., 2025) according to the Japanese value sets). Fig. 4 shows decrements of each item: those for PD, MO, and BO were the largest, while EN, WS, and II had smaller influences on AP-7D values. Fig. 5 illustrates the distribution of utilities for all AP-7D health states compared with the EQ-5D-5L and SF-6Dv2 value sets in Japan. This can be attributed to the complex interaction between differences in number of health states described, the dimensions included, and the valuation methods used (composite time-trade off [cTTO] and DCE).

Table 3
Results by the conditional logit model.

		Mixed effects model								Fixed effects model			
		Model 1 (unconstrained)		Model 2 (Constrained)		Model 3 (unconstrained, N4)		Model4 (unconstrained, N34)		Model5 (unconstrained)		Model6 (panel logit)	
Effect		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Time		1.265	0.021	1.214	0.029	1.279	0.026	1.066	0.022	0.369	0.007	0.225	0.007
PD x Time	2	−0.113	0.014	−0.114	0.010	−0.081	0.012	−0.102	0.013	−0.018	0.003	0.014	0.003
	3	−0.332	0.011	−0.319	0.010	−0.287	0.010	−0.297	0.010	−0.065	0.003	−0.026	0.003
	4	−0.413	0.013	−0.430	0.014	−0.369	0.010	−0.363	0.011	−0.097	0.003	−0.022	0.003
AD x Time	2	−0.105	0.011	−0.109	0.012	−0.093	0.011	−0.062	0.012	−0.044	0.004	−0.014	0.004
	3	−0.209	0.010	−0.198	0.011	−0.228	0.011	−0.209	0.010	−0.064	0.003	−0.019	0.003
	4	−0.258	0.012	−0.280	0.011	−0.269	0.011	−0.211	0.013	−0.088	0.004	−0.030	0.004
EN x Time	2	−0.020	0.011	−0.013	0.011	−0.036	0.010	−0.020	0.011	0.007	0.004	0.004	0.004
	3	−0.079	0.012	−0.080	0.012	−0.084	0.013	−0.113	0.011	−0.003	0.003	−0.028	0.003
	4	−0.125	0.012	−0.134	0.011	−0.156	0.012	−0.111	0.012	−0.020	0.003	−0.028	0.003
MO x Time	2	−0.147	0.013	−0.144	0.011	−0.165	0.010	−0.112	0.011	−0.040	0.003	−0.029	0.003
	3	−0.339	0.012	−0.326	0.013	−0.333	0.011	−0.296	0.011	−0.086	0.003	−0.076	0.003
	4	−0.353	0.010	−0.339	0.012	−0.383	0.013	−0.308	0.016	−0.088	0.003	−0.086	0.003
WS x Time	2	−0.004	0.012	−0.005	0.010	−0.027	0.013	−0.037	0.009	0.003	0.003	−0.023	0.003
	3	−0.059	0.011	−0.027	0.011	−0.057	0.012	−0.087	0.010	−0.016	0.004	−0.050	0.004
	4	−0.136	0.012	−0.110	0.013	−0.183	0.013	−0.155	0.010	−0.023	0.003	−0.046	0.003
II x Time	2	−0.049	0.011	−0.030	0.013	−0.052	0.011	−0.017	0.012	−0.030	0.003	−0.021	0.003
	3	−0.162	0.011	−0.154	0.009	−0.185	0.011	−0.138	0.011	−0.050	0.003	−0.031	0.003
	4	−0.152	0.011			−0.214	0.011	−0.145	0.011	−0.051	0.004	−0.051	0.004
BO x Time	2	−0.028	0.012	−0.043	0.013	−0.047	0.012	−0.044	0.014	−0.015	0.004	−0.007	0.004
	3	−0.297	0.012	−0.288	0.014	−0.320	0.011	−0.300	0.010	−0.064	0.003	−0.072	0.003
	4	−0.349	0.012	−0.337	0.011	−0.380	0.010	−0.356	0.011	−0.081	0.003	−0.077	0.003
N4						0.947	0.026						
N34								1.783	0.059				
Number of inconsistencies		1		0		0		1		2		4	
Number of observations		201,075		201,075		201,075		201,075		201,075		201,075	
DIC		91,282.1		90,768.7		88,401.8		84,556.2		NA		NA	
AIC		NA		NA		NA		NA		131971.5		125722.1	

Table 4
Results for sensitivity analysis about exclusion time.

Exclusion time		Mixed effects model (Model1)									
		All respondents		<5 s		<8 s		<12.5 s		<15 s	
Effect		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Time		1.124	0.018	1.179	0.022	1.142	0.019	1.265	0.021	1.273	0.022
PD x Time	2	−0.115	0.010	−0.115	0.011	−0.120	0.010	−0.113	0.014	−0.128	0.010
	3	−0.285	0.010	−0.289	0.010	−0.296	0.011	−0.332	0.011	−0.328	0.013
	4	−0.363	0.013	−0.364	0.011	−0.366	0.012	−0.413	0.013	−0.436	0.013
AD x Time	2	−0.069	0.012	−0.080	0.012	−0.101	0.012	−0.105	0.011	−0.097	0.012
	3	−0.166	0.010	−0.176	0.009	−0.197	0.010	−0.209	0.010	−0.206	0.011
	4	−0.234	0.011	−0.236	0.009	−0.241	0.010	−0.258	0.012	−0.285	0.014
EN x Time	2	0.010	0.010	−0.015	0.010	−0.011	0.011	−0.020	0.011	−0.032	0.012
	3	−0.064	0.009	−0.067	0.012	−0.071	0.011	−0.079	0.012	−0.074	0.011
	4	−0.110	0.011	−0.114	0.012	−0.122	0.011	−0.125	0.012	−0.144	0.010
MO x Time	2	−0.121	0.009	−0.116	0.009	−0.126	0.009	−0.147	0.013	−0.179	0.011
	3	−0.301	0.010	−0.279	0.011	−0.289	0.011	−0.339	0.012	−0.353	0.010
	4	−0.294	0.010	−0.282	0.010	−0.298	0.012	−0.353	0.010	−0.365	0.013
WS x Time	2	0.011	0.011	0.003	0.010	0.004	0.009	−0.004	0.012	0.007	0.010
	3	−0.026	0.012	−0.017	0.011	−0.020	0.010	−0.059	0.011	−0.030	0.010
	4	−0.124	0.013	−0.124	0.013	−0.121	0.011	−0.136	0.012	−0.120	0.012
SR x Time	2	−0.012	0.011	−0.028	0.010	−0.017	0.011	−0.049	0.011	−0.013	0.010
	3	−0.151	0.009	−0.133	0.010	−0.135	0.012	−0.162	0.011	−0.169	0.011
	4	−0.129	0.010	−0.119	0.010	−0.124	0.012	−0.152	0.011	−0.159	0.010
BO x Time	2	−0.031	0.012	−0.040	0.011	−0.032	0.013	−0.028	0.012	−0.051	0.017
	3	−0.229	0.010	−0.255	0.011	−0.250	0.009	−0.297	0.012	−0.311	0.012
	4	−0.312	0.010	−0.329	0.011	−0.312	0.009	−0.349	0.012	−0.373	0.010
Number of inconsistencies		4		2		2		1		2	
Number of observations		230,550		230,400		222,000		201,075		190,425	
DIC		108,205.8		107,318.5		103,034.7		91,282.1		86,491.1	

3.4. Comparison of responses to EQ-5D-5L and AP-7D

Suzuki. et al. (2025) collected data from 528 Japanese respondents, and 21 outliers (4.0 %), as defined in section 2.4, were deleted. Table 6 presents the correlation matrix between the EQ-5D-5L and AP-7D responses. The table indicates concepts overlapping with the EQ-5D-5L, and many items have small correlations between the two. There was no issue with some concepts between the AP-7D and EQ-5D-5L, which are important regardless of cultural differences (mobility, pain, and mental health), because we measured the same concepts. Excluding these items,

Table 5
Japanese scoring algorithm of AP-7D.

Item	Level	Coefficient
PD	2	−0.089
	3	−0.259
	4	−0.333
AD	2	−0.076
	3	−0.157
	4	−0.220
EN	2	−0.014
	3	−0.062
	4	−0.102
MO	2	−0.119
	3	−0.264
	4	−0.277
WS	2	−0.003
	3	−0.024
	4	−0.103
II	2	−0.024
	3	−0.130
	4	−0.130
BO	2	−0.037
	3	−0.235
	4	−0.283

The worst score: −0.448.
The second-best score: 0.997.

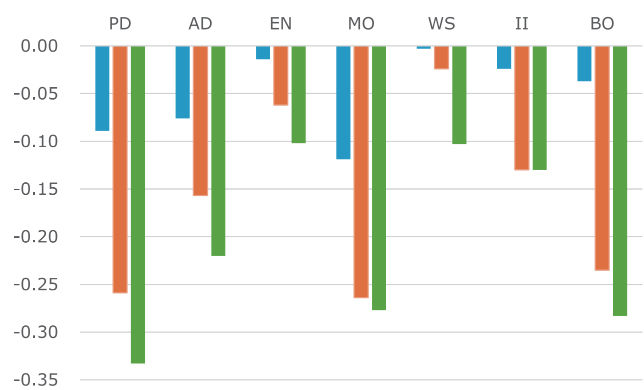

Fig. 4. Decrements of each item PD: pain/discomfort, MH: mental health, EN: energy, MO: mobility, WS: work/school, II: interpersonal interactions, BO: burden to others.

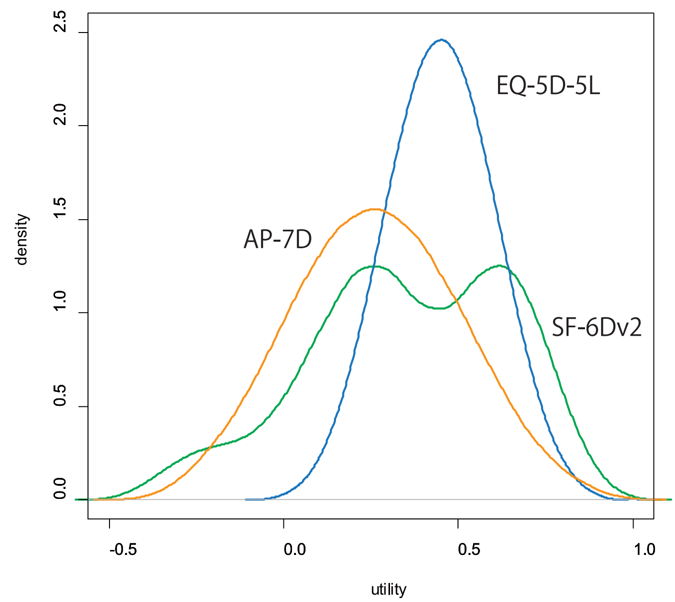

Fig. 5. Value distribution for AP-7D.

Table 6
Correlation coefficients between EQ-5D-5L and AP-7D.

AP-7D	EQ-5D-5L				
	Mobility	Self-care	Usual activities	Pain/ discomfort	Anxiety/ depression
Pain/ discomfort	0.009	0.077	0.131	0.578	0.188
Anxiety/ depression	0.011	0.133	0.120	0.136	0.495
Energy	0.109	−0.004	0.173	0.197	0.368
Mobility	0.411	−0.013	0.118	0.159	−0.009
Work/ school	0.069	−0.022	0.275	0.328	0.171
Social relations	0.036	−0.017	0.187	0.131	0.240
Burden to others	0.139	−0.020	0.104	0.245	0.175

the correlation coefficients for the following are 0.3–0.4 (week or moderate correlation): (a) “energy” (AP-7D) and “anxiety/depression” (EQ-5D-5L) as well as (b) “work/school” (AP-7D) and “pain/discomfort” (EQ-5D-5L).

4. Discussion

4.1. The limitation from a pluralistic viewpoint

We established the first value set of AP-7D. According to Table 5, there are small correlations between many EQ-5D-5L and AP-7D items. This suggests that AP-7D captures different aspects of health to the EQ-5D-5L. Although this may not be definitive evidence, but it supports a health pluralistic view and the role of this new instrument. However, in this study, we applied the standard valuation method commonly used for other instruments. However, debate is needed over whether conventional approaches are suitable for capturing Asian people's preferences. For example, in Japan, the worst EQ-5D-5L score (Shiroiwa et al., 2016) based on the TTO method is much higher than in other countries, whereas the SF-6Dv2 (Shiroiwa et al., 2025) and/or EORTC QLQ-C10D (Shiroiwa et al., 2024a) scores based on DCE with duration are lower. This discrepancy may partly reflect artifacts such as survey mode differences (e.g. face-to-face or web-based and/or methods of quality control), but it could also reflect biases inherent in the valuation methods. “Death (or dead)” is a sensitive topic everywhere, and attitudes toward it are strongly influenced by cultural and religious contexts. Thus, it is no wonder that risk preferences regarding the trade-off between health states and life expectancy vary across cultures. Additionally, there was some discussion about whether “dead” should be anchored to utility = 0 (Sampson et al., 2024). Our analysis used the assumption that the utility of “dead” is zero. If this restriction is removed, it will affect the value set and its international comparisons. At present, no clear alternative method exists for valuing health states more appropriately for Asian populations. Instead, we selected triplet DCE rather than the more common pairwise DCE, taking into account the considerations discussed in the Methods section. This remains an important issue for future research.

The mode (face-to-face or web-based surveys) of the valuation survey may need to be reconsidered. While the equivalent of face-to-face and web-based surveys is observed in some countries (Jiang et al., 2021, 2023), it has not been established in others. Of course, differences in nationality influence the most suitable mode for valuation surveys. In the first place, the AP-7D was developed based on face-to-face interviews. However, it is unclear whether the face-to-face interview format is appropriate for extracting health concepts. Some Asian people find it difficult to express themselves and are reluctant to insist strongly on their ideas and feelings. It is possible that we should have paid more attention to psychological safety when developing our new instrument. In Australia, for example, "yarning circles" (Kennedy et al., 2022) are sometimes used to conduct as focus group interview for First Nations Australian (Aboriginal and Torres Strait Islander) groups. Currently, we cannot decide which methods are appropriate because of limited empirical evidence; however, these methods suggest a way to collect the voices of non-Western people.

Some critics argue that our instrument does not adequately capture key characteristics of Asian populations, because East and Southeast Asia are too heterogeneous to be treated as a single cultural environment. This critic may be valid. Although our current instrument does not satisfy everyone, we believe that our pluralistic approach is a novel and important attempt to reconsider health universalism and inform the future development of PBMs. Our instrument was developed based on qualitative analyses in Asian countries; therefore, we call it a "culture-specific" PBM. Conversely, it may be better for health pluralists to regard each PBM as a culture-specific PBM. We believe that conventional PBMs will continue to play an important role in the future, but we must also avoid bringing invisible Western centralism to the measurement of utilities.

As is well known, utility values derived from the same health states differ depending on the instrument used. It may cause concerns that the use of culture-specific PBMs can cause practical difficulties compared to measured utilities. It is also difficult to determine whether the difference in utility value is due to cultural differences, the instruments used, or both. Instruments, for which there is no accumulated usage experience, cause problems when it comes to comparing measured utility with that of other health states, interpreting the value, and considering its validity. Using instruments that are not well-established may result in poor-quality data. In particular, for HTA-related decision-makers, we assume it is problematic that the utility value for economic evaluation varies among cultural groups in addition to other problems. How these instruments constructed by different concepts should be used differently is not easy to answer. In practice, the selection of PBMs may require a particularistic approach[2] rather than a generalistic one, if we refer again to the words of bioethics. We need a case-by-case selection of instruments depending on the situation, instead of a consistent principal.

4.2. Interpretation and limitations of our empirical survey

Next, we need to return to a discussion about the empirical results of the valuation surveys. This is the first study to report the value set for AP-7D, a culture-specific PBM for East and Southeast Asia. We constructed the Japanese value set based on the parameters estimated using the mixed logit model. Only one inconsistency was observed (levels 3 and 4 of WS) in model 1. Having a small number of logical inconsistencies is important for PBM value sets, which can be used for QALY calculation. To construct an AP-7D value set for Japan, we selected model 2 (constrained model), as its inconsistency count is limited, whereas model 1 is the simplest model (user-friendly), and mixed effects models are theoretically superior to fixed effects models, as they account for the heterogeneity of health preferences.

Reflecting non-linear discount rates (Jonker et al., 2018) to value sets is a future issue, which seems theoretically valid but is not yet a standard valuation method. No other valuation survey uses a non-linear discount rate. We need to pay deliberate attention to other empirical studies and ongoing discussions.

When the immediate death card is included in the choice task based on the split-triplet design, Pullenayegum et al. (2025) recommend anchoring using duration and/or non-linear discount to obtain a consistent value set by cTTO. However, our results do not need to be consistent with those of cTTO. Practically, our value set shows that the worst score was much less than zero at −0.448, which is different from the findings of Pullenayegum et al. In their case, their worst score is much higher than zero, if the anchoring immediate death and linear discounting method is used. In addition, the results show an estimated discount rate of 23.4 %, which is higher than the results of another empirical study (West et al., 2003), although the comparison has significant uncertainty. If the discount rate was overestimated, estimators by non-linear modeling may also have biased. Further discussion is needed on this issue.

As described in the Results section, PD, MO, and BO exerted the greatest influence on the AP-7D values. The strong association of pain and mobility (or physical function) items with the utility of health states is consistent with other PBMs. Notably, BO had one of the largest decrements on the AP-7D values, comparable with PD and MO, which may reflect values prevalent in Japanese society (and some other Asian societies) that emphasize family and community over the individual, and consensus and harmony over conflict.

Our valuation survey was based on an established international protocol, enhancing the international comparability of AP-7D value sets under development for other Asian countries, which is a strength of our study. Meanwhile, a limitation of our survey was its sampling method, which did not follow a rigid random sampling from the Japanese general population. Although the distribution of respondents' backgrounds was close to Japanese population norms, unknown characteristics of people registered to the web panel may have influenced our results. As shown in the Results section, a greater number of men and younger respondents were excluded. This might have led to biased results.

We assumed multiplicative utility function to estimate value sets. However, an empirical study does not support this function (Jonker and Norman, 2022). The study showed that most respondents did not use the multiplicative function. If our assumption is inappropriate, our estimators may be biased, which is also a limitation of our results. Respondents also reported subjectively experiencing a heavy cognitive burden while performing DCE tasks, although this was the same situation for the web-based valuation surveys of other PBMs. It was difficult to estimate how cognitive burden might have influenced the participants' responses. Finally, although the value set had only one inconsistency, it was nevertheless observable, which is not desirable for the value set and should be avoided if possible.

We summarized the results of our empirical survey, concluding that it had produced the first value set for AP-7D in accordance with the international protocol. Now we can convert responses to AP-7D value for QALY calculation. The next step of our study involves establishing value sets in other countries and conducting international comparisons of them. This may also contribute to clarifying the difference in preferences among countries and the necessity of culture-specific PBMs. This is a new attempt to consider and capture cultural differences in utility measurement.

Ethics committee

This study was performed in line with the principles of the Declaration of Helsinki. The ethics committee of the National Institute of Public Health, to which the first author belongs (NIPH-IBRA#12264) approved. Informed consent was obtained from all individual participants included in the study. Participants signed informed consent regarding publishing their anonymous data.

[2] The word "particularism" is used in moral ethics. According to Dancy (2017), "Moral Particularism, at its most trenchant, is the claim that there are no defensible moral principles, that moral thought does not consist in the application of moral principles to cases, and that the morally perfect person should not be conceived as the person of principle." (DANCY, J. 2017. Moral Particularism. *In:* EDWARD, N. Z. (ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2017 ed, ibid.

Funding

This study funded by National Institute of Public Health.

CRediT authorship contribution statement

Takeru Shiroiwa: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft. **Yasuhiro Morii:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Eri Hoshino:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Tatsunori Murata:** Formal analysis, Investigation, Writing – review & editing. **Richard Norman:** Methodology, Supervision, Writing – review & editing. **Brendan Mulhern:** Methodology, Supervision, Writing – review & editing. **Nan Luo:** Methodology, Supervision, Writing – review & editing. **Donna Rowen:** Methodology, Supervision, Writing – review & editing. **Takashi Fukuda:** Conceptualization, Supervision, Writing – review & editing.

Declaration of interest statement

The authors have no conflicts of interests to declare that are relevant to the content of this article.

Acknowledgment

We would like to thank the members of HTAsiaLink for joining the project and for their support.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

Bahrampour, M., Byrnes, J., Norman, R., Scuffham, P.A., Downes, M., 2020. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. Eur. J. Health Econ. 21, 983–992.
Bansback, N., Brazier, J., Tsuchiya, A., Anis, A., 2012. Using a discrete choice experiment to estimate health state utility values. J. Health Econ. 31, 306–318.
Beauchamp, T.L., Childress, J.F., 2019. Principles of Biomedical Ethics. Oxford University Press, New York.
Brazier, J., Peasgood, T., Mukuria, C., Marten, O., Kreimeier, S., Luo, N., Mulhern, B., Pickard, A.S., Augustovski, F., Greiner, W., Engel, L., Belizan, M., Yang, Z., Monteiro, A., Kuharic, M., Gibbons, L., Ludwig, K., Carlton, J., Connell, J., Rand, S., Devlin, N., Jones, K., Tsuchiya, A., Lovett, R., Naidoo, B., Rowen, D., Rejon-Parrilla, J.C., 2022. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. Value Health 25, 482–491.
Brazier, J., Usherwood, T., Harper, R., Thomas, K., 1998. Deriving a preference-based single index from the UK SF-36 Health Survey. J. Clin. Epidemiol. 51, 1115–1128.
Brazier, J.E., Mulhern, B.J., Bjorner, J.B., Gandek, B., Rowen, D., Alonso, J., Vilagut, G., Ware, J.E., 2020. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. Med. Care 58, 557–565.
Cheung, Y.B., Thumboo, J., 2006. Developing health-related quality-of-life instruments for use in Asia: the issues. Pharmacoeconomics 24, 643–650.
Coast, J., Flynn, T.N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J.J., Peters, T.J., 2008. Valuing the ICECAP capability index for older people. Soc. Sci. Med. 67, 874–882.
Dancy, J., 2017. Moral particularism. In: Edward, N.Z. (Ed.), The Stanford Encyclopedia of Philosophy. Winter 2017.
Devlin, N.J., Tsuchiya, A., Buckingham, K., Tilling, C., 2011. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Econ. 20, 348–361.
EUROQOL Group, 1990. EuroQol–a new facility for the measurement of health-related quality of life. Health Policy 16, 199–208.
Fayers, P.M., Machin, D., 2016. Quality of Life: the Assessment, Analysis and Interpretation of Patient-Reported Outcomes. Wiley-Blackwell, Chichester.
Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C.H., Zhu, Z., Depauw, S., Denton, M., Boyle, M., 2002. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med. Care 40, 113–128.
Flynn, T.N., Louviere, J.J., Marley, A.A., Coast, J., Peters, T.J., 2008. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. Popul. Health Metr. 6, 6.
Forder, J.E., Caiels, J., 2011. Measuring the outcomes of long-term care. Soc. Sci. Med. 73, 1766–1774.
Hawthorne, G., Richardson, J., Osborne, R., 1999. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Qual. Life Res. 8, 209–224.
Horsman, J., Furlong, W., Feeny, D., Torrance, G., 2003. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual. Life Outcome 1, 54.
Jiang, R., Pullenayegum, E., Shaw, J.W., MüHlbacher, A., Lee, T.A., Walton, S., Kohlmann, T., Norman, R., Pickard, A.S., 2023. Comparison of preferences and data quality between discrete choice experiments conducted in online and face-to-face respondents. Med. Decis. Mak. 43, 667–679.
Jiang, R., Shaw, J., MüHlbacher, A., Lee, T.A., Walton, S., Kohlmann, T., Norman, R., Pickard, A.S., 2021. Comparison of online and face-to-face valuation of the EQ-5D-5L using composite time trade-off. Qual. Life Res. 30, 1433–1444.
Jonker, M.F., Donkers, B., De Bekker-Grob, E., Stolk, E.A., 2019. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health Econ. 28, 350–363.
Jonker, M.F., Donkers, B., De Bekker-Grob, E.W., Stolk, E.A., 2018. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. Value Health 21, 993–1001.
Jonker, M.F., Norman, R., 2022. Not all respondents use a multiplicative utility function in choice experiments for health state valuations, which should be reflected in the elicitation format (or statistical analysis). Health Econ. 31, 431–439.
Kennedy, M., Maddox, R., Booth, K., Maidment, S., Chamberlain, C., Bessarab, D., 2022. Decolonising qualitative research with respectful, reciprocal, and responsible research practice: a narrative review of the application of Yarning method in qualitative Aboriginal and Torres Strait Islander health research. Int. J. Equity Health 21, 134.
Kind, P., Brook, B., Rabin, R., 2005. EQ-5D Concepts and Methods: a Developmental History. Springer, Dordrecht.
King, M.T., Costa, D.S., Aaronson, N.K., Brazier, J.E., Cella, D.F., Fayers, P.M., Grimison, P., Janda, M., Kemmler, G., Norman, R., Pickard, A.S., Rowen, D., Velikova, G., Young, T.A., Viney, R., 2016. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Qual. Life Res. 25, 625–636.
King, M.T., Norman, R., Mercieca-Bebber, R., Costa, D.S.J., Mctaggart-Cowan, H., Peacock, S., Janda, M., MüLler, F., Viney, R., Pickard, A.S., Cella, D., 2021. The functional assessment of cancer therapy eight dimension (FACT-8D), a multi-attribute utility instrument derived from the cancer-specific FACT-general (FACT-G) quality of life questionnaire: development and Australian value set. Value Health 24, 862–873.
King, M.T., Viney, R., Simon Pickard, A., Rowen, D., Aaronson, N.K., Brazier, J.E., Cella, D., Costa, D.S.J., Fayers, P.M., Kemmler, G., Mctaggart-Cowen, H., Mercieca-Bebber, R., Peacock, S., Street, D.J., Young, T.A., Norman, R., 2018. Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. Pharmacoeconomics 36, 225–238.
Kraut, R., 2022. Aristotle's ethics. In: Edward, N.Z., URI, N. (Eds.), The Stanford Encyclopedia of Philosophy. Fall 2022.
Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 26, 661–677.
Mchorney, C.A., Ware Jr., J.E., Raczek, A.E., 1993. The MOS 36-Item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med. Care 31, 247–263.
Ministry of Health Labour and Welfare, 2023. Vital Statistics.
Mulhern, B.J., Bansback, N., Norman, R., Brazier, J., 2020. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. Med. Care 58, 566–573.
Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.M., Brazier, J., Flynn, T., Forder, J., Wall, B., 2012. Outcomes of social care for adults: developing a preference-weighted measure. Health Technol. Assess. 16, 1–166.
Norman, R., Mulhern, B., Viney, R., 2016. The impact of different DCE-based approaches when anchoring utility scores. Pharmacoeconomics 34, 805–814.
Perkins, M.R., Devlin, N.J., Hansen, P., 2004. The validity and reliability of EQ-5D health state valuations in a survey of Māori. Qual. Life Res. 13, 271–274.
Ribeiro Santiago, P.H., Haag, D., Macedo, D.M., Garvey, G., Smith, M., Canfell, K., Hedges, J., Jamieson, L., 2021. Psychometric properties of the EQ-5D-5L for aboriginal Australians: a multi-method study. Health Qual. Life Outcome 19, 81.
Robinson, A., Spencer, A., 2006. Exploring challenges to TTO utilities: valuing states worse than dead. Health Econ. 15, 393–402.
Roudijk, B., Jonker, M.F., Bailey, H., Pullenayegum, E., 2024. A direct comparison between discrete choice with duration and composite time trade-off methods: do they produce similar results? Value Health 27, 1280–1288.
Rowen, D., Brazier, J., Ara, R., Azzabi Zouraq, I., 2017. The role of condition-specific preference-based measures in health technology assessment. Pharmacoeconomics 35, 33–41.
Sampson, C., Parkin, D., Devlin, N., 2024. Is anchoring at 'dead' a theoretical requirement for health state valuation? Health Econ. 33, 1929–1935.
SAS Institute Inc, 2024. SAS/STAT® 13.1 User's Guide the BCHOICE Procedure. SAS Institute Inc, Cary, NC.
Shiroiwa, T., Ikeda, S., Noto, S., Igarashi, A., Fukuda, T., Saito, S., Shimozuma, K., 2016. Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. Value Health 19, 648–654.
Shiroiwa, T., King, M.T., Norman, R., MüLler, F., Campbell, R., Kemmler, G., Murata, T., Shimozuma, K., Fukuda, T., 2024a. Japanese value set for the EORTC QLU-C10D: a

multi-attribute utility instrument based on the EORTC QLQ-C30 cancer-specific quality-of-life questionnaire. Qual. Life Res. 33, 1865–1879.
Shiroiwa, T., Murata, T., Ahn, J., Li, X., Nakamura, R., Teerawattananon, Y., Kun, Z., Shafie, A.A., Valverde, H., Lam, H., Ng, K., Nadjib, M., Pwu, R.F., Nugraha, R.R., Chen, Y.C., Fukuda, T., 2022. Developing a new region-specific preference-based measure in east and Southeast Asia. Value Health Reg Issues 32, 62–69.
Shiroiwa, T., Murata, T., Morii, Y., Hoshino, E., Fukuda, T., 2024b. Comparison of four value sets derived using different TTO and DCE approaches: application to the new region-specific PBM, AP-7D. Health Qual. Life Outcome 22, 16.
Shiroiwa, T., Yamamoto, Y., Murata, T., Mulhern, B., Bjorner, J., Brazier, J., Fukuda, T., Rowen, D., Fukuhara, S.I., 2025. Valuation survey for SF-6Dv2 in Japan based on the international protocol. Qual. Life Res. 34, 445–455.
Shiroiwa, T., Fukuda, T., 2024. Are web-based valuation surveys for preference-based measures as reliable as face-to-face surveys? TTO, DCE and DCE with duration. Appl. Health Econ. Health Pol. 22, 391–400.
Sintonen, H., 2001. The 15D instrument of health-related quality of life: properties and applications. Ann. Med. 33, 328–336.
Stevens, K., 2009. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual. Life Res. 18, 1105–1113.
Suzuki, Y., Shiroiwa, T., Murata, T., Kamono, E., Morii, Y., Fukuda, T., 2025. Convergent validity, reliability and responsiveness of the AP-7D and EQ-5D-5L based on the survey for the general people in five countries: a new cultural-specific preference-based measure developed in east and southeast Asia. Health Qual. Life Outcome 23, 61.
Torrance, G.W., Furlong, W., Feeny, D., Boyle, M., 1995. Multi-attribute preference functions. Health utilities index. Pharmacoeconomics 7, 503–520.
Turner, L., 2003. Bioethics in a multicultural world: medicine and morality in pluralistic settings. Health Care Anal. 11, 99–117.
Ware Jr., J.E., Sherbourne, C.D., 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med. Care 30, 473–483.
West, R.R., Mcnabb, R., Thompson, A.G., Sheldon, T.A., Grimley Evans, J., 2003. Estimating implied rates of discount in healthcare decision-making. Health Technol. Assess. 7, 1–60.
Wille, N., Badia, X., Bonsel, G., BurströM, K., Cavrini, G., Devlin, N., Egmar, A.C., Greiner, W., Gusi, N., Herdman, M., Jelsma, J., Kind, P., Scalone, L., Ravens-Sieberer, U., 2010. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual. Life Res. 19, 875–886.
Wu, J., He, X., Chen, P., Xie, S., Li, X., Hu, H., Zhao, K., Xie, F., 2022. China health related outcomes measures (CHROME): development of a new generic preference-based measure for the Chinese population. Pharmacoeconomics 40, 957–969.