Original Article

Cultural adaptation and psychometric assessment of the Japanese version of the Treatment Satisfaction Questionnaire for Medication (TSQM)

Anzu YAKUSHIJI*, Ayumi MORISHITA*, Koki IDEHARA*, Louis Patrick WATANABE*, Takeshi AKIYAMA* and Ana Maria RODRIGUEZ-LEBOEUF[2*]

Objective Patients with high treatment satisfaction are known to have better treatment compliance and persistence, and the Treatment Satisfaction Questionnaire for Medication (TSQM) is a questionnaire widely used to assess treatment satisfaction. A Japanese version of this questionnaire is available; however, its application is limited. In this study, we aimed to culturally adapt and assess the psychometric properties of a newly developed Japanese version of the TSQM in patients with diabetes or migraine who were prescribed oral and injectable medications in Japan.

Methods The study included two phases; phase I, retranslation and cultural adaptation of the Japanese TSQM, and phase II, psychometric validation of the updated Japanese TSQM. Japanese patients with diabetes or migraine receiving oral and/or injection treatment were involved in the cognitive debriefing interview and psychometric validation. The retranslation and cultural adaptation of the Japanese TSQM were conducted following the process recommended by the ISPOR Task Force. The updated Japanese TSQM, finalized after amendments based on cognitive debriefing interviews, was distributed to patients for psychometric validation. The psychometric properties of the finalized Japanese TSQM were assessed using classical test and Rasch measurement theories.

Results Minor modifications were made to the Japanese TSQM after cognitive debriefing interviews with eight patients. In total, 512 patients completed the final version of the Japanese TSQM. All versions of the Japanese TSQM had moderate reliability, with Cronbach's alpha ranging 0.6–0.9 for each domain and intraclass correlation coefficients for test-retest reliability between 0.5 and 0.6 for most domains. High item-convergent and item-discriminant validity were observed, in which all items had higher correlations with their respective domains than with other domains of the Japanese TSQM. The Rasch measurement theory added additional support to the reliability and validity of the measures as they fit the Rasch models. Additionally, high internal consistency reliability, unidimensionality of the domains, and scale invariance were supported by the values of the person separation index, Chi-square value of each domain, and differential item functioning across patient subgroups, respectively.

Conclusion The Japanese TSQM was successfully updated by revising the complex wording and removing wording specific to oral medications, making it usable for patients regardless of medication type. The updated Japanese TSQM had sufficient reliability and validity and could be used regardless of the disease and mode of administration of medication.

Key words : Treatment Satisfaction Questionnaire for Medication (TSQM), treatment satisfaction, cultural adaptation, psychometric validation, patient reported outcome

Nihon Koshu Eisei Zasshi　2026; 73(3): 237–246.　doi:10.11236/jph.25-018

I. BACKGROUND

Patient satisfaction with treatment is a crucial concept for healthcare. High treatment satisfaction is associated with compliance and persistence with treatment[1]. One of the measures to assess treatment satisfaction is the Treatment Satisfaction Questionnaire for Medication (TSQM). The TSQM is a measure that has been developed based on a conceptual framework of treatment satis-

* Real-World Evidence Solutions & HEOR IQVIA Solutions Japan G.K.
2* Patient Centered Solutions, IQVIA.
Correspondence author: Ayumi Morishita
Keikyu Dai-1 Building 4–10–18 Takanawa, Minato-ku, Tokyo 108–0074 JAPAN
E-mail : ayumi.morishita@iqvia.com

faction[1]. There are three versions of the TSQM: TSQM version 1.4 (TSQM-1.4), TSQM version II (TSQM-II), and TSQM version 9 (TSQM-9). All versions have gone through rigorous psychometric assessment, and have demonstrated satisfactory construct validity for treatment satisfaction, as well as known-group validity in relation to adherence[2~4].

All versions of the TSQM have previously been translated into Japanese. However, the Japanese TSQM had not been widely adopted. Inputs from stakeholders identified three main issues with the original Japanese TSQM. Firstly, patients found terminology used in the measure difficult to understand. Secondly, the terminology was restricted to oral medication posing challenges for patients using non-oral treatment when responding. Lastly, there was no openly available evidence on the methodology used for the linguistic validation, and no psychometric validation had been conducted. This raised concerns on its reliability and validity.

The objective of the study was to retranslate, culturally adapt, and assess the psychometric properties of the Japanese versions of the TSQM-1.4, TSQM-II, TSQM-9 using a sample of Japanese patients with diabetes and migraine. An exploratory objective was to assess the psychometric properties of the Japanese versions of the TSQM for different modes of treatment administration, namely oral and injectable modes of administration.

II. METHODS

1. Study Design

The study was a combination of a cross-sectional qualitative cognitive debriefing study (Phase I) and a longitudinal cohort study (Phase II).

2. Sample Populations for Phase I and Phase II

For both Phase I and Phase II, the target populations consist of patients in Japan having either diabetes or migraine. Diabetes and migraine were chosen as diseases of interest as they are diseases with oral and injectable medication that are common amongst the Japanese population. The prevalence of patients with diabetes and with migraine are estimated to be 14.6% and the 8.4% of the Japanese population, respectively[5,6]. Moreover, both diseases are chronic conditions which require long-term treatment. Therefore, treatment satisfaction is especially important for the adherence and compliance of treatment. As the TSQM is a generic measure[2~4], focusing on chronic diseases common in Japan was considered as a good rationale for selecting these target populations.

For both phases, the sampling strategy was a convenience sampling using an online patient panel owned by Rakuten Insight, Inc.[7] An invitation to the study was shared via email to patients that self-identified as having diabetes or migraine. If patients accessed the link, eligibility was confirmed using screening questions and online consent was obtained. When consenting, patients were asked to either provide contact information to schedule a qualitative interview or were sent a link to the online Japanese TSQM.

The eligibility criteria included being 18 years and older, being diagnosed with diabetes or migraine, and being prescribed at least one treatment (oral or injection) for diabetes or migraine within the last three months of the study period.

3. Phase I: Assessment of Content Validity and Cultural Adaptation

The cultural adaptation of the Japanese TSQM versions was conducted following the process described in the "Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures", published by an ISPOR task force in 2005[8]. The process describes a 10-step process including 1) preparation; 2) forward translation; 3) reconciliation; 4) back translation; 5) back translation review; 6) harmonization; 7) cognitive debriefing; 8) review of results from cognitive debriefing; 9) proofreading and 10) final report[8].

Two independent bilingual translators conducted forward translations of all TSQM versions from English to Japanese. The two translations were reconciled into one version. Subsequently, two additional independent bilingual translators conducted the back-translation of the reconciled versions of the TSQM versions into English. All translators translated into their native language. The back-translated and reconciliation versions were reviewed by a scientific lead responsible for the TSQM (AMR), which led to the draft Japanese TSQM versions.

Cognitive debriefing interviews were conducted with Japanese patients with migraine and/or diabetes who were prescribed oral or injectable medication. Patients were prioritized based on their demographic and clinical characteristics (such as age, gender, and area of residency) from the list of eligible patients who consented to participate. They were approached by email or telephone to schedule an online interview session. During the interview, the participants were asked to complete the draft versions of the Japanese TSQM and were debriefed on their comprehension of the instructions, questions and response options. Once interviews were completed, the Japanese TSQM versions were finalized with the scientific team leading the translation of the TSQM versions and the scientific lead responsible for the TSQM during a final consolidation meeting. The results of the translations, cognitive debriefing interviews, and amendments were reviewed, leading to the final Japanese versions of the TSQM.

4. Phase II: Survey Deployment and Psychometric Assessment

The final Japanese TSQM versions were distributed amongst patients with migraine and/or diabetes. The survey data was used for the psychometric assessments of the Japanese TSQM versions.

Once the patients were identified as eligible and consented to participate, patients were directed to the main survey. Patients divided into four groups with equal sample sizes (i.e., 150 participants per group): (1) diabetes patients treated with oral medication, (2) diabetes patients treated with injectable medication, (3) migraine patients treated with oral medication, and (4) migraine patients treated with injectable medication. Each patient was randomly assigned to one of the TSQM versions. The participants were asked to complete the same survey 8 to 14 days after the first completion. This time interval was used based on previous studies which showed that the most commonly used time interval for assessing test-retest reliability was 1 to 2 weeks (25%)[9]. After careful assessment for the context of diabetes and migraine, it was judged to be a period where we would not anticipate change in disease severity or treatment satisfaction and where the potential for recall bias would be minimized between the two survey administrations.

The study (including both phases) was reviewed and approved by an independent ethics committee (SOUKEN Co., Ltd)[10] on August 10th, 2023.

5. Measures

Participating patients answered questions on sociodemographic and clinical characteristics, and were asked to also complete the following measures:

- The TSQM: Japanese TSQM-1.4, Japanese TSQM-II and Japanese TSQM-9. TSQM-1.4 and TSQM-II includes 14 and 11 questions, respectively, grouped into four domains: effectiveness, side effects, convenience, and global satisfaction. TSQM-9 include nine questions and consists of three domains, effectiveness, convenience, and global satisfaction[2~4].
- The Patient Global Impression of Severity (PGIS) scale: The PGIS includes one question on the severity of disease, using a Likert scale with five response options: *"None", "Mild", "Moderate", "Severe", "Very Severe"*[11]. The PGIS was included to assess the potential change in disease severity at the two time points when the Japanese TSQM versions were administered.

6. Statistical Analyses

The samples were described using means and standard deviations (SD) for continuous variables and frequencies and percentages for categorical ones. The targeting of the measure to the sample was assessed using ceiling and floor effects frequencies.

The psychometric properties of the Japanese TSQM versions were assessed through Classical Test Theory (CTT) and Rasch Measurement Theory (RMT).

Reliability was assessed using internal consistency reliability and test-retest reliability. Internal consistency was assessed using Cronbach's alpha. For test-retest reliability, samples were grouped based on change in disease severity, as measured by the PGIS, from the first and second administration of the Japanese TSQM versions. To minimize potential changes in disease severity due to recall bias, the PGIS was regrouped into 3 categories: 1. *Severe and Very Severe*, 2. *Mild and Moderate*, 3. *None*. Intraclass correlation coefficient (ICC) (3,1) was used to assess test-retest reliability in the sample of respondents who had not changed disease severity between the two administrations of the Japanese TSQM versions[12]. Item-convergent validity and item-discriminant validity were assessed to evaluate how well each question within the domains measure the intended construct, and how distinct each question was from the questions measuring different constructs, respectively[13,14] using Spearman's correlation coefficient. A value of ≥0.4 was considered acceptable for item-convergent validity, while item-discriminant validity was supported when an item correlated more strongly with its own domain compared to other domains[14,15].

RMT was also used to assess unidimensionality construct validity and scale invariance (differential item functioning (DIF)) of the Japanese TSQM versions.

All analyses for CTT were completed using R version 4.3 and RMT was completed using RUMM 2030.

III. RESULTS

1. Cognitive Debriefing

Cognitive debriefing interviews were conducted with eight Japanese patients with migraine and/or diabetes who were prescribed oral or injectable medication. The eight patients included four diabetic patients, two migraine patients and two patients with both diseases. The measure was amended based on the cognitive debriefing interviews. Questions amended included the questions which involved the patients to answer the level of confidence and level of certainty. Many of the patients felt that these questions are hard to answer, and so were amended accordingly. Moreover, majority of the patients felt uncomfortable with the term "dissatisfied" in questions asking, "how satisfied or dissatisfied". Therefore, these questions were changed to "how satisfied".

2. Patient Characteristics

Out of 10,000 registered members of the Rakuten patient panel, 7,849 (78.5%) consented to participate. Of these, 2,868 (28.7%) met the eligibility criteria. A total of 512 participants were eligible and agreed to participate to the psychometric assessments of the Japanese TSQM versions. One hundred sixty-six participants completed Japanese TSQM-1.4, 179 completed Japanese TSQM-II, and 167 completed the Japanese TSQM-9, respectively.

As seen in Table 1, within the three cohorts, most of the participants were male; and were aged between 40 and 64 years old. Among the participants in the Japanese TSQM-1.4 cohort, three patients had diagnosis of both diabetes and migraine, and 51 patients were using both modes of administration at the first survey completion. Within participants who completed the Japanese TSQM-II, eight patients had a diagnosis of both diseases, and 64 were using treatment with both mode of administration. Eleven patients were diagnosed with both diseases in the Japanese TSQM-9 cohort, and 54 were using treatment of both mode of administration. These patients were asked to focus on a specific disease and specific mode of administration when answering the measure.

Table 1 Demographic and clinical characteristics of participants

	Japanese TSQM-1.4 $N=166$	Japanese TSQM-II $N=179$	Japanese TSQM-9 $N=167$
Sex (*n* (%))			
Female	52 (31.33%)	66 (36.87%)	56 (33.53%)
Male	114 (68.67%)	113 (63.13%)	111 (66.47%)
Age (*n* (%))			
<40 years old	19 (11.45%)	28 (15.64%)	15 (8.98%)
40–<65 years old	124 (74.70%)	125 (69.83%)	125 (74.85%)
≥65 years old	23 (13.86%)	26 (14.53%)	27 (16.17%)
Education level (*n* (%))			
Graduated from middle school	5 (3.01%)	2 (1.12%)	4 (2.40%)
Student at high school	0 (0.00%)	0 (0.00%)	1 (0.60%)
Graduated from high school	42 (25.30%)	44 (24.58%)	50 (29.94%)
Student at junior college or technical college	4 (2.41%)	8 (4.47%)	6 (3.59%)
Graduated from junior college or technical college	33 (19.88%)	39 (21.79%)	20 (11.98%)
Student at university or graduate school	4 (2.41%)	3 (1.68%)	6 (3.59%)
Graduated from university or graduate school	77 (46.39%)	83 (46.37%)	79 (47.31%)
Unknown	1 (0.60%)	0 (0.00%)	1 (0.60%)

Frequency of response (*n* (%)) presented for demographic characteristics of participants included in psychometric validation of each Japanese TSQM versions.

3. Classical Test Theory Psychometric Properties

Descriptive statistics and content validity

The description of the different cohorts and the targeting of the study samples to the Japanese TSQM measures are included in Table 2. The means (SD) for each domain of Japanese TSQM-1.4, TSQM-II and TSQM-9 had similar scores during the first administration.

Most patients who completed both survey administrations (94.6% of Japanese TSQM-1.4 cohort, 91.6% of in Japanese TSQM-II cohort, and 94.6% of Japanese TSQM-9 cohort).

No significant change was observed in the mean scores of domains for each version and the trend of disease severity at the second survey administration.

1) Internal consistency

As seen in Table 3, internal consistency using Cronbach's alpha for the Japanese TSQM-1.4 ranged from 0.72 to 0.92, from 0.81 to 0.90 for the Japanese TSQM-II, and 0.67 to 0.83 for Japanese TSQM-9, all of which are generally considered to indicate acceptable internal consistency[16].

2) Test-retest reliability

The assessment of test-retest reliability was completed with 93 patients for Japanese TSQM-1.4, 100 patients for Japanese TSQM-II and 106 patients for Japanese TSQM-9 who did not demonstrate change in disease severity between the two survey administrations, as measured by the rescored PGIS. All TSQM versions resulted in having ICC estimate lower than 0.75, which is generally recognized as limit of good or high reliability[17], between the two survey administrations. The estimated ICC was found to be especially low for the side effects domain of Japanese TSQM-II with a score of 0.37 (95% confidence interval (CI): 0.19–0.53). Test-retest reliability between the two survey administrations for Japanese TSQM-1.4, Japanese TSQM-9 and the other domains of the Japanese TSQM-II was moderate, with ICC estimate for the Japanese TSQM-1.4 varying from 0.54 to 0.67, 0.61 to 0.63 for the Japanese TSQM-9, and 0.60 to 0.69 for Japanese TSQM-II (for domains other than side effects).

3) Intra-test validity

Intra-test validity was supported by the high correlation between items and their corresponding domain, which exceeded the commonly used correlation coefficient threshold of 0.7[18] for all items in all Japanese TSQM versions. As observed in Table 4, items had correlations ranging from 0.90 to 0.93 with the effectiveness domain, 0.70 to 0.97 for the side effects domain, 0.75 to 0.86 for the convenience domain, and 0.84 to 0.86 for the global satisfaction domain for the Japanese TSQM-1.4. Similarly, correlation coefficients ranged from 0.70 to 0.95 between all items and their corresponding domain for the Japanese TSQM-II. A similar high degree of intra-test reliability was seen for the Japanese TSQM-9, with correlation coefficients ranging from 0.70 to 0.91 between items and their corresponding domain. Intra-test validity was further supported by the fact that all items had higher correlations with their respective domains than with other domains of the Japanese TSQM.

4. Rasch Measurement Theory

As observed in Table 5, the RMT analyses supported a

Table 2　Descriptive statistics of sociodemographic and clinical characteristics of participants and content validity (ceiling and floor effects)

	Japanese TSQM-1.4			Japanese TSQM-II			Japanese TSQM-9		
	1st survey administration (all patients)	Patients completing both survey administrations	2nd survey administration	1st survey administration	Patients completing both survey administrations	2nd survey administration	1st survey administration	Patients completing both survey administrations	2nd survey administration
	$N=166$	$N=157$	$N=157$	$N=179$	$N=164$	$N=164$	$N=167$	$N=158$	$N=158$
Allocation of group (*n* (%))									
Diabetes	100 (60.24%)	—	96 (57.83%)	100 (55.87%)	—	95 (53.07%)	100 (59.88%)	—	95 (56.89%)
Using oral medication	50 (30.12%)	—	47 (28.31%)	50 (27.93%)	—	47 (26.26%)	50 (29.94%)	—	47 (28.14%)
Using injection medication	50 (30.12%)	—	49 (29.52%)	50 (27.93%)	—	48 (26.82%)	50 (29.94%)	—	48 (28.74%)
Migraine	66 (39.76%)	—	61 (36.75%)	79 (44.13%)	—	69 (38.55%)	67 (49.12%)	—	63 (37.72%)
Using oral medication	50 (30.12%)	—	47 (28.31%)	50 (27.93%)	—	46 (25.70%)	50 (29.94%)	—	46 (27.54%)
Using injection medication	16 (9.64%)	—	14 (8.43%)	29 (16.20%)	—	23 (12.85%)	17 (10.18%)	—	17 (10.18%)
Severity of disease (The Patient Global Impression of Severity (PGIS) scale – recategorized into three groups) (*n* (%))									
None	56 (33.73%)	55 (35.03%)	68 (43.31%)	57 (31.84%)	55 (33.54%)	68 (41.46%)	61 (36.53%)	56 (35.44%)	63 (39.87%)
Mild and moderate	103 (62.05%)	95 (60.51%)	82 (52.23%)	106 (59.22%)	97 (59.15%)	86 (52.44%)	91 (54.49%)	88 (55.70%)	83 (52.53%)
Severe and very severe	7 (4.22%)	7 (4.46%)	7 (4.46%)	16 (8.94%)	12 (7.32%)	10 (6.10%)	15 (8.98%)	14 (8.86%)	12 (7.59%)
Japanese TSQM (Mean (SD))									
Effectiveness	51.27 (14.51)	51.13 (14.72)	52.87 (16.08)	50.09 (16.17)	50.30 (16.35)	52.39 (17.14)	51.50 (14.95)	51.44 (14.77)	52.07 (15.64)
Side effects	93.22 (15.73)	93.31 (15.71)	90.49 (19.49)	93.02 (19.15)	92.33 (18.74)	90.29 (20.10)	—	—	—
Convenience	61.08 (15.74)	61.32 (15.72)	61.04 (16.29)	52.89 (16.65)	53.18 (17.11)	52.64 (14.99)	58.08 (15.25)	58.12 (15.42)	57.88 (18.15)
Global satisfaction	57.79 (17.77)	57.83 (17.99)	57.19 (18.54)	52.98 (16.49)	53.56 (16.57)	52.69 (15.08)	57.27 (17.80)	57.01 (17.74)	55.74 (19.07)
Ceiling effect (*n* (%))									
Effectiveness	2 (1.20%)	—	—	3 (1.68%)	—	—	3 (1.80%)	—	—
Side effects	137 (82.53%)	—	—	153 (85.47%)	—	—	—	—	—
Convenience	10 (6.02%)	—	—	6 (3.35%)	—	—	9 (5.39%)	—	—
Global satisfaction	14 (8.43%)	—	—	4 (2.23%)	—	—	15 (8.98%)	—	—
Floor effect (*n* (%))									
Effectiveness	3 (1.81%)	—	—	3 (1.68%)	—	—	2 (1.20%)	—	—
Side effects	0 (0.00%)	—	—	1 (0.56%)	—	—	—	—	—
Convenience	0 (0.00%)	—	—	2 (1.12%)	—	—	0 (0.00%)	—	—
Global satisfaction	1 (0.60%)	—	—	3 (1.68%)	—	—	1 (0.60%)	—	—

Allocation of groups, ceiling effect and floor effect, and PGIS scores are presented in frequency of response (*n* (%)) and the scores for each domain of each Japanese TSQM version are presented as mean (SD).

Table 3 Reliability estimates of the Japanese TSQM versions

	Japanese TSQM-1.4	Japanese TSQM-II	Japanese TSQM-9
Internal consistency (Cronbach's alpha)			
Effectiveness	0.917	0.875	0.832
Side effects	0.808	0.805	—*
Convenience	0.721	0.895	0.675
Global satisfaction	0.782	0.874	0.762
Test-retest reliability (ICC (95% CI))			
Patients with no rescored PGIS change (n)	93	100	106
Effectiveness	0.672 (0.543–0.770)	0.596 (0.454–0.709)	0.625 (0.494–0.729)
Side effects	0.545 (0.385–0.673)	0.369 (0.187–0.527)	—*
Convenience	0.633 (0.493–0.740)	0.617 (0.479–0.725)	0.613 (0.479–0.719)
Global satisfaction	0.670 (0.540–0.768)	0.693 (0.575–0.782)	0.630 (0.500–0.733)

* Japanese TSQM-9 does not include a side effects domain
Results of internal consistency and test-retest reliability presented as Cronbach's alpha and ICC (95% CI), respectively, for each domain of each Japanese TSQM version.

Table 4 Item-to-domain correlation for each Japanese TSQM versions*

Item number		Effectiveness	Side effects	Convenience	Global satisfaction
Japanese TSQM-1.4	Q1	**0.910**	0.089	0.380	0.541
	Q2	**0.927**	0.073	0.405	0.540
	Q3	**0.902**	0.108	0.401	0.497
	Q4	0.090	**0.974**	−0.042	−0.051
	Q5	−0.046	**0.695**	0.534	0.179
	Q6	−0.030	**0.901**	0.460	0.244
	Q7	0.115	**0.768**	0.318	0.234
	Q8	0.115	**0.818**	0.119	0.026
	Q9	0.293	−0.053	**0.755**	0.363
	Q10	0.297	0.016	**0.767**	0.342
	Q11	0.328	−0.031	**0.859**	0.438
	Q12	0.449	−0.093	0.413	**0.839**
	Q13	0.397	0.035	0.257	**0.855**
	Q14	0.605	−0.031	0.605	**0.836**
Japanese TSQM-II	Q1	**0.942**	0.068	0.449	0.551
	Q2	**0.948**	0.078	0.476	0.570
	Q3	0.026	**0.699**	0.239	0.156
	Q4	0.333	**0.701**	0.311	0.327
	Q5	0.210	**0.897**	0.273	0.238
	Q6	0.239	**0.906**	0.109	0.192
	Q7	0.465	0.204	**0.842**	0.652
	Q8	0.399	0.223	**0.904**	0.712
	Q9	0.399	0.182	**0.930**	0.713
	Q10	0.488	0.126	0.680	**0.924**
	Q11	0.625	0.183	0.764	**0.952**

Table 4　Item-to-domain correlation for each Japanese TSQM versions* (continued)

Item number		Effectiveness	Side effects	Convenience	Global satisfaction
Japanese TSQM-9**	Q1	**0.869**	—	0.390	0.545
	Q2	**0.913**	—	0.486	0.577
	Q3	**0.800**	—	0.516	0.599
	Q4	0.361	—	**0.702**	0.429
	Q5	0.418	—	**0.840**	0.470
	Q6	0.328	—	**0.816**	0.464
	Q7	0.559	—	0.553	**0.868**
	Q8	0.343	—	0.316	**0.796**
	Q9	0.647	—	0.654	**0.866**

* Item-to-domain item correlations of ≥0.7 are shown in bold.
** Japanese TSQM-9 does not include a side effects domain.
Results of item-convergent validity and item-discriminant validity presented as item-to-domain correlation for each Japanese TSQM version.

Table 5　Summary Rasch fit statistics and psychometric criteria of the Japanese TSQM

Japanese TSQM version	Domain	Item residual mean (SD)	Person residual mean (SD)	Chi-square χ^2 (df)	Chi-square P	Reliability (PSI)	DIF*
Japanese TSQM-1.4	Effectiveness	−0.400 (0.966)	−1.324 (0.531)	6.050 (9)	0.735	0.869	No DIF
	Side effects	0.145 (1.141)	−0.383 (0.865)	11.269 (12)	0.506	0.818	No DIF
	Convenience	0.347 (0.724)	−0.390 (0.940)	3.707 (9)	0.930	0.79	No DIF
	Global satisfaction	0.443 (0.873)	−0.350 (0.985)	12.052 (9)	0.210	0.767	No DIF
Japanese TSQM-II	Effectiveness	−0.258 (0.055)	−0.658 (0.769)	0.888 (6)	0.989	0.823	No DIF
	Side effects	0.968 (1.474)	−0.553 (1.371)	11.955 (9)	0.216	0.506	No DIF
	Convenience	0.095 (1.266)	−0.492 (0.970)	3.274 (9)	0.952	0.905	No DIF
	Global satisfaction	1.492 (0.349)	0.117 (0.324)	3.821 (6)	0.080	0.622	No DIF
Japanese TSQM-9	Effectiveness	−2.323 (2.256)	−2.023 (2.035)	21.203 (9)	0.118	0.825	No DIF
	Convenience	−0.649 (1.856)	−1.124 (1.635)	6.777 (9)	0.660	0.707	No DIF
	Global satisfaction	0.266 (1.844)	−0.902 (1.244)	23.107 (9)	0.059	0.730	No DIF

* DIF was assessed for disease, sex, age group, education, time of administration, and mode of administration (oral or injectable).
Abbreviations: DIF: Differential item functioning, PSI: Person separation index
Results of psychometric analysis using Rasch measurement theory presented for each domain of each Japanese TSQM version.

good model fit, with mean (SD) of the item residuals and person residuals for all domains of all 3 Japanese TSQM versions ranging between −2.5 and 2.5. Moreover, none of the chi-square values obtained were statistically significant, supporting evidence for the unidimensionality of each domain.

The Person Separation Index (PSI), which supports reliability, had values above the accepted threshold of 0.7

for most domains of the Japanese TSQM versions. Moderate reliability was obtained from the Japanese TSQM-II global satisfaction domain (0.622) and side effects domain (0.506).

Assessment of DIF supported scale invariance between patient groups compared by disease, sex, age group, education. Moreover, scale invariance was supported by comparing the time of administration, further supporting test-retest reliability. Finally, DIF scale invariance was assessed for all domains of the Japanese versions of TSQM-1.4, the TSQM-II, and the TSQM-9 between mode of administration (oral or injectable), further supporting similar psychometric properties of the TSQM for different modes of administrations.

IV. DISCUSSION

This study consisted of a cultural adaptation and psychometric assessment of the newly translated TSQM versions. The Japanese TSQM was updated through re-translation and cognitive debriefing interviews to include a more patient friendly wording that can used regardless of the mode of administration.

The translation and cultural adaptation of all three versions of the TSQM followed standard guidelines and were cognitively debriefed with Japanese patients. The Japanese terminology was amended to adapt to the Japanese context based on the eight interviews. For example, questions on confidence level and level of certainty were changed as patients expressed difficulty in responding.

Internal consistency was assessed using Cronbach's alpha and supported a high internal reliability for all domains within Japanese TSQM-1.4, Japanese TSQM-II and Japanese TSQM-9, except for the convenience domain of the TSQM-9 which had a moderate reliability. The score for Cronbach's alpha was lower than the previously conducted validation study of Japanese TSQM, which had internal consistency values of 0.8 or higher.

Test-retest reliability assessed using ICC for patients who did not change disease severity between both times of administration. Moderate correlations were obtained for all versions of the Japanese TSQM, except the side effects domain of Japanese TSQM-II, which had a low value of 0.369. The possible explanation for the low ICC estimate shown for the side effects domain of Japanese TSQM-II lies in the selection of the anchor criteria to define the absence of change from the first and second time of administration. The PGIS was selected as an anchor to assess disease severity as is typically done and has been recommended by regulatory bodies[19]. The PGIS was reclassified into three categories to prevent any potential influence of recall bias between the times of administration. The TSQM is however a measure of treatment satisfaction, which can very well be related to disease severity, but is probably more directly related to the effectiveness of treatments in managing symptoms, regardless of disease severity. In retrospect, the PGIS may therefore not be a fit-for-purpose anchor. A potential alternative would have been to create a satisfaction-specific version of the PGIS (asking whether treatment satisfaction had changed, rather than disease severity).

ICC values for the assessment of test-retest reliability were lower than the original TSQM, in particular for Japanese TSQM-9[4]. This could be due to the differences between the sample populations used to assess each TSQM versions. Migraine is a disease that repeatedly undergoes exacerbations and remissions and may perhaps have less stable treatment satisfaction than some of the therapeutic areas that were selected for the original TSQM validation.

High intra-test validity of Japanese version of TSQM was shown for all domains, supporting the structural validity and the scoring frame of the Japanese TSQM version.

A strength of this study is the use of RMT, in addition to CTT Analyses, to assess the psychometric properties of a measure. The paradigm of CTT proposes that observed scores consist of an error-free true score component and a random error component[20]. Over the recent decades, measurement in health have undergone a transformation. With the advent of modern test theory, new perspectives have emerged[21]. Rasch measurement theory, models the response to an item as a probabilistic function of person and item parameters (and random error)[22]. RMT not only provides a better basis for connecting the conceptual theory of the measure and empirical statistical evidence, it also provides an opportunity to generalize psychometric properties of measures beyond the sample population selected for its assessment.

High PSI scores were supported for all domains of the TSQM versions, except for the global satisfaction domain of the Japanese TSQM-9 and the side effects domain of the Japanese TSQM-II. The reason for these lower values could very well be the lack of variability obtained from the sample population used, most patients reporting high level of treatment satisfaction. One important condition for stable psychometric assessment, regardless of the theory used, is a variable sample population. Moreover, there are only two questions within the global satisfaction domain of TSQM-9 and the side effects domain of the Japanese TSQM-II. This, coupled, with the lack of variability in the sample, could very well affect the PSI score.

This limitation in sample variability is related to the sampling strategy. Patients were self-referred through a commercial patient panel company. This sampling bias limits attrition to more severely affected individuals, who may be more severely affected and may not be satisfied with their current treatment regimens as not effectively controlling their symptoms or due to important side effects. The very high scores obtained throughout the side effect domains in all Japanese versions, and the high ceiling effect of 85% for these domains, supports that lack of variability.

An additional limitation that could contribute to this lower variability seen in treatment satisfaction in the sample relate to the sample sizes for the different subgroups of patients. Of relevance, the therapeutic area and mode of administration were deemed to be important sampling characteristics to assess the psychometric properties of the TSQM versions. Although useful exploratory work using Rasch analysis with a small sample is feasible, item calibrations or person measures stable within 0.5 a logit and 99% confidence need sample sizes of 108 to 243 per subgroup of interest (a sample size of 150 individuals per subgroup is recommended for most purposes). This sample size was not achieved for the RMT assessment[23~25].

Future work should focus on reassessing the psychometric properties of the Japanese TSQM measures with a larger sample of more varied sample of patients in terms of treatment satisfaction, and potentially varied target therapeutic areas.

Despite the limited variability of the sample, differential item functioning assessment supported scale invariance when subgroups of patients by disease, sex, age group, education, time of administration, and mode of administration (oral or injectable). The absence of DIF by time of administration and mode of administration further supports test-retest reliability. The absence of DIF by mode of administration is new evidence, supporting the use of the TSQM for various modes of administration.

V. CONCLUSION

The Japanese versions of TSQM (TSQM-1.4, TSQM-II, and TSQM-9) are measures with sufficient reliability and validity, which can be used regardless of disease and mode of administration. Additional studies are needed to determine the extent of the psychometric properties for modes of administration other than oral or injectable.

AY, AM, KI, LPW, and TA are employees of IQVIA Solutions Japan G.K., who are the funders and is the Japanese branch of IQVIA which owns the TSQM. AMR is an employee of IQVIA which owns the TSQM.

This work was fully funded by IQVIA Solutions Japan G.K. The study was fully conducted by the funder.

All authors were involved in the development of the concept and design of the study. AY and AM contributed to data acquisition. KI, and AMR, analyzed the data, and all authors were involved in interpretation of the data. AY, AMR, and KI were the primary writers, and all authors read and approved the final manuscript.

We thank YL and AT from IQVIA Solutions G.K. for participating in the forward translation and analysis. We also thank patients who participated in this study and gracefully gave their time for us to assess the cultural adaptation of the Japanese TSQM versions.

References

1) Barbosa CD, Balp MM, Kulich K, et al. A literature review to explore the link between treatment satisfaction and adherence, compliance, and persistence. Patient Prefer Adherence 2012; 6: 39–48.

2) Atkinson MJ, Sinha A, Hass SL, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. Health Qual Life Outcomes 2004; 2: 12.

3) Atkinson MJ, Kumar R, Cappelleri JC, et al. Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. Value Health 2005; 8(Suppl 1): S9–S24.

4) Bharmal M, Payne K, Atkinson MJ, et al. Validation of an abbreviated Treatment Satisfaction Questionnaire for Medication (TSQM-9) among patients on antihypertensive medications. Health Qual Life Outcomes 2009; 7: 36.

5) Sakai F, Igarashi H. Prevalence of migraine in Japan: a nationwide survey. Cephalalgia 1997; 17: 15–22.

6) The Japan Diabetes Society. [Clinical Guidelines of Diabetes 2024] (in Japanese). 2024. http://www.jds.or.jp/modules/publication/index.php?content_id=4 (accessed May 24, 2024).

7) Rakuten Insight. [Disease panel] (in Japanese). https://insight.rakuten.co.jp/download/disease_panel.pdf (accessed May 24, 2024).

8) Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005; 8: 94–104.

9) Quadri N, Wild D, Skerritt B, et al. PRM153. A literature review of the variance in interval length between administrations for assessment of test retest reliability and equivalence of pro measures. Value Health 2013; 16: A40–A41.

10) SOUKEN. [Ethical Commitee] (in Japanese). https://www.souken-lab.co.jp/service/ethics/#gsc.tab=0 (accessed April 05, 2024).

11) Mapi Research Trust. PATIENT GLOBAL IMPRESSION OF SEVERITY (PGIS). 2023. https://eprovide.mapi-trust.org / instruments / patient-global-impressions-scale-change-improvement-severity (accessed May 24, 2024).

12) Steiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. 6th ed. Oxford, UK: Oxford University Press. 2024; 1–480.

13) Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959; 56: 81–105.

14) Regnault A, Balp MM, Kulich K, et al. Validation of the Treatment Satisfaction Questionnaire for Medication in patients with cystic fibrosis. J Cyst Fibros 2012; 11: 494–501.

15) Trask PC, Tellefsen C, Espindle D, et al. Psychometric validation of the cancer therapy satisfaction questionnaire. Value Health. 2008; 11: 669–679.

16) Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2018; 48: 1273–1296.

17） Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016; 15: 155–163.
18） Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. Qual Life Res 1993; 2: 441–449.
19） U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research（CDER）, Center for Biologics Evaluation and Research（CBER）, Center for Devices and Radiological Health（CDRH）. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints For Regulatory Decision-Making Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. 2023. https://www.fda.gov/media/166830/download（accessed May 24, 2024）.
20） Lord FM, Novick MR（eds）. Statistical Theories of Mental Test Scores. Menlo Park, CA, USA: Addison-Wesley Educational Publishers Inc. 1968; 1–592.
21） Embretson SE, Reise SP. Item Response Theory. New York, NY, USA: Psychology Press. 2013; 1–384.
22） Andrich D, Pedler P. A law of ordinal random error: the Rasch measurement model and random error distributions of ordinal assessments. Measurement 2019; 131: 771–781.
23） Bamber D, van Santen JPH. How many parameters can a model have and still be testable. J Math Psychol 1985; 29: 443–473.
24） Wright BD, Douglas GA. Best test design and self-tailored testing. Research Memorandum No.19. Chicago, IL, USA: Statistical Laboratory, Department of Education, University of Chicago. 1975; 1–55.
25） Azizan NH, Mahmud Z, Rambli A. Rasch rating scale item estimates using maximum likelihood approach: effects of sample size on the accuracy and bias of the Estimates. Int J Adv Sci Technol 2020; 29: 2526–2531.

（Received 2025.4.2; Accepted 2025.8.22; Advance Publication 2025.12.17）