BMJ Open

Validation of the predictive accuracy of health-state utility values based on the Lloyd model for metastatic or recurrent breast cancer in Japan

Tsuguo Iwatani ,[1] Eisuke Inoue,[2] Koichiro Tsugawa[3]

To cite: Iwatani T, Inoue E, Tsugawa K. Validation of the predictive accuracy of health-state utility values based on the Lloyd model for metastatic or recurrent breast cancer in Japan. *BMJ Open* 2021;**11**:e046273. doi:10.1136/bmjopen-2020-046273

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-046273).

Received 25 October 2020
Accepted 02 November 2021

[1]Breast Surgery, National Cancer Center-Hospital East, Kashiwa, Japan
[2]Research Administration Center, Showa University, Shinagawa-ku, Tokyo, Japan
[3]Department of Breast and Endocrine Surgery, St Marianna University School of Medicine, Kawasaki, Kanagawa, Japan

Correspondence to
Dr Tsuguo Iwatani;
t.iwatani@marianna-u.ac.jp

ABSTRACT

Introduction Although there is a lack of data on health-state utility values (HSUVs) for calculating quality-adjusted life-years in Japan, cost–utility analysis has been introduced by the Japanese government to inform decision making in the medical field since 2016.

Objectives This study aimed to determine whether the Lloyd model which was a predictive model of HSUVs for metastatic breast cancer (MBC) patients in the UK can accurately predict actual HSUVs for Japanese patients with MBC.

Design The prospective observational study followed by the validation study of the clinical predictive model.

Setting and participants Forty-four Japanese patients with MBC were studied at 336 survey points.

Methods This study consisted of two phases. In the first phase, we constructed a database of clinical data prospectively and HSUVs for Japanese patients with MBC to evaluate the predictive accuracy of HSUVs calculated using the Lloyd model. In the second phase, Bland-Altman analysis was used to determine how accurately predicted HSUVs (based on the Lloyd model) correlated with actual HSUVs obtained using the EuroQol 5-Dimension 5-Level questionnaire, a preference-based measure of HSUVs in patients with MBC.

Results In the Bland-Altman analysis, the mean difference between HSUVs estimated by the Lloyd model and actual HSUVs, or systematic error, was −0.106. The precision was 0.165. The 95% limits of agreement ranged from −0.436 to 0.225. The t value was 4.6972, which was greater than the t value with 2 degrees of freedom at the 5% significance level (p=0.425).

Conclusions There were acceptable degrees of fixed and proportional errors associated with the prediction of HSUVs based on the Lloyd model for Japanese patients with MBC. We recommend that sensitivity analysis be performed when conducting cost-effectiveness analyses with HSUVs calculated using the Lloyd model.

Strengths and limitations of this study

- We investigated whether health-state utility values (HSUVs) calculated using the Lloyd model can accurately predict the actual HSUVs for Japanese patients with metastatic breast cancer (MBC).
- Even in Japan, it is possible to estimate the predictive value of HSUVs from clinical data and perform a cost-effectiveness analysis.
- Our methods can be applied to assess the accuracy of other clinical prediction models as well.
- Our clinical database for validating the HSUVs predictive model, at a single institution, was relatively small (336 survey points of 44 Japanese patients with MBC) and included repeated measurements.
- The Lloyd model can predict an acceptable degree of the actual HSUVs for Japanese patients with MBC; however, we recommend that sensitivity analysis be performed when conducting cost-effectiveness analyses with HSUVs calculated using the Lloyd model for uncertainty of predictive HSUVs data.

INTRODUCTION

Breast cancer is a common malignant disease among Japanese women. In 2017, approximately 91 000 Japanese women had newly diagnosed with breast cancer, and 15 000 Japanese women died from breast cancer.[1]

Although metastatic breast cancer (MBC) is difficult to cure, advances in treatment since the 1990s, especially with the advent of new innovative drugs, have led to gradual improvements in survival after recurrence.[2] While these innovative drugs provide significant benefits to patients with MBC, they are extremely expensive. The budget impact is a key concern in Japan, which has a universal healthcare system.[3] Increasing treatment costs are also a challenge for patients and healthcare providers in the USA and Europe. The US Centres for Medicare and Medicaid Services estimates that spending on healthcare in America will increase from US$3.6 trillion in 2018 to nearly US$6.0 trillion by 2027, with the cost of drugs, including anticancer agents, expected to contribute significantly to this increase.[4]

Against this backdrop, a system was established in Japan since 2016 to reflect the results of a cost–utility analysis using quality-adjusted

life-years (QALYs) as an outcome in determining the effectiveness of high-cost pharmaceuticals and medical devices chosen by the government. These are new decision-making processes for the pricing of health technologies. The cost per QALYs threshold is set at ¥5 million (nearly fifty thousand dollars) per QALYs. This should be determined with reference to certain factors such as the opportunity cost of the healthcare system, the cost of already-reimbursed technologies (eg, haemodialysis cost for end-stage renal disease), incremental cost-effectiveness ratio thresholds and gross domestic product per capita in foreign countries and willingness-to-pay surveys in Japan.[3 5] In 2016, two anticancer agents, including trastuzumab emtansine for the treatment of patients with MBC, were included in a pilot of the Japanese version of health technology assessment.[3 5]

For effective allocation of healthcare resources, it is necessary to evaluate the clinical and cost-effectiveness.[5] Model-based analysis plays an important role in determining the cost effectiveness of chemotherapy agents because it requires a combination of multiple data sources.[6 7] However, clinical researchers in Japan have not had to collect basic data, such as health-state utility values (HSUVs), for health technology assessment because, unlike in Europe and in many non-European countries, health technology assessment has not been actively applied to health policy decision making since the 1990s.[5] Therefore, in Japan, basic data for CUA of MBC treatment is needed. This is especially needed to estimate HSUVs and calculate QALYs for CUA use.

The most common measure of the outcome in CUA is QALYs. In the QALY method, quality adjustment is based on a set of values called HSUVs, which suggests the relative desirability of the health condition. These utilities reflect the value of the health-state and improvement in health condition. The Japanese Guideline for Preparing Cost-Effectiveness Evaluation to the Central Social Insurance Medical Council recommends that when QALYs is calculated, HSUVs should be reflective of the value in a general population (using a preference-based measure (PBM) or direct methods, such as the standard gamble (SG) and the time trade-off (TTO)). Moreover, if Japanese HSUVs are newly collected for CUA, the use of PBMs with a value set developed in Japan using TTO (eg, the EuroQol 5-Dimension 5-Level (EQ-5D-5L) questionnaire[8]) is recommended as the first choice.[5] As described in detail in the Methods section, in our study, data for HSUVs were collected using EQ-5D-5L, following the guidelines.

In 2016, we were mandated to use CUA to determine the treatment of MBC, using the earlier database before the Central Social Insurance Medical Council established the Japanese Guidelines for the Preparation of Cost-Effectiveness Assessments. Therefore, we were faced with a situation of available clinical data, but no HSUV. We included existing clinical data into a mathematical model to estimate the HSUVs in patients with MBC and used the results for the CUA. Lloyd *et al*[9] developed a predictive model of HSUVs for MBC patients in the United Kingdom. The Lloyd's predictive model established as a base, one hundred members of the UK general public who rated the burden of progressive, responding and stable disease on treatment; and also, febrile neutropenia (FN), stomatitis; diarrhoea/vomiting; fatigue; hand-foot syndrome (grade 3/4 toxicities) and hair loss, using SG to determine HSUVs. We predicted that the Lloyd model may be useful for estimating HSUVs for patients registered in databases containing clinical data only, including several MBC databases in Japan. However, the Lloyd model's accuracy in predicting HSUVs for Japanese patients with MBC has not yet been validated.

This study aimed to determine whether the Lloyd model (an HSUV prediction model) can accurately predict the actual HSUVs for Japanese patients with MBC.

METHODS

We validated the predictive accuracy of HSUVs estimated by including clinical data from Japanese patients with MBC into the Lloyd model. The study consisted of two phases. In the first phase, we constructed a database of clinical data and HSUVs for Japanese patients with MBC in a real-world setting to evaluate the predictive accuracy of HSUVs calculated using the Lloyd model. In the second phase, we assessed how accurately predicted HSUVs (based on the Lloyd model) correlated with actual HSUVs obtained using preference-based health status measures in Japanese patients with MBC.

HSUVs and patient-reported outcomes

The first phase of our study involved developing a comprehensive database of HSUVs and patient-reported outcomes (PROs) for Japanese patients with MBC, which is linked to patients' social background and treatment history, and PRO surveys of adverse events from anticancer agents using a questionnaire. The study sample included patients who attended the Outpatient Breast Clinic at the Department of Breast and Endocrine Surgery, St. Marianna University School of Medicine, Kawasaki, Japan, between May 2016 and September 2018. The inclusion criteria were (1) Japanese women aged >20 years, (2) a histopathological diagnosis of breast cancer and (3) provision of written informed consent for study participation. Exclusion criteria were (1) undergoing active treatment for mental disorders and (2) participation in other clinical trials.

We developed this longitudinal study to collect clinical data on patients' disease conditions and treatments. Patients' social background factors (age, educational level, marital status, residential environment, employment status and household income) were examined; and breast cancer condition survey (including breast cancer subtypes, number of metastatic organs, metastatic organ sites and treatment response, and HSUVs (measured using the EQ-5D-5L questionnaire) were completed at entry, as per the study schedule (figure 1).

The EQ-5D-5L is a PBM scale developed by the EQ groups,[8] and this is a target measure of how accurately the Lloyd's model predicts the HSUV, measured by this questionnaire,

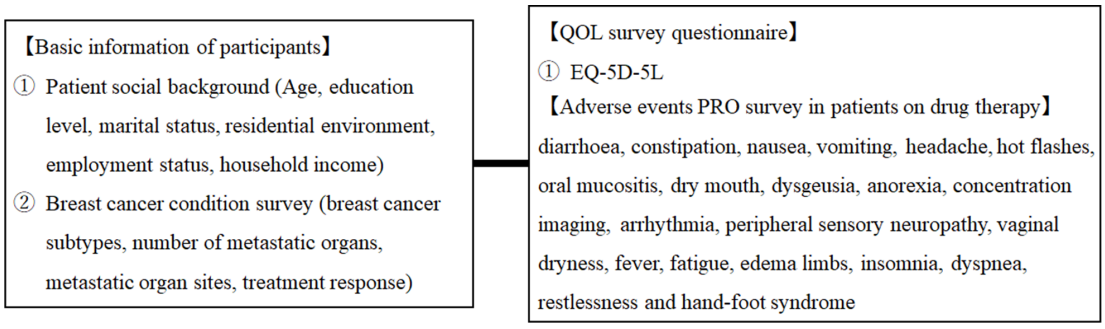

Figure 1 Characteristics of Japanese patients with metastatic breast cancer in the health utility values and patient-reported outcomes database. EQ-5D-5L, EuroQol 5-Dimension 5-Level; PRO, patient-reported outcome; QOL, quality of life.

in our study. This descriptive system comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has five levels: no problems, slight problems, moderate problems, severe problems and extreme problems. Using this questionnaire form, 3125 health-state patterns, ranging from 11111 (representing the best health-state) to 55555 (worst health-state), can be defined. These 3125 health-state patterns may be converted into a country-specific single index value (so-called HSUVs) using country-specific value sets, which have been derived from large country-specific validation studies using time-trade-off/discrete choice methodology and which anchor 1 for 'perfect health' and 0 for 'dead,' respectively.[10] In other words, the HSUVs calculated using the EQ-5D-5L are a value of the respondent's health status from the general public's perspective.

The results of a PRO survey of drug therapy-related (hormone therapy, chemotherapy, and targeted molecular therapy) adverse events were also evaluated with treatment response. Drug therapy-related adverse events questionnaires were completed following the Common Terminology Criteria for Adverse Events V.4.0.[11] The questionnaire included five levels of severity: 'none' to grades 1–4. The 21 items surveyed were: diarrhoea, constipation, nausea, vomiting, headache, hot flashes, oral mucositis, dry mouth, dysgeusia, anorexia, concentration imaging, arrhythmia, peripheral sensory neuropathy, vaginal dryness, fever, fatigue, limb oedema, insomnia, dyspnoea, restlessness and hand-foot syndrome (figure 2).

Patients receiving hormone therapy for the treatment of MBC were surveyed every 9±3 weeks using HSUVs and PRO surveys of adverse events. Patients receiving chemotherapy or molecular targeted therapy were interviewed every 6±3 weeks using HSUVs and PRO surveys of adverse events. Patients receiving radiation therapy were evaluated once during the treatment period using HSUVs.

Research assistants distributed the questionnaires to the participants before the physician's examination and collected them approximately 30 min later. All data were collected in the same manner. All data were collected in accordance with the protocol. There were no missing data in the repeated measures.

Validation of the predictive accuracy of HSUVs based on the Lloyd model

The Lloyd model evaluated the clinical data of Japanese patients with MBC to predict HSUVs.[9] The model consisted of population-based societal preferences for distinct stages of MBC and six common toxicities. Health states were developed based on literature review, iterative cycles of interviews and a focus group with clinical experts. This predictive model established as a base, 100 members of the UK general public who rated the burden of progressive, responding and stable disease on treatment; and also, FN, stomatitis; diarrhoea/vomiting; fatigue; hand-foot syndrome (grade 3/4 toxicities); and hair loss, using SG to determine HSUVs. The Lloyd model was based on the logistic model using the linear combination of disease states and toxicities. Stable disease on treatment had a utility value of 0.72, with a corresponding gain of p=0.07 following a treatment response and a decline by 0.27 for disease progression. Toxicities led to declines in utility between 0.10 (diarrhoea/vomiting) and 0.15 (FN). The Lloyd model's estimated HSUVs were compared with those measured using the EQ-5D-5L questionnaire to verify the Lloyd model's reliability.

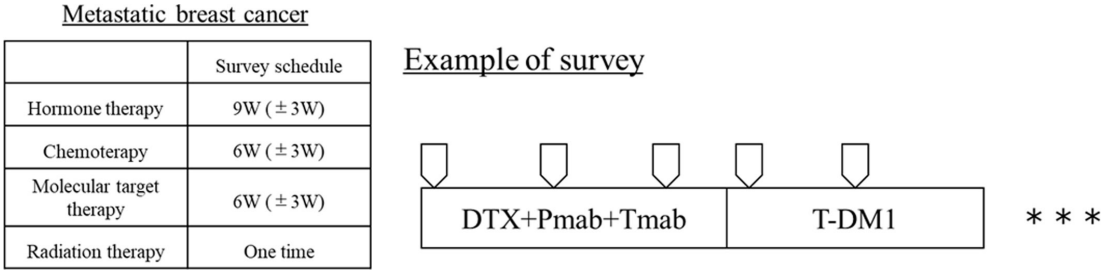

Figure 2 Longitudinal study schedule based on patients' disease condition and treatment. DTX, docetaxel; Pmab, pertuzumab; T-DM1, trastuzumab emtansine; Tmab, trastuzumab; W, weeks.

Statistical analyses

Correlation analysis between the predicted HSUVs based on the Lloyd model and actual HSUVs obtained using the EQ-5D-5L questionnaire

We first examined the distribution of both data from the predicted HSUVs based on the Lloyd's model and the actual HSUVs derived from the Japanese value set obtained using the EQ-5D-5L questionnaire. When the data were normally distributed, we used Pearson's product rate correlation coefficient, and when they were skewed, we used Spearman's rank correlation coefficient.

Validation of the predictive accuracy of HSUVs based on the Lloyd model

The Bland-Altman analysis was used to determine how accurately predicted HSUVs (based on the Lloyd model) correlated with actual HSUVs obtained using the EQ-5D-5L questionnaire. The presence or absence of fixed or systematic errors (ie, bias) and precision were determined by calculating the SD of the difference between measured and predicted values. The 95% limits of agreement (LOA) were calculated as bias ±2SD. Confidence intervals at the upper end of the 95% LOA were calculated as $\text{bias} + 2\ \text{SD} \pm \text{t} \times \sqrt{\frac{3\text{SD}^2}{\text{n}}}$, where t is defined as the 97.5%-tile value of t distribution and n is defined as the number of samples. CIs at the lower end of the 95% LOA were calculated as $\text{bias} - 2\ \text{SD} \pm \text{t} \times \sqrt{\frac{3\text{SD}^2}{\text{n}}}$.

Since we repeatedly used and reviewed data from the same subjects, accuracy and the 95% LOA were corrected for repeated measurements using the analysis of variance (ANOVA) method.[12] Results obtained from the same subject contained both patient-specific and method-dependent errors that were not completely independent and needed to be corrected. This is because the mean bias is not affected by repeated measurements; however, precision and the 95% LOA are. The method is described in the following steps:

Step 1: Perform a one-way ANOVA with the patient included as a random factor.

Step 2: Determine the variance of the differences between repeated measurements in the same patient.

Step 3: Determine the variance of the differences in measurements between patients.

Step 4: Since the variance of the differences in measurements between patients determined in step 3 depends on the number of samples (n), we divided by $\frac{(\sum \text{mi})^2 - \sum \text{mi}^2}{(\text{n}-1)\sum \text{mi}}$, where m is defined as the number of repeated measurements in each patient to obtain the variance of the differences in measurements between the corrected patients, which equals the variance of the differences in measurements derived using the corrected measurement method.

Step 5: The 'variance of the differences between repeated measurements in the same patient' determined in step 2 plus the 'variance of the differences in measurements between the corrected patients' (ie, 'the variance of the differences in measurements derived using the corrected measurement method') determined in step 4 equals the variance of the differences in measurements obtained by repeated measurements.

Step 6: The square root of the variance obtained in step 5 is used to calculate precision (the SD of the differences in measurements taking into account repeated measurements).

Finally, a correlation significance test was performed to determine the presence of a proportional error.[13] Proportional agreement analysis was used to determine the presence or absence of a proportional error by testing the correlation's significance. Significance level was set at 0.05. All statistical analyses were performed using JMP V.15.1 (SAS Institute).

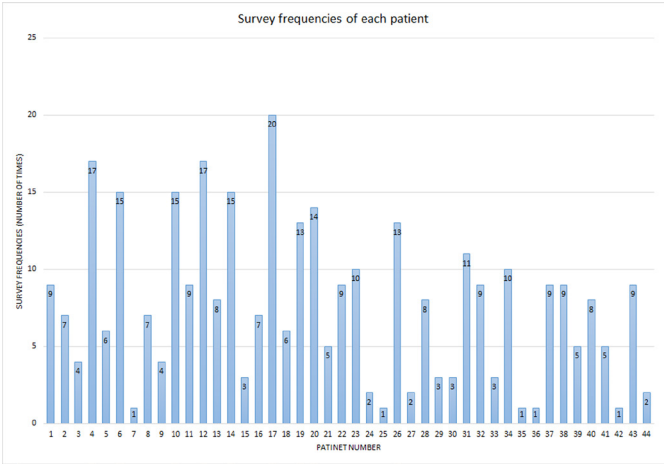

Figure 3 Survey frequencies of each of the 44 patients with metastatic breast cancer.

Patient and public involvement

It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

HSUVs and PROs

Forty-four Japanese patients with MBC or recurrent breast cancer were studied at 336 survey points. Figure 3 shows the details of the survey frequencies of the 44 patients. The clinical characteristics are summarised in table 1 and adverse events are summarised in table 2.

Correlation analysis between the predicted HSUVs based on the Lloyd model and actual HSUVs obtained using the EQ-5D-5L questionnaire

The results of the distribution of both data from the predicted HSUVs based on the Lloyd model and the actual HSUVs obtained using the EQ-5D-5L questionnaire are shown in figure 4. The predicted HSUVs based on the Lloyd model and the actual HSUVs obtained using the EQ-5D-5L questionnaire were both highly skewed towards 1.0. We, therefore, carried out a correlation

Table 1 The clinical characteristics in Japanese patients with MBC

Variable	Patients (n=44)	
Age at enrolment (years), mean±SD (range)	57.4±11.6 (29–80)	
Breast cancer subtype, survey frequency		N (%)
HR+/HER2– (luminal)	30	68.1
HR–/HER2+ (HER2)	4	9.1
HR+/HER2+ (luminal–HER2)	4	9.1
HR–/HER2- (triple-negative)	5	11.1
Unknown	1	2.3
No of metastatic organs, survey frequency		N (%)
1	19	46.3
2	7	17.1
≥3	15	36.6
Potentially life-threatening organ metastases survey frequency (liver, lung, brain)		N (%)
+	27	61.4
–	17	38.6
Metastatic organs (including duplicates), survey frequency		N (%)
Liver	13	29.5
Lung	17	38.6
Brain	4	9.1
Bone	16	36.4
Distant LNs	17	38.6
Breast/skin	22	50.0

HER2, human epidermal growth factor receptor 2; HR, hormone receptor; LN, lymph node; MBC, metastatic breast cancer.

Table 2 The adverse events in Japanese patients with MBC

Adverse events (44 eligible patients, 336 survey points) (including duplicates)	N (%)
Diarrhoea	46 (13.7)
Vomiting	21 (6.3)
Stomatitis	49 (14.6)
FN	0 (0.0)
Fatigue	218 (64.9)
Hand-foot syndrome	18 (5.4)
Hair loss	96 (28.6)
Disease states (44 eligible patients, 336 survey points)	
Treatment response (stable disease)	334 (99.4)
Disease progression	2 (0.6)
Mean HSUVs derived from EQ-5D-5L (44 eligible patients, 336 survey points)	0.83172

EQ-5D-5L, the EuroQol 5-Dimension 5-Level questionnaire; FN, febrile neutropenia; HSUVs, health-state utility values; MBC, metastatic breast cancer.

analysis using Spearman's rank correlation coefficient ρ, which was 0.1379. This means that there was no simple correlation between the predicted HSUVs based on the Lloyd model and the actual HSUVs obtained using the EQ-5D-5L questionnaire.

Validation of the predictive accuracy of HSUVs based on the Lloyd model

The results of the Bland-Altman analysis are shown in figure 5 and table 3. We first examined the fixed error. The mean difference between HSUVs estimated by the Lloyd model and actual HSUVs measured using the EQ-5D-5L questionnaire, or fixed error (bias), was −0.106. The precision was 0.165. The 95% LOA ranged from −0.436 to 0.225. The upper end of the 95% LOA CI ranged from 0.042 to 0.407, while the lower end of the CI ranged from −0.618 to −0.253.

We corrected for the accuracy of repeated measurements using a one-way ANOVA with the patient included as a random factor. The results are shown in table 4. The variance in measurement differences caused by repeated measurements in the same patient was 0.014. The variance of the differences in measurements between patients was 0.116–0.014=0.102. The variance of the differences in measurements between the corrected patients was 0.014.

The sum of the 'variance of the differences between repeated measurements in the same patient' and the 'variance of the differences in measurements between the corrected patients' (ie, the variance of the differences in measurements obtained by repeated measures) was 0.0141+0.0135 = 0.0276. The square root of this variance,

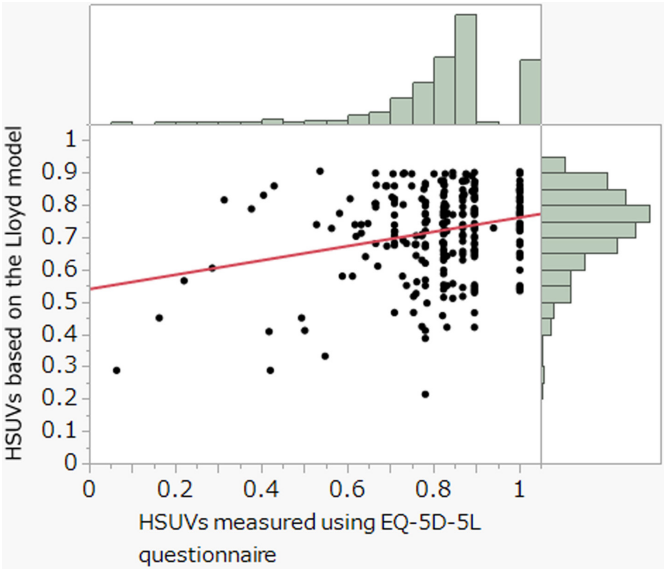

Figure 4 Correlation analysis of between the predicted HSUVs based on the Lloyd model and actual HSUVs obtained using the EQ-5D-5L questionnaire. EQ-5D-5L, EuroQol 5-Dimension 5-Level; HSUV, health-state utility values.

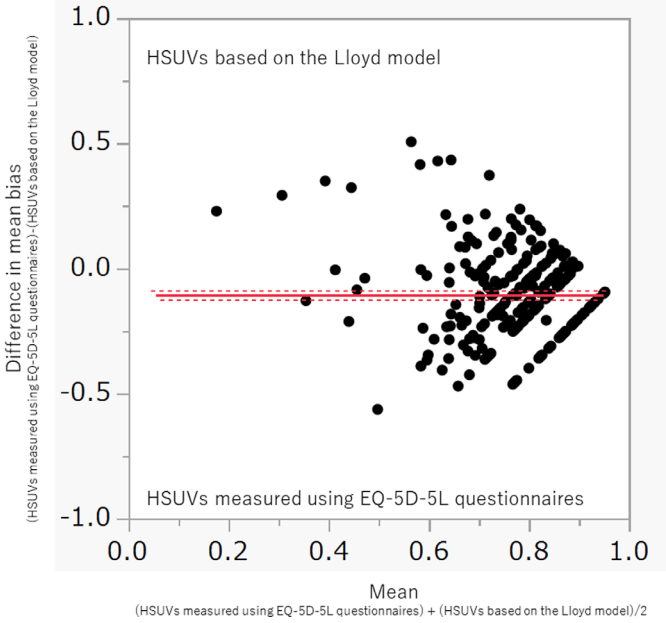

Figure 5 Validation of the predictive accuracy of HSUVs based on the Lloyd model by the Bland-Altman analysis. EQ-5D-5L, EuroQol 5-Dimension 5-Level; HSUV, health-state utility values.

which is the precision (ie, the SD of the differences in measurements, taking into account repeated measurements) (table 5), was 0.166.

Second, a correlation significance test was performed to determine the presence of a proportional error. For a sample size of 336, the t value for 2 ddf at the 5% significance level was 4.30. In the Bland-Altman analysis, where γ=0.24894 and n=336, the formula $t = \gamma\sqrt{\frac{n-2}{1-\gamma^2}}$ yielded a t value of 4.697, which was greater than the t value with 2 df at the 5% significance level (p=0.425). We concluded that there was a significant correlation, suggesting that a proportional error exists.

DISCUSSION

This study determined whether the Lloyd model, which was the only prediction model that can infer HSUVs from clinical data, can accurately predict the actual HSUVs for Japanese patients with MBC. Our results showed that there were fixed and proportional errors associated with the prediction of HSUVs based on the Lloyd model for Japanese patients with MBC. We, therefore, investigated the clinical significance of a mean difference (bias) of 0.106 between them. McClure *et al*[14] reported that the simulation-based instrument-defined minimally important difference (MID) estimate (mean±SD) for each Japanese-specific scoring algorithm was 0.048±0.004. Differences in MID estimates reflect differences in population preferences in valuation techniques used, as well as in modelling strategies. After excluding the maximum-valued scoring parameters, the MID estimate was 0.044±0.004. These results suggest that the difference between our estimates of HSUVs based on the Lloyd model and the actual measurements of HSUVs based on the EQ-5D-5L is larger than the MID of the EQ-5D-5L. In other words, the mean difference (bias) of 0.106 was found to be clinically meaningful.

Although we showed an error between the predicted HSUVs and the actual HSUVs, we still considered the Lloyd model to be useful for predicting HSUVs in Japanese patients with MBC. The first reason is that HSUVs obtained in this study are in close agreement with previous studies using PBMs, such as EQ-5D, or direct methods, such as the SG and the TTO.[15–22] The mean HSUV for Japanese patients with MBC in this study was 0.726 based on the Lloyd model and 0.831 using the EQ-5D-5L questionnaire. The difference between predicted and actual HSUVs was 0.105. For example, Tachi *et al*[15] reported HSUVs calculated using the EQ-5D ranging from 0.73±0.18 to 0.84±0.17 for Japanese patients with breast

Table 3 The results of the Bland-Altman analysis

HSUVs based on the lloyd model	0.72627
HSUVs measured using EQ-5D-5L questionnaires	0.83172
Difference in mean bias	−0.1055
SE	0.009
No of samples	336
Precision	0.165
Correlation	0.24894
t value	11.7207
95% LOA	−0.436,0.225

EQ-5D-5L, EuroQol 5-Dimension 5-Level; HSUVs, health-state utility values; LOA, limits of agreement.

Table 4 ANOVA for the calculation of agreement with multiple observations per individual

ANOVA	df	SS	MS
Subject (between columns)	43	5.003	0.116
Residual (within columns)	292	4.108	0.014
Total	335	9.111	

ANOVA, analysis of variance; MS, mean square; SS, sum of squares.

Table 5 Fixed error between HSUVs estimated by the Lloyd model and those measured using the EuroQol 5-Dimension 5-Level questionnaire

Fixed error	No correction for repeated measurements	Correction for repeated measurements
Mean difference (bias)	−0.106	
Precision	0.165	0.166
95% LOA	−0.436 to 0.225	−0.438 to 0.227

HSUV, health-state utility value; LOA, limits of agreement.

cancer, including those receiving chemotherapy for MBC. The range of variability they report for these HSUVs is comparable to the fixed error found in our study. Tachi *et al*[15] and Chou *et al*[23] also reported that adverse events during chemotherapy were associated with lower HSUVs, as did Lloyd *et al*.[9] Second, the guidelines for conducting cost-effectiveness analyses in Japan's healthcare sector recommend that sensitivity analysis be conducted. Cost-effectiveness analysis involves uncertainty in the analytical framework and the data itself. The results are reanalysed under different conditions to see how they change. Since HSUVs are also assessed for uncertainty under different conditions, we suggest that the Lloyd model can be applied to Japanese patients with MBC. An error of approximately 0.1 between predicted and actual HSUVs in our study is within the sensitivity analysis range. Finally, as far as we know, the Lloyd's model was the only mathematical model that allowed us to infer HSUVs from the available clinical data. We used the results of the quality of life (QOL) questionnaire, which we previously studied, to estimate HSUVs based on the report of a mapping algorithm. For example, Hagiwara *et al*[24] developed mapping algorithms, which can be used to generate the EQ-5D-5L index from the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 or The Functional Assessment of Cancer Therapy-General in cost-effectiveness analyses for Japanese MBC patients. Thus, mapping is a useful method when data on disease-specific QOL measures are available, but when only clinical data are available, the Lloyd's model is currently the only way to estimate HSUVs in patients with MBC.

This study has several limitations. First, the initial phase of this study was to construct a database of clinical data and HSUVs for Japanese patients with MBC in a real-world setting to evaluate the predictive accuracy of HSUVs calculated using the Lloyd model. However, because our database was constructed at a single institution, it was relatively small (336 survey points of 44 Japanese patients with MBC) and included repeated measurements. Therefore, a simple Bland-Altman analysis was not sufficient to evaluate the accuracy of the Lloyd model. Instead, we adopted the Bland and Altman method[12] to correct the results. The precision with and without correction for repeated measurements was 0.165 and 0.166, respectively, with a difference of 0.001. The difference in precision was small, even after correction, because the number of repeated measurements was less than the number of subjects. Second, there is a potential bias in the frequency of adverse events, a parameter that was substituted into the Lloyd model. For example, our adverse event data did not show the occurrence of FN. One reason for the absence of FN, despite the reported frequency of FN in our study regimen of up to 68.8%, is that we treated our high-risk FN patients with 'pegfilgrastim' (recombinant human granulocyte colony-stimulating factor analogue filgrastim).[25–30] Thus, the fact that the frequency of adverse events changed over time due to the development of supportive care may be key in determining which adverse events to select as parameters when building a predictive model of HSUVs in the future. In fact, we predicted that since the Lloyd model was developed in the UK, there could be some differences in the coefficients and their impact between the UK and Japan. Therefore, it would be better to develop a Japanese model, rather than to apply the UK-based model. We attempted to develop a model using our data and compared the significant factors and their coefficients with those of the Lloyd's model. However, we were unable to develop a comparable Japanese model because of the substantial differences in the frequencies of adverse events and the extent of disease progression between our data and the background data of the Lloyd's model. Finally, we were unable to verify whether the HSUV estimates from the Lloyd model, which models HSUV using clinical factors alone, may potentially differ from those measured using a PBM, such as EQ-5D-5L. This is because HSUV is generally affected independently by physical, mental, and social factors. In other words, the difference between the EQ-5D-5L-measured values of HSUV and the estimated values of HSUV using the Lloyd model found in our study may have been caused by a combination of (1) the limitations of the Lloyd model itself, as described above, and (2) the differences in the background of the development of the model, since the Lloyd model was created for the general UK population, whereas our cohort included Japanese breast cancer patients. Another limitation of our examination was that it was not possible to determine the factor that caused this difference.

CONCLUSION

In conclusion, we showed that there were fixed and proportional errors associated with the prediction of HSUVs based on the Lloyd model for Japanese patients with MBC. When cost-effectiveness analyses are conducted using HSUVs calculated using the Lloyd model, we recommend that sensitivity analysis be performed, assuming an error in the HSUVs. In the future, the authors plan to apply the methods of Lloyd *et al* to build a predictive model of HSUVs in Japanese MBC patients.

Acknowledgements We would like to thank Editage (www.editage.com) for their writing support.

Contributors TI and EI conceived and designed the study. TI, KT participated in the acquisition. TI and EI analysed the data, and TI, EI and KT interpreted the data. TI made the draft writing of the manuscript, and TI, EI and KT made critical revision for important intellectual content, and the decision to submit the manuscript for publication. TI, EI and KT have read and approved the final version and agree to be accountable for all aspects of the work. TI takes the responsibility for the overall content as the guarantor.

Funding This work was supported by a Grant-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JSPS) (JP16K19180).

Competing interests KT received honoraria from AstraZeneca, Chugai Pharmaceutical, Daiichi Sankyo, Eizai, Eli Lilly Japan KK, Nippon Kayaku, Pfizer, Taiho Pharmaceutical and Takeda Pharmaceutical.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval All procedures were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study was approved by the Institutional Review Board and Ethics Committee of St. Marianna University School of Medicine, Kanagawa, Japan (No.3146). All participants provided written informed consent to participate.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Extra data can be accessed via the Dryad data repository at http://datadryad.org/ with the doi:10.5061/dryad.r2280gbbp.

ORCID iD

Tsuguo Iwatani http://orcid.org/0000-0001-7162-1515

REFERENCES

1 Cancer Registry and Statistics. Cancer information service, National cancer center, Japan (monitoring of cancer incidence in Japan (MCIJ)). Available: https://ganjoho.jp/reg_stat/statistics/dl/index.html#incidence4pref [Accessed 30 Mar 2020].
2 Gennari A, Conte P, Rosso R, *et al*. Survival of metastatic breast carcinoma patients over a 20-year period: a retrospective analysis based on individual patient data from six consecutive studies. *Cancer* 2005;104:1742–50.
3 Shiroiwa T, Fukuda T, Ikeda S, *et al*. New decision-making processes for the pricing of health technologies in Japan: the FY 2016/2017 pilot phase for the introduction of economic evaluations. *Health Policy* 2017;121:836–41.
4 Vokinger KN, Hwang TJ, Grischott T, *et al*. Prices and clinical benefit of cancer drugs in the USA and Europe: a cost-benefit analysis. *Lancet Oncol* 2020;21:664–70.
5 Shiroiwa T, Fukuda T, Ikeda S, *et al*. Development of an official guideline for the economic evaluation of Drugs/Medical devices in Japan. *Value Health* 2017;20:372–8.
6 Shih Y-CT, Halpern MT. Economic evaluations of medical care interventions for cancer patients: how, why, and what does it mean? *CA Cancer J Clin* 2008;58:231–44.
7 Tappenden P, Chilcott J, Ward S, *et al*. Methodological issues in the economic analysis of cancer treatments. *Eur J Cancer* 2006;42:2867–75.
8 Herdman M, Gudex C, Lloyd A, *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
9 Lloyd A, Nafees B, Narewska J, *et al*. Health state utilities for metastatic breast cancer. *Br J Cancer* 2006;95:683–90.
10 Shiroiwa T, Fukuda T, Ikeda S, *et al*. Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D. *Qual Life Res* 2016;25:707–19.
11 Common terminology criteria for adverse events: CTCAE v4.0. Available: http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc_40 [Accessed 1 Aug 2015].
12 Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007;17:571–82.
13 Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* 2002;29:527–36.
14 McClure NS, Sayah FA, Xie F, *et al*. Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value Health* 2017;20:644–50.
15 Tachi T, Teramachi H, Tanaka K, *et al*. The impact of outpatient chemotherapy-related adverse events on the quality of life of breast cancer patients. *PLoS One* 2015;10:e0124169.
16 Hagiwara Y, Shiroiwa T, Shimozuma K, *et al*. Impact of adverse events on health utility and health-related quality of life in patients receiving first-line chemotherapy for metastatic breast cancer: results from the select bc study. *Pharmacoeconomics* 2018;36:215–23.
17 Paracha N, Thuresson P-O, Moreno SG, *et al*. Health state utility values in locally advanced and metastatic breast cancer by treatment line: a systematic review. *Expert Rev Pharmacoecon Outcomes Res* 2016;16:549–59.
18 Hao Y, Wolfram V, Cook J. A structured review of health utility measures and elicitation in advanced/metastatic breast cancer. *Clinicoecon Outcomes Res* 2016;8:293–303.
19 Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. *Expert Rev Pharmacoecon Outcomes Res* 2010;10:553–66.
20 Kuchuk I, Bouganim N, Beusterien K, *et al*. Preference weights for chemotherapy side effects from the perspective of women with breast cancer. *Breast Cancer Res Treat* 2013;142:101–7.
21 Earle CC, Chapman RH, Baker CS, *et al*. Systematic overview of cost-utility assessments in oncology. *J Clin Oncol* 2000;18:3302–17.
22 Pickard AS, Wilke CT, Lin H-W, *et al*. Health utilities using the EQ-5D in studies of cancer. *Pharmacoeconomics* 2007;25:365–84.
23 Chou T-C, Chiang S-C, Ko Y. Health state utilities for metastatic breast cancer in Taiwan. *Breast* 2020;51:57–64.
24 Hagiwara Y, Shiroiwa T, Taira N, *et al*. Mapping EORTC QLQ-C30 and FACT-G onto EQ-5D-5L index for patients with cancer. *Health Qual Life Outcomes* 2020;18:354.
25 Roché H, Fumoleau P, Spielmann M, *et al*. Sequential adjuvant epirubicin-based and docetaxel chemotherapy for node-positive breast cancer patients: the FNCLCC PACS 01 trial. *J Clin Oncol* 2006;24:5664–71.
26 Toi M, Nakamura S, Kuroi K, *et al*. Phase II study of preoperative sequential FEC and docetaxel predicts of pathological response and disease free survival. *Breast Cancer Res Treat* 2008;110:531–9.
27 Sparano JA, Wang M, Martino S, *et al*. Weekly paclitaxel in the adjuvant treatment of breast cancer. *N Engl J Med* 2008;358:1663–71.
28 Citron ML, Berry DA, Cirrincione C, *et al*. Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of intergroup trial C9741/Cancer and leukemia group B trial 9741. *J Clin Oncol* 2003;21:1431–9.
29 Jones SE, Savin MA, Holmes FA, *et al*. Phase III trial comparing doxorubicin plus cyclophosphamide with docetaxel plus cyclophosphamide as adjuvant therapy for operable breast cancer. *J Clin Oncol* 2006;24:5381–7.
30 Kosaka Y, Rai Y, Masuda N, *et al*. Phase III placebo-controlled, double-blind, randomized trial of pegfilgrastim to reduce the risk of febrile neutropenia in breast cancer patients receiving docetaxel/cyclophosphamide chemotherapy. *Support Care Cancer* 2015;23:1137–43.