**Refining and Testing Four Bolt-ons in Patients with Sleep Disturbances using Mixed Methods.**

**Authors:** Ai-Ping Chua[1], Aureliano Finch[2], Shaffinaz Rahman[1], MaryAnne[1], Kaayathiri Ravichandran[1], Emelita Montaniel[1], Nan Luo[3]

[1.]National University Health System, Singapore [2.]EuroQoL Office [3.]National University of Singapore, Singapore

***Background:*** The EQ-5D-5L may not adequately capture the health-related quality of life (HRQoL) of patients with sleep disturbances. We examined whether the addition of four sleep-related bolt-ons could enhance the psychometric properties of EQ-5D-5L.

***Methods*:** We employed a mixed methodology involving in-depth interviews with 23 patients and clinicians to test the face validity of "*sleep*", "c*ognition*", "*tiredness*" and "*relationship*" and refine the bolt-ons. We then administered the four bolt-ons, appended to EQ-5D-5L, together with three condition-specific patient-reported outcome measures (cPROMs), i.e., Sleep Apnea Quality of Life Index (SAQLI), the Functional Outcome Sleep Questionnaire (FOSQ) and the Pittsburgh Sleep Quality Index (PSQI) questionnaires during two clinical visits. We compared ceiling effects and construct validity by testing *a priori* hypotheses in relation to the cPROMs and polysomnographic characteristics via correlation and areas under the curves (AUC) analyses, respectively. We examined responsiveness among "*treated/improved*" participants using standardized response means (SRM) and AUC analysis, and reliability among "*untreated/no change*" participants using intra-class correlation coefficient (ICC) or Cohen's Kappa (k).

***Results:*** 110 participants (mean [SD] age: 47[13]) were recruited of 90 returned for their review assessments (mean [SD] interval: 2.2 [2.1] months). The absolute ceiling reduction was 42.7%. The bolt-ons were better correlated with cPROMs and possessed higher discriminatory power and responsiveness, with comparable reliability to EQ-5D-5L. A combined module of the four bolt-ons provided better results than individual bolt-ons.

***Conclusions*:** Adding the four sleep-related bolt-ons improved EQ-5D-5L's psychometric properties among patients with sleep disturbances. (246 words)

## INTRODUCTION

The EQ-5D is one of the most commonly used HRQoL measures worldwide. It comprises of five single-items dimensions, "mobility", "self-care", "usual activities", "pain/discomfort", and "anxiety/depression". The descriptive system is accompanied by a visual analog scale (i.e. EQ VAS), which can be used to assess overall health. The EQ-5D scoring system is based on preferences, which allow for quality-adjusted life years (QALY) calculation and use in health economic evaluation. The large success of the instrument is mostly due to this setting of application, albeit it is also commonly used in other contexts such as population health surveys and clinical[1-3].

Sleep disorders such as sleep apnea (SA) are a rapidly growing problem, afflicting millions of people worldwide[4]. They are associated with adverse health outcomes and exert significant socioeconomic burdens on societies[5-6]. In addition, sleep disturbances exert significant transversal impact across diseases. Disturbed sleep with awake consequences of sleepiness, low energy levels and fatigue are common problems reciprocally associated with chronic ailments across multiple therapeutic areas (e.g., cancers, psychiatric and mental health problems, stroke and neurodegenerative conditions, heart diseases, rheumatological conditions, chronic kidney failure and chronic obstructive airway diseases) leading to poorer health-related quality of life (HRQoL) when assessed by patient-reported outcome measures (PROM)[7-14].

The EQ-5D, as a generic PROM and preference-based measure, is found generally to be valid and responsive across multiple health conditions, including diabetes, osteoarthritis, rheumatoid arthritis, cancer, psoriasis, lupus, Parkinson's disease, multiple sclerosis, dementia, stroke, lung disease and haemophilia, although there is a call for more responsiveness studies and robust psychometric testing and reporting[15-16]. Yet, there is also evidence showing poor responsiveness in patients affected by sleep disorders, particularly in the context of its use as a PROM[17]. For example, clinical studies have found that the EQ-5D may not adequately capture the severity of the condition, and it is not able to detect clinical improvement[18-21]. There is also evidence reporting that sleep/ fatigue are important dimensions missing from the EQ-5D but included in other generic HRQoL and preference-based measures.

Bolt-ons are dimensions that can be appended to the descriptive system of the EQ-5D. This may allow for a more comprehensive description of patients' health, thereby making it a more valid PROM for specific populations[22]. Some bolt-ons had been shown to increase the sensitivity of the EQ-5D, for example, hearing and cognition significantly influenced the valuation of health preference in the general population, respiratory bolt-on in chronic

obstructive pulmonary disease and vision bolt-on in cataract improved its known-group validity whereas informativity and variance were better captured by the addition of "itch" in burn patients, and "skin irritation"/"self-confidence" bolt-ons in psoriasis, respectively[23-27]. While bolt-on research started two decades ago, there are still pending questions, which include the availability of different bolt-on versions developed using different approaches. For example, there are at least four different versions of sleep bolt-on (Appendices.docx Table 1)[28-31]. The earliest version "*difficulties with sleep*" developed by Perneger et al was tested in a general population sample together with other bolt-ons and was based on a review of the existing general health status and utility questionnaires, with no prior qualitative work[29]. This was followed by Yang et al who used "*problems with sleep*" and tested it in a valuation study[30]. Subsequently, there were two studies of bolt-ons on EQ-5D-5L[28, 31]. One of these studies was conducted in a general population sample in South Korea with five-level response descriptors[31]. The other was a discrete choice experiment testing the impact on preferences of using "*problem sleeping*" and four other bolt-ons (relationships, cognition, energy, hearing) in a large general population sample[28].

There are some limitations in existing sleep bolt-on which had been developed. First, Mulhern et al recently reported a set of criteria for developing bolt-on[32], but none of the existing bolt-on studies meet all the criteria, with some studies fulfilling varying number of recommendations[28, 31, 33-34]. Second, except for one study where focus groups were used to collect input on the wording among patients and member of the public, none of the other sleep bolt-ons was developed using patients' input[28]. Third, most of the bolt-on research for sleep developed single item bolt-ons, except for one study which developed multiple items for energy and sleep[28]. Lastly, none of the studies tested sleep bolt-on in any patient group in the hospital setting who suffered from sleep problems or involved the healthcare professionals, thus findings may not be generalizable to patient populations with sleep disturbances

A more general issue with bolt-on research to date is that studies have generally taken a dimension-centric approach, by developing or testing bolt-ons individually. There is paucity of studies looking comparatively at bolt-ons in a chosen population. Among the few, Finch and colleague tested cognition, vision, hearing, social relationships and energy impact on a proxy of HRQoL, and on preferences[23,28]. They found energy/vitality and relationships had the largest impact on HRQoL followed by cognition/speech, vision and hearing[23]. Also, they found the largest shift in preference for hearing, cognition and relationship, followed by energy and sleep for level 5 of the bolt-on[28]. In Rencz et al, it was found that multiple bolt-on items may be

useful to improve the EQ-5D properties in a general population health survey in Hungary[33]. The authors showed that other than sleep, cognition, social relationships and tiredness bolt-ons significantly correlated with the corresponding items from prominent generic PROMs such as Patient-Reported Outcomes Measurement Information System and SF-6D[33].

Further empirical studies are therefore warranted to confirm the properties of different bolt-ons in patient specific populations, including those affected by sleep disorders. This study takes the approach of investigating possible refinements for existing bolt-on for use in clinical setting and as PROMs. Specifically, we investigated how sleep and related bolt-ons of energy, cognition and relationship perform in patients with sleep disturbances. We first ascertained the importance of these four dimensions of "*sleep*", "*energy*", "*cognition*" and "*relationship*" in populations affected by sleep disorders, and the ease of understanding and responding to these bolt-ons for sleep patients and clinicians, using qualitative evidence. We use this evidence to refine the bolt-ons, and we subsequently test the bolt-ons psychometric properties among chronic disease patients with sleep disturbances. For this study, we selected four existing bolt-ons (sleep, energy, cognition and relationship), as these had been developed in the same study, had involved patients and members of the general public for their development, and had been tested using larger sample sizes compared to other studies[28]. They also adhered to the formatting style of EQ-5D-5L, with five-level response descriptors of "*no/slight/moderate/severe/extreme problems*".

## METHODS

The study used a mix-method design. It was carried out in two phases, both taking place in a tertiary public hospital in Singapore, from the 28th of February 2023 to the 30th of April 2024.

### *Qualitative component*

In the first qualitative phase, we assessed the face validity of the four candidate bolt-ons ("*sleep*", "*cognition*", "*energy*", and "*relationships*") using the COREQ (Consolidated criteria for Reporting Qualitative research) checklist[34]. We had chosen these four bolt-ons to be further tested as they were key dimensions impacting HRQoL of patients with sleep

disorder, in a qualitative evidence synthesis[35]. Permission was earlier sought from the lead author to use and modify (if necessary) the bolt-ons, and the IP of these bolt-ons and possible subsequent modifications was transferred to the EQ Foundation.

We purposively sampled both English proficient patients with different chronic health conditions and sleep disturbances, and clinicians who managed such patients routinely, using a convenience sampling technique. In addition, we also included participants without known chronic illnesses to seek their views given that sleep disturbances can also occur in healthy individuals and the EQ-5D is also intended for this population. While we did not use any quotas, the recruitment aimed at identifying participants based on the abovementioned characteristics of interest to ensure a good mix. Each recruited participant underwent an in-depth semi-structured face-to-face interview session in English by trained personnel (AP and SAR/MA/EM/KR), using a standard protocol incorporating a topic guide & field notetaking over a 30 to 40 minutes time interval. The interview guide was pre-tested in three volunteers and within the research team to clarify open-ended and concept elicitation questions and enhance consistent and neutral cognitive debriefing style.

The interviews consisted of three parts. In part one, the participants were first invited to share what health and being unhealthy mean to them, to elicit their definitions of "health" and "illness". They were then requested to self-complete the EQ-5D-5L questionnaire. After completing the questionnaire, they were invited to mention if there were important aspects of health or illness that were not already covered by EQ-5D-5L. In part two, responders were presented with the four bolt-on items and probed for their views and understanding of the items. Their opinions and suggestions were also sought with respect to the phrasing of the bolt-on' labels and descriptors. Probes were used to solicit their views on the appropriateness, acceptability and relevance of the bolt-on items to their conditions, and the ease of understanding and answering the bolt-on items. Lastly in part three, the participants were asked to rank the nine items (five EQ-5D standard items and the four bolt-on items) in the order of one to five ($r_1$, $r_2$, $r_3$, $r_4$ & $r_5$) with one indicating the highest importance and relevance to their health using a two-step approach. In the first step, they were first instructed that they could assign the same number ranking to two or more items if they felt that the items are of equal importance including all nine items. In the second step, they were then requested to choose only five out of the nine items and assign a unique ranking number (one to five) to each of the five items. Participants were further probed for their reasons behind their rankings. Sound recording of the interviews was carried out with consenting participants.

*Analyses*

We applied framework analysis to analyze the data. Interpretation of findings was guided by the World Health Organization (WHO)'s definition of health[36]. Interviews were manually transcribed and participants pseudonymized, and recorded field notes on part one of the interview were reported. The remaining content of the audio-transcripts and manually recorded field notes were aggregated, summarized and presented for iterative rounds of discussion among team members until consensus was achieved on the final version of bolt-ons for psychometric testing in the second phase of the study. We pilot-tested the final versions among ten chronic disease patients with sleep disturbances in the clinic prior to commencing the second phase.

***Quantitative component***

The second quantitative phase of the study was for assessing the psychometric properties of the final versions of bolt-on items alongside the EQ-5D-5L. We recruited consecutive patients referred to the outpatient sleep disorder clinic or inpatient sleep laboratory of the study site. Patients were referred for management of their sleep disturbances and prescribed an overnight polysomnography as part of their routine clinical care. The inclusion criteria were (1) age ≥ 21 years, (2) clinically confirmed diagnosis of sleep disturbance or disorder, (3) at least one underlying physician-diagnosed chronic illness, (4) adequate cognitive and communicative (visual, aural & linguistic) ability, (5) English literate, and (6) informed consent.

During the *index visit*, each participant answered a set of standard questionnaires, prior to undergoing a polysomnography. The questionnaire set comprised the EQ-5D-5L followed by the four bolt-on items and the EQ VAS, three condition-specific PROMs (cPROM) comprising of the Sleep Apnea Quality of Life Index (SAQLI), the Functional Outcome Sleep Questionnaire (FOSQ) and the Pittsburgh Sleep Quality Index (PSQI) questionnaires, and questions relating to their socio-demographics and health conditions. All questionnaires were administered by trained research staff using the Research Electronic Data Capture (REDCap®) platform, which had previously been tested among five patient volunteers for feasibility and quality check. Polysomnographic data of the recruited participants were also collected. All participants signed the informed consent and the approval to conduct the study was obtained from the Domain-Specific Review Board (2022/00385). During the *review visit*, participants who returned for follow-up consultation for their sleep disturbance, underwent a similar procedure as their *index visit* and were administered the same set of questionnaires.

*Instruments*

The ***SAQLI*** is a sleep apnoea specific QoL instrument consisting of 35-items that capture the adverse impact of sleep apnoea on four subscales: daily functioning, social interactions, emotional functioning, and symptoms. Items are scored on a seven-point Likert scale, ranging from "all of the time" to "not at all"[37]. The item scores for each subscale are averaged to generate a total score for each subscale. The four subscale scores are then averaged to generate a global score ($SAQLI_{tot}$) which ranges between one and seven. Higher scores represent better HRQoL. The instrument recall period is set at "past four weeks". The instrument also contains a fifth subscale and two visual analogue scales (VAS) which are situationally administered in patients who had received treatment for sleep apnoea to capture potential negative impacts exerted by treatment side-effects on quality of life. The two VAS reflect positive and negative impacts from sleep apnoea treatment respectively: each range from 0 ("no impact") to 10 ("extremely large impact"). The fifth subscale and VAS scores can be used to compute net QoL impact with treatment.

The ***FOSQ-10*** is an abbreviated version of the original FOSQ, a disease-specific QoL questionnaire assessing the impact of sleepiness on awake activities which are classified into five subscales comprising of 1) activity level (two items), 2) vigilance (three items), 3) intimacy and sexual relationships (one item), 4) general productivity (two items), and 5) social outcomes (one item), and scored on a four-point Likert response format with "extreme difficulty" and "no difficulty" being the extreme answers[38]. The five subscales will yield five mean scores which are then averaged and multiplied by five to produce a total score ($FOSQ_{tot}$) ranging from five to twenty, with higher scores indicating better functional status. The recall period is not specified although instruction suggests current difficulty.

The ***PSQI*** is a sleep quality questionnaire with 19 self-rated questions of both Likert- and open-ended types assessing across seven domains over past month period: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medications, and daytime dysfunction[39]. The answers to the open-ended questions are converted to scaled scores per developer's scoring guidelines. Each domain generates a score ranging from 0 to 3 with higher scores indicating more severe sleep disruptions. The domain scores are then summed to produce overall sleep quality score ($PSQI_{tot}$). A sleep quality score of $\leq 5$ or $>5$ defined "good" vs "poor" sleeper, respectively.

*Analyses*

Data of the index visit were used to compare the psychometric performance (distributional properties, convergent validity, discriminatory power and explanatory power) of the standard EQ-5D-5L (5L) and its five modified versions, that is, the four bolt-on item tagged to it individually plus a bolt-on module including all four items. Distributional properties were compared by looking at both ceiling (absolute and relative) and floor effects. Ceiling effects described the best possible health state and were determined as the proportion of participants who reported "*no problem*" in all EQ-5D items, that is, "*11111*" descriptive health profile for the five core items and "*111111111*" if there are nine (five core and four bolt-on) items. A threshold of below 15% was set for ceiling effects as per COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN)'s recommendations[40-41]. On the contrary, floor effects described the worst possible health state and were determined by the proportions of participants who reported "extreme problem/unable" in all items, that is, "*55555*" descriptive health profile for the five core items and "*555555555*", if there are nine items. Additionally, we also examined the proportions of responses with top twenty health profiles. We then evaluated the convergent validity of individual items among themselves and with the selected cPROMs. The latter were assessed via correlations between the level sum score (LSS) of each version against $SAQLI_{tot}$, $FOSQ_{tot}$ and $PSQI_{tot}$ using Spearman's rank correlation coefficient (ρ). Weak, moderate and strong correlations were defined as ρ 0.20 to less than 0.35, 0.35 to less than 0.50 and ≥0.50, respectively[42].

Following that, we compared their discriminatory power by assessing their ability to distinguish participants known to differ in the severity of their disease conditions based on clinical parameters, using c-statistics (c-stat) that were computed from area under the receiver operating characteristic curve (AUC) analyses and processing LSS of the five versions as continuous variables. Seven *a priori* selected and binarized clinical (polysomnographic) parameters were used in the analyses as the outcomes: total sleep time (TST), sleep efficiency (SE), sleep latency (SL), arousal index (AI), apnea-hypopnea index (AHI), oxygen desaturation index (ODI), and percentage TST with pulse oximetric oxygen saturation <90% (%TST SpO2<90). A higher c-stat value denotes a greater discriminatory power. Cut-offs were guided by severity threshold used clinically for AHI, ODI and percentage TST, while cut-offs for the rest of the selected parameters were determined by their respective medians[43-44]. Lastly, we compared the explanatory power of the bolt-on items using linear regression models for the variation in the EQ VAS score. This was performed by extending the core model regressing EQ VAS over the five standard EQ-5D items using their LSS and including each added bolt-

on -individually which then generated an adjusted R-square value ($R^2$). Greater explanatory power is represented by a higher $R^2$.

We used the data collected from both index visit and review visit to compare the reliability and responsiveness of the EQ-5D-5L +S, +E, +C, +R, and +SECR. We allocated the participants who showed up for their routine review visit and completed the questionnaires into two categories ("treated" vs. "not treated) based on their treatment status, i.e., whether they had received the definitive treatment for their sleep disturbance, and HRQoL statuses ("i*mproved*" vs, "*no change/deteriorated*") defined by the cPROM (SAQLI, FOSQ, PSQI) scores thresholds. The thresholds of SAQLI, FOSQ and PSQI were set at 0.5, 1 and 2, respectively[45-49]. Participants were assigned "*improved*" if their review visit cPROM scores compared to their index visit scores increased by more than 0.5 for SAQLI and more than 1 for FOSQ and for PSQI, if the scores decreased by more than 2; otherwise, they were assigned "*no change/deteriorated*". Participants whose scores did not change (increase or decrease) at or beyond the set thresholds were assigned to "*no change*" category. Both "*treated*" and "i*mproved*" categories were used to assess responsiveness of the LSS for each EQ-5D version, by applying area under the receiver operating characteristic curve (AUC) analyses to generate c-statistics (c-stat) and by computing the standardized response means (SRM). Both "*untreated*" and "*no change*" categories were used to assess reliability with weighted Cohen's Kappa (k) for each EQ item score and intra-class correlation coefficient (ICC) for LSS. ICC and $\kappa$ values of 0.4 to $< 0.6$ and $\geq 0.6$ were considered moderate and good agreement, respectively[50-51].

We summarized all descriptive data with normal distribution in mean (standard deviation, SD) and those with skewed distribution in median (interquartile range, IQR). Pearson's chi-square (x2) statistical testing was used for comparing proportions whereas unpaired t-test / Mann-Whitney U, one-way analysis of variance (ANOVA) / Kruskal-Wallis test, or paired t-test / Wilcoxon's Sign Rank test were used for comparing continuous variables. We used R Statistical Software (version 4.3.0; R Foundation for Statistical Computing, Vienna, Austria) to perform all statistical analyses. A p-value of less than 0.05 specified statistical significance.

## RESULTS

### *Qualitative component*

Overall, 23 participants comprising of 12 chronic disease patients (p), 3 healthy individuals (h) and eight clinicians (c) (age ranges of 26 -78, 21-76, and 30- 65 years) completed the interviews. Characteristics of the participants including their health states are presented (Appendices.docx Table 2).

In the first part of the interviews, some participants generally conceptualized HRQoL as related to physical health, usual activities and pain. For example, PF4, a male stroke patient, stated that health is *"being independent, mobile and able to take care of your own needs and do life daily routine and basic activities"*, while *PF12*, a female atopic dermatitis patient, that being healthy means being *"free from physical and mental illness, from bodily discomfort and pain…able to breathe easily, not having to suffer constantly from itch, stuffy nose".* Some participants evidence the importance of mental health. For example, PF14, a female participant, stated that being healthy means *"not having to worry and feeling overwhelmed by stress all the time…not having further panic attack and being able to feel happy".* Other themes emerging from the interviews fitted the "sleep (S)", "cognition (C)" (mental), "energy/fatigue/tired/stamina (E)" and "relationships/interactions/community participation (R)" (social) dimensions, as shown in the analysed transcripts 9 (12 patients, one healthy participants and six clinicians), 17 (nine patients, one healthy participants and seven clinicians), and 12 (seven patients, three healthy participants and two clinicians) participants, respectively. We present key examples of the illustrative quotes retrieved for additional dimensions supporting the four bolt-ons (Appendices.docx Table 3).

In part two of the interview, all the participants were able to understand and respond to the bolt-on questions when administered and explained by the interviewer. They also found the four types of dimensions to be relevant to patients with sleep disturbances. When probe to provide recommendations to further improve the clarity of the items, suggestions were given on the label and phrasings of the four initial candidates.

For sleep, participants highlighted that "… *problems sleeping*" may be subject to interpreting the items as solely "*difficulty falling asleep*", which may not include the broader aspects of sleep. Participants suggested to replace this with "*problems with sleep*". Participants also suggested to list examples of specific problems with sleep for reference and clarity (Appendices.docx Table 3).

For energy, there were few synonymous terms raised by participants conveying the same concept such as *"tired" "fatigue" "low vitality" "sluggish"*. However colloquially, most

participants used and preferred the term "*tired*" which was thought to be simpler to understand and most expressed.

For cognition, most patient participants were not familiar with the term in the bolt-on head label, however they were able to come up and elaborate on different components when they were informed that it referred to mental abilities. Most patient participants raised "*memory*" and/or "*focus/concentration*" as key components constituting "cognition" or "mental abilities". They were also able to understand and answer the cognition item without need for prompting or clarification after reading the response level descriptors.

For relationships, participants suggested adding primary relationship types, such as family and friends to better illustrate social relationships.

In the ranking exercise, , 14 participants (p=8, c=6) expressed difficulty with the ranking exercise attributed to perceived equipoise and/or proximate inter-connectedness among the nine items, therefore ranking them equally. Among patient participants, "*PD*" (n=4), "*UA*" (n=4), "*MO*" (n=4) and "*S*" (n=3) were most frequently endorsed as the most important items, with "*S*", "*E*", "*C*" & "*R*" in tie (n=2) second place. Among clinicians, "*PD*" (n=6), "*SC*" (n=5), "*C*" (n=3) & "*AD*" (n=3) were considered the first, second, third, fourth and fifth most important dimensions. We tabulated the results of the ranking exercise using both approaches in Appendices.docx Table 4.

Based on the suggestions provided by the participants, the team further modified the label and phrasings of the four initial candidate bolt-ons. For sleep, we synthesized the examples of sleep issues elicited in our exploration to "*difficulty falling*", "*staying asleep*" and "*waking up unrefreshed*". For parsimony, these were positioned within parenthesis in the label. To ensure consistency with the EQ-5D, the label "*sleep problems*" was used in place of "*sleep*", with response descriptors being '*no/slight/moderate/severe/extreme sleep problems'*.

For energy, we adopted an existing "tiredness" bolt-on in the literature given "tired" was the more frequently preferred term in colloquial English (Singapore).

For cognition, the team replaced the level descriptors with "memory" and "concentration". As all participants had no difficulty responding to the specified level descriptors, the team decided to retain "cognition" in the bolt-on head label.

For relationships, the team further added "*family and friends*" within the parenthesis of the bolt-on label as some examples of key relationship types.

Therefore, the final four bolt-on versions were: (1) **S** - "***sleep problems*** *(e.g., difficulty falling or staying asleep, waking up unrefreshed, etc.)*" with "*no/slight/moderate/severe/extreme sleep problems*"; (2) **C** - "***cognition***" with "*no/slight/moderate/severe/extreme problems with memory or concentration*"; (3) **T** - "***tiredness***" with "*not/slightly/moderately/severely/extremely tired*"; (4) **R** - "***relationships*** *(e.g., family, friends, etc.)*" with "*no/slight/moderate/severe/extreme problems with social relationships*". During the pilot testing, patients provided affirmative responses to the appropriateness, saliency, acceptability, comprehensiveness, and comprehensibility of the four final bolt-on items. Consequently, the final six EQ versions used for phase 2 quantitative testing were 5L (standard EQ-5D-5L), +S, +C, +T, +R (5L tagged with each of the four single developed bolt-on), and +SCTR (5L tagged with the "module", i.e., the four bolt-ons bundled).

***Quantitative component***

We recruited 110 patients. Their mean age was 47 years old, with 76.4% being males and 29.1% being of non-Chinese ethnicity (Malay, 21.8%; Indian, 4.6%; Others, 2.7%). 81.8% of participants were actively employed, while 84.6% resided in public housing, and vast majority received formal education (99.1%). Their clinical characteristics are presented in table 1. Proportions of patients who reported full health, that is, "*no problems*" in all the EQ-5D items, was 42.7%. When considering the EQ-5D +SCTR, all participants reported some problems. We summarized the ceilings and health state profiles for all EQ versions (Table 2) and item-level responses (Appendices.docx Fig. 1).

Table 3 showed a correlation matrix of the items/EQ VAS. All bolt-ons were moderately correlated with FOSQ, PSQI and SQALI, while +R was moderately correlated with FOSQ, but weakly with PSQI and SQALI. Correlations of all bolt-ons were higher than the correlation of the EQ-5D-5L i.e., weak with the three cPROMs. When all bolt-ons were considered simultaneously, the highest correlations were reported with the three cPROMs (Table 4). Also known-group validity of +S, +C, +T, +R and +SCTR were all higher than known group validity for the EQ-5D-5L, with +S, +T and +SCTR producing significantly higher mean c-stat (table 4). The adjusted $R^2$ generated in the multiple linear regression model were 0.114, 0.166. 0.112. 0.136, 0.159 and 0.200 for 5L, +S, +C, +T, +R and +SCTR, respectively (all p-values<0.001).

Ninety participants returned for their second interview at 2.2 (mean) ± 2.1 (SD) months. Their mean age was 47 years old (SD 13). 77.8% of them were males and 25.6% were of non-Chinese ethnicity. Among these participants, 64.4% received definitive treatment for their sleep disturbance, and 48.9%, 48.9% and 62.2% had improvement in their HRQoL status based on SAQLI, PSQI and FOSQ scores, respectively; they were utilized for responsiveness testing.

For responsiveness, both SRM and c-stat results showed similar trends with +SCTR, +S, +T, +C and +R (in order of declining magnitude) producing larger effect sizes than 5L in both the "*treated*" and SAQLI- and FOSQ-based "*improved*" groups (Table 5). In the PSQI-based improved group, SRM and c-stat values in +T exceeds marginally those of +S (table 4).

For reliability, approximately one-third (35.6%) of participants had not undergone their treatment. Each of them was interviewed at one month of their index visit. Among those participants who received treatment, 42.2%, 41,1% and 32.2% were categorized as "*no change*" based on the cut-offs of the SAQLI, PSQI and FOSQ, respectively. The ICCs of LSS in all versions were at least in the moderate range except in 5L and +R for FOSQ which were weak (Table 6). Similarly, for all items, k values were within at least fair to moderate limits and statistically significant ($p<0.05$), except for T item in PSQI and PD item in FOSQ which were insignificantly weak.

## DISCUSSION

Our qualitative findings lent further support to the content validity of the four bolt-on items. We showed that sleep, cognition, energy and relationships bolt-ons were relevant dimensions, and we adapted them and demonstrated that they improved the psychometric properties of the EQ-5D-5L in our study cohort of chronic disease patients with sleep disturbances with respect to the convergent validity, discriminatory power and explanatory power of EQ VAS. Our results were aligned with previous publications on bolt-ons in specific disease population groups. Some of the better-known examples are the bolt-ons on breathing, vision, cognition and itch.

Several studies that had examined the sleep dimension had shown varying results in enhancing the psychometric properties of EQ-5D[23, 28-31, 52-57]. The most recently published paper by Rencz et al showed that among nine bolt-on items, sleep bolt-on reduced the ceiling effects of EQ-5D by 30% [33]. In our study, it was reduced by 42.7%. The varying results could arise from many reasons. In the first place, Rencz et al, used a general public sample in their

analyses, while we examined diseased population affected by sleep disorders. Besides, there may be differences in the research methodology, phrasings and labels of the bolt-ons, as well as clinical elements. For example, unless they have bed partners who observe and report them, some patients are simply not aware of the sleep disrupting symptoms such as breathing (e.g. snoring), teeth grinding, dream enactment etc, which are occurring during their sleep. Rather these patients experience awake consequences in physical, mental and social domains, especially the latter two because of their fragmented sleep.

In comparing the individual bolt-ons, it is not surprising that "*sleep*" bolt-on outperformed the other three bolt-ons in our study given all participants recruited suffered from sleep disturbances that led them to seek medical attention. The measurement performance of "*tiredness*" and "*cognition*" bolt-ons were marginally behind "*sleep*" bolt-on. "*Sleepiness*" is often perceived and expressed mutually and loosely with "*tiredness*"/ "*fatigue*"/ "*low energy*" by patients in the phase-1 study or literature although they are clinically and pathologically distinguishable[35]. Patients who feel sleepy in the day due to their sleep disturbances generally report feeling tired although the converse may not be true concerning these two symptoms. Both symptoms can certainly affect one's mental functioning therefore impaired cognition is also often concurrently reported. Among the four bolt-ons, performance of "*relationship*" was weakest, and subpar in all measurement properties compared to the other three. Although social relationships or connectedness emerged as a relevant dimension in our qualitative interview, it is nonetheless an adverse by-product effect of sleep disturbances collaterally linked to their physical and mental sequelae. Moreover, there could exist multitude of factors, besides the severity and the kind of sleep problem, that possibly influence the impact on "*relationships*" in our study population. These included unmeasurable factors such as patients' personality types/social skills, values, motivation levels, as well as social norms. For example, impacts might be attenuated if patients were introvert, placed less value on social interactions and were less motivated in the need to want to socialize, or if they already had very weak existing social support and sense of social belonging. Conversely, it might also be possible that impacts were substantially mitigated to below perceivable thresholds if patients had strong and/or dense existing social networks. Unfortunately, our study was not designed to further determine and explore these factors. Except for discriminatory power where the bolt-on module was inferior to the "*sleep*" bolt-on, the convergent validity of the bolt-on module appeared to be higher than each of the four individual bolt-ons. Nevertheless, our study findings provided empirical evidence to support incorporating key multi-faceted single-item dimensions to EQ-5D-5L

which will allow the clinicians to better capture the health states of patients with sleep disturbance without unduly lengthening the questionnaire and exacerbating respondents' fatigue, while retaining the genericity advantage conferred by the instrument.

Our mixed study design enabled us to explore further the saliency of "*sleep*", "*cognition*", "*energy*" and "*relationships*" dimensions through the lenses of both patients and clinicians, and to verify the face and content validity of the four bolt-on items. Based on the qualitative input, we also revised the phrasings of the bolt-ons extracted from the literature to fit for purpose within the local cultural context. In testing their performances, we employed a combination of subjective data from self-reporting condition-specific PROMs and objective measures using polysomnographic parameters. The team had designedly included chronic disease as an enrolment criterion in our study so that its findings can provide broader applicability extending beyond just the population with sleep disturbances, given that sleep and energy/fatigue were identified as relevant generic constructs absent in EQ-5D across different population groups and had been built-into major condition specific PROM such as EORTC for cancer[23,28, 52, 58]. However, there were some limitations in our study that should be kept in mind when interpreting our results or generalising our findings. We were only able to compare the performances of the six EQ-5D versions using level sum scores and not the utility scores, as the latter will require a value set through eliciting the general public's health preferences of EQ-5D-5L containing the bolt-on(s) which is currently not available. Nevertheless, it can be argued that LSS may be more suitable than the preference-based index score when EQ-5D is used as a PROM in the clinical setting. The differences in the effect sizes between the bolt-on versions and standard EQ-5D-5L might have been underestimated due to the inclusion of patients with other chronic diseases that might have elevated the sensitivity of standard EQ-5D-5L. Also given that our study on the four bolt-ons was conducted in a specialist setting among Asian patients, the results may not be extended to other types of bolt-ons in dissimilar population groups. Further work can consider duplicating the study with larger sample sizes, investigating in other target populations and contextures to verify the internal and external validity of the results, and deriving social preference valuations set for health states of the bolt-on(s) containing EQ-5D-5L versions.

## CONCLUSIONS

The addition of sleep and related bolt-ons abolished the ceiling effects of EQ-5D-5L and enhanced its construct validity and responsiveness while preserving its reliability among chronic disease patients with sleep disturbances attending a multidisciplinary sleep clinic. A

module of bolt-ons including “*sleep*”, “*cognition*”, “*tiredness*” and “*relationship*” dimensions provided better results than individual bolt-ons.

**1**

**Table 1: Characteristics of study sample (N=110).**

| | |
|---|---|
| **Age (mean±SD)** | 46.7 (13.0) |
| **Male** | 84 (76.4) |
| **Ethnicity** | |
| Chinese | 78 (70.9) |
| Malay | 24 (21.8) |
| Indian | 4 (4.6) |
| Others | 3 (2.7) |
| **Marital Status** | |
| Single | 37 (33.6) |
| Married | 64 (58.2) |
| Divorced/Widowed | 9 (8.2) |
| **Employment** | |
| Employed | 90 (81.8) |
| Housewife/unemployed/retiree | 16 (14.6) |
| Student | 2 (3.6) |
| **Residential** | |
| Public (HDB) | 93 (84.6) |
| Private (condo/landed) | 17 (15.5) |
| **Education level** | |
| None | 1 (0.9) |
| Secondary & below | 20 (18.2) |
| 'A' level & equivalent | 47 (42.7) |
| Tertiary | 42 (38.2) |
| **Charlson comorbidity index** | |
| ≥ 1 | 65 (59.1) |
| 0 | 45 (40.9) |
| **Total sleep time, minutes** | |
| < 350 | 55 (50.0) |
| ≥ 350 | 55 (50.0) |
| **Sleep efficiency, percentage** | |
| < 80 | 58 (52.7) |
| ≥ 80 | 52 (47.3) |
| **Sleep latency, minutes** | |
| ≥ 20 | 50 (45.5) |
| < 20 | 60 (54.5) |
| **Arousal index, events per hour** | |
| ≥ 40 | 49 (44.6) |
| < 40 | 61 (55.5) |
| **Apnea-hypopnea index, events per hour** | |
| ≥ 30 | 80 (72.7) |
| < 30 | 30 (27.3) |
| **Oxygen desaturation index, events per hour** | |

| | |
|---|---|
| ≥ 30 | 57 (51.8) |
| < 30 | 53 (48.2) |
| **%TST SpO2<90, percentage** | |
| ≥ 25 | 36 (32.7) |
| < 25 | 74 (67.3) |

AHI; apnea-hypopnea index; AI; arousal index; CCI, Charlson comorbidity index; ODI, oxygen desaturation index; N, sample size of recruited patients in step 2 of study; SE, sleep efficiency; SL, sleep latency; TST, total sleep time; %, proportions in percentages; %TST SpO2<90; percentage TST with pulse oximetric oxygen saturation <90%

**Table 2: Ceilings and health state profiles of EQ-5D-5L and bolt-ons (N=110).**

| | Ceilings | | | | Health state profiles | |
|---|---|---|---|---|---|---|
| **Item** | **n (%)** | **EQ-5D-5L versions** | **n (%)** | ***Absolute /^relative ceiling reduction (%)** | **n/#maximum (proportions, %)** | ****20 top profiles (%)** |
| **MO** | 99 (99.0) | **5L** | 47 (42.7) | | 25/3125 (0.8) | 95.5 |
| **SC** | 107 (97.3) | | | | | |
| **UA** | 94 (85.5) | | | | | |
| **PD** | 63 (57.3) | | | | | |
| **AD** | 74 (67.3) | | | | | |
| **S** | 22 (20.2) | **+S** | 0 | 42.7/100.0 | 41/15625 (0.3) | 80.9 |
| **C** | 53 (48.2) | **+C** | 0 | 42.7/100.0 | 35/15625 (0.2) | 86.4 |
| **T** | 17 (15.5) | **+T** | 0 | 42.7/100.0 | 39/15625 (0.3) | 82.7 |
| **R** | 86 (78.2) | **+R** | 0 | 42.7/100.0 | 33/15625 (0.2) | 88.2 |
| **SCTR** | 7 (6.4) | **+SCTR** | 0 | 42.7/100.0 | 81/15625 (0) | 44.5 |

5L, EuroQoL-five dimensions-five levels; %, percentage; +C, 5L plus *cognition* bolt-on item; +T, 5L plus *tiredness* bolt-on item; +R, 5L plus *relationship bolt-on* item; +S, 5L plus *sleep bolt-on* item; +SCTR, 5L plus *sleep, cognition, tiredness and relationship bolt-on* items; AD, *anxiety/depression* item; C, *cognition* bolt-on item; MO, *mobility* item; N, total sample size; n, count; PD, *pain/discomfort* item; R, *relationship* bolt-on item; S, *sleep* bolt-on item; SC, *self-care* item; T, *tiredness* bolt-on item; UA, *usual activities* item.

* $ceiling_{5L} - ceiling_{5L+bolton}$

^ $(ceiling_{5L} - ceiling_{5L+bolton}) / ceiling_{5L}$

\# maximum possible health state for the EQ-5D-5L version

**proportion of counts for 20 top health profiles out of 110 total sample count

**Table 3: [*]Correlation coefficients between EQ-5D-5L items, bolt-ons and EQ VAS.**

| | MO | SC | UA | PD | AD | S | C | T | R | VAS |
|---|---|---|---|---|---|---|---|---|---|---|
| **MO** | 1 | 0.32 | 0.21 | 0.44 | 0.20 | 0.10 | 0.11 | 0.24 | 0.12 | -0.21 |
| **SC** | | 1 | 0.42 | 0.22 | 0.11 | 0.03 | 0.07 | 0.06 | 0.07 | -0.12 |
| **UA** | | | 1 | 0.30 | 0.15 | 0.26 | 0.16 | 0.19 | 0.13 | -0.17 |
| **PD** | | | | 1 | 0.32 | 0.16 | 0.10 | 0.20 | 0.01 | -0.28 |
| **AD** | | | | | 1 | 0.25 | 0.33 | 0.30 | 0.40 | -0.36 |
| **S** | | | | | | 1 | 0.43 | 0.62 | 0.10 | -0.38 |
| **C** | | | | | | | 1 | 0.46 | 0.22 | -0.15 |
| **T** | | | | | | | | 1 | 0.14 | -0.36 |
| **R** | | | | | | | | | 1 | -0.29 |

AD, *anxiety/depression* item; C, *cognition* bolt-on item; EQ-5D-5L, EuroQoL-five dimensions-five levels; EQ VAS, EuroQoL visual analogue scale; SC, *self-care* item; MO, *mobility* item; PD, *pain/discomfort* item; R, *relationship bolt-on* item; S, *sleep bolt-on* item; T, *tiredness bolt-on* item; UA, *usual activities* item.

[*]All $p<0.05$.

| **Table 4: [*]Construct validity of EQ-5D-5L and its five bolt-on(s) versions.** | | | | | | |
|---|---|---|---|---|---|---|
| | **5L** | **+S** | **+C** | **+T** | **+R** | **+SCTR** |
| **[**]Correlations** | | | | | | |
| *FOSQ* | -0.30 | -0.45 | -0.44 | -0.45 | -0.36 | -0.57 |
| *PSQI* | 0.29 | 0.51 | 0.42 | 0.41 | 0.30 | 0.54 |
| *SAQLI* | -0.24 | -0.44 | -0.34 | -0.38 | -0.28 | -0.52 |
| **[§]Known-groups** | | | | | | |
| *TST* | 0.523 | 0.601 | 0.553 | 0.543 | 0.510 | 0.586 |
| *SE* | 0.509 | 0.567 | 0.530 | 0.520 | 0.528 | 0.562 |
| *SL* | 0.549 | 0.578 | 0.518 | 0.569 | 0.541 | 0.561 |
| *AI* | 0.516 | 0.556 | 0.522 | 0.553 | 0.510 | 0.521 |
| *AHI* | 0.517 | 0.518 | 0.553 | 0.524 | 0.535 | 0.547 |
| *ODI* | 0.497 | 0.546 | 0.481 | 0.536 | 0.497 | 0.540 |
| *%TSTSpO2<90* | 0.516 | 0.542 | 0.555 | 0.553 | 0.483 | 0.524 |
| *[§§]Mean (SD)* | 0.518 (0.016) | [^]0.558 (0.027) | 0.530 (0.027) | [^]0.543 (0.017) | 0.515 (0.021) | ^0.549 (0.023) |

5L, EuroQoL-five dimensions-five levels; +C, 5L plus *cognition* bolt-on item; +T, 5L plus *tiredness* bolt-on item; +R, 5L plus *relationship bolt-on* item; +S, 5L plus *sleep bolt-on* item; +SCTR, 5L plus *sleep, cognition, tiredness and relationship bolt-on* items; %TST SpO2<90, percentage TST with pulse oximetric oxygen saturation <90%; AD, *anxiety/depression* item; AHI; apnea-hypopnea index; AI; arousal index; FOSQ, Functional Outcome Sleep Questionnaire; SC, *self-care* item; MO, *mobility* item; ODI, oxygen desaturation index; PD, *pain/discomfort* item; PSQI, Pittsburgh Sleep Quality Index; SAQLI, Sleep Apnea Quality of Life Index; SD, standard deviation; SE, sleep efficiency; SL, sleep latency; TST, total sleep time; UA, *usual activities* item.

[*] Presented data were on level sum scores.

[**]Spearman's rho correlations analyses, all p-values<0.001.

[§]C-statistics analyses of *a prior* determined known-groups.

[§§]Mean (standard deviation) of C-statistics data for the seven clinical known-groups.

[^]P-values of comparisons against 5L <0.05.

| **Table 5: *Responsiveness of EQ-5D-5L and its five bolt-on(s) versions.** | | | | | | |
|---|---|---|---|---|---|---|
| | **5L** | **+S** | **+C** | **+T** | **+R** | **+SCTR** |
| **Standardized response means** | | | | | | |
| *Treated (n=58)* | -0.59 (-0.87 to -0.31) | -1.34 (-1.70 to -0.98) | -0.90 (-1.21 to -0.59) | -1.36 (-1.72 to -1.00) | -0.69 (-0.98 to -0.40) | -1.83 (-2.25 to -1.40) |
| *SAQLI (n=44)* | -0.50 (-0.82 to -0.18) | -1.13 (-1.51 to -0.74) | -0.85 (-1.20 to -0.50) | -1.16 (-1.54 to -0.77) | -0.59 (-0.92 to -0.27) | -1.56 (-2.00 to -1.11) |
| *PSQI (n=44)* | -0.65 (-0.98 to -0.32) | -1.30 (-1.70 to -0.89) | -0.93 (-1.29 to -0.57) | -1.41 (-1.83 to -0.98) | -0.85 (-1.20 to -0.50) | -1.86 (-2.36 to -1.37) |
| *FOSQ (n=56)* | -0.61 (-0.89 to -0.32) | -1.34 (-1.71 to -0.78) | -0.99 (-1.31 to -0.67) | -1.33 (-1.69 to -0.96) | -0.74 (-1.04 to -0.44) | -1.86 (-2.30 to -1.42) |
| **C-statistics** | | | | | | |
| *Treated* | 0.576 | 0.776 | 0.662 | 0.739 | 0.598 | 0.846 |
| *SAQLI* | 0.504 | 0.647 | 0.625 | 0.606 | 0.537 | 0.704 |
| *PSQI* | 0.578 | 0.720 | 0.631 | 0.698 | 0.648 | 0.773 |
| *FOSQ* | 0.585 | 0.753 | 0.721 | 0.690 | 0.632 | 0.836 |

(), 95% CI; 5L, EuroQoL-five dimensions-five levels; +C, 5L plus *cognition* bolt-on item; +T, 5L plus *tiredness* bolt-on item; +R, 5L plus *relationship bolt-on* item; +S, 5L plus *sleep bolt-on* item; +SCTR, 5L plus *sleep, cognition, tiredness and relationship bolt-on* items; FOSQ, Functional Outcome Sleep Questionnaire; n, number of participants in the specified assessment category; PSQI, Pittsburgh Sleep Quality Index; SAQLI; Sleep Apnea Quality of Life Index.

* Presented data on level sum scores.

| **Table 6: [*]Reliability of EQ-5D-5L and its five bolt-on(s) versions.** | | | | |
|---|---|---|---|---|
| | **Untreated** (n=32) | **SAQLI** (n=38) | **PSQI** (n=37) | **FOSQ** (n=29) |
| **Intraclass correlations** | | | | |
| *EQ-5D-5L* | 0.81 (0.65-0.90) | 0.77 (0.53-0.88) | 0.51 (0.23-0.72) | 0.25 (-0.09-0.55) |
| *+S* | 0.78 (0.58-0.89) | 0.66 (0.14-0.86) | 0.54 (0.06-0.78) | 0.57 (0.24-0.78) |
| *+C* | 0.80 (0.59-0.91) | 0.78 (0.43-0.90) | 0.51 (0.15-0.73) | 0.49 (0.17-0.72) |
| *+T* | 0.76 (0.53-0.88) | 0.67 (0.21-0.85) | 0.41 (0.01-0.68) | 0.47 (0.12-0.71) |
| *+R* | 0.76 (0.56-0.88) | 0.75 (0.49-0.87) | 0.53 (0.26-0.73) | 0.37 (0.03-0.63) |
| *+SCTR* | 0.75 (0.49-0.88) | 0.62 (0.05-0.84) | 0.70 (0.35-0.86) | 0.70 (0.35-0.86) |
| **Cohen's kappa** | | | | |
| Mobility[^] | 0.78 | 0.77 | 0.68 | [##] |
| Self-care[^] | ^^ | ^^ | ^^ | ^^ |
| Usual activities[^] | 0.64 | 0.47 | [**] | [**] |
| Pain/discomfort[^] | 0.47 | 0.67 | 0.40 | 0.19 |
| Anxiety/depression[^] | 0.78 | 0.54 | 0.59 | 0.76 |
| Sleep[^] | 0.47 | 0.34 | 0.36 | 0.42 |
| Cognition[^] | 0.59 | 0.53 | 0.39 | 0.77 |
| Tiredness[^] | 0.44 | 0.30 | 0.16 | 0.39 |
| Relationship[^] | 0.43 | 0.50 | 0.39 | 0.54 |

(), 95% CI; 5L, EuroQoL-five dimensions-five levels; +C, 5L plus *cognition* bolt-on item; +T, 5L plus *tiredness* bolt-on item; +R, 5L plus *relationship bolt-on* item; +S, 5L plus *sleep bolt-on* item; +SCTR, 5L plus *sleep, cognition, tiredness and relationship bolt-on* items; FOSQ, Functional Outcome Sleep Questionnaire; n, number of untreated or "no change" participants in the specified assessment category; PSQI, Pittsburgh Sleep Quality Index; Rx, "*treated*" category; SAQLI; Sleep Apnea Quality of Life Index.

[*]Intraclass correlations and cohen's kappa analyses.

[^] Cohen's kappa analyses; all $p<0.05$.

[##]All participants rated 1 at baseline and follow-up except for two participants who rated 2 at baseline.

[**]All participants rated 1 at baseline and follow-up except for one participant who rated 2 at baseline.

[^^]All participants rated 1 at baseline and follow-up.

## REFERENCES


1. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Applied health economics and health policy. 2017;15(2):127-37.
2. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D Data: Cham (CH): Springer; 2020.
3. Wang A, Rand K, Yang Z, Brooks R, Busschbach J. The remarkably frequent use of EQ-5D in non-economic research. Eur J Health Econ. 2022;23(6):1007-14.
4. Motamedi KK, McClary AC, Amedee RG. Obstructive sleep apnea: a growing problem. Ochsner J. 2009;9(3):149-53.
5. Ferrie JE, Kumari M, Salo P, Singh-Manoux A, Kivimäki M. Sleep epidemiology--a rapidly growing field. Int J Epidemiol. 2011;40(6):1431-7.
6. Wickwire EM, Albrecht JS, Towe MM, Abariga SA, Diaz-Abad M, Shipper AG, et al. The Impact of Treatments for OSA on Monetized Health Economic Outcomes: A Systematic Review. Chest. 2019;155(5):947-61.
7. Abad VC, Sarinas PS, Guilleminault C. Sleep and rheumatologic disorders. Sleep Med Rev. 2008;12(3):211-28.
8. Cappuccio FP, Miller MA. Sleep and Cardio-Metabolic Disease. Curr Cardiol Rep. 2017;19(11):110.
9. Chen D, Yin Z, Fang B. Measurements and status of sleep quality in patients with cancers. Support Care Cancer. 2018;26(2):405-14.
10. Freeman D, Sheaves B, Waite F, Harvey AG, Harrison PJ. Sleep disturbance and psychiatric disorders. Lancet Psychiatry. 2020;7(7):628-37.
11. Khot SP, Morgenstern LB. Sleep and Stroke. Stroke. 2019;50(6):1612-7.
12. Lin CH, Lurie RC, Lyons OD. Sleep Apnea and Chronic Kidney Disease: A State-of-the-Art Review. Chest. 2020;157(3):673-85.
13. Malhotra RK. Neurodegenerative Disorders and Sleep. Sleep Med Clin. 2018;13(1):63-70.
14. McNicholas WT, Hansson D, Schiza S, Grote L. Sleep in chronic respiratory disease: COPD and hypoventilation disorders. Eur Respir Rev. 2019;28(153).
15. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res. 2021;30(3):647-73.
16. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. The European Journal of Health Economics. 2018;19(4):557-70.
17. Abma IL, van der Wees PJ, Veer V, Westert GP, Rovers M. Measurement properties of patient-reported outcome measures (PROMs) in adults with obstructive sleep apnea (OSA): A systematic review. Sleep Med Rev. 2016;28:18-31.
18. Schmidlin M, Fritsch K, Matthews F, Thurnheer R, Senn O, Bloch KE. Utility indices in patients with the obstructive sleep apnea syndrome. Respiration. 2010;79(3):200-8.
19. Jenkinson C, Stradling J, Petersen S. Comparison of three measures of quality of life outcome in the evaluation of continuous positive airways pressure therapy for sleep apnoea. J Sleep Res. 1997;6(3):199-204.
20. Jenkinson C, Stradling J, Petersen S. How should we evaluate health status? A comparison of three methods in patients presenting with obstructive sleep apnoea. Qual Life Res. 1998;7(2):95-100.
21. Weaver TE, Mathias SD, Crosby RD, Bron M, Bujanover S, Menno D, et al. Relationship between sleep efficacy endpoints and measures of functional status and health-related quality of life in participants with narcolepsy or obstructive sleep apnea treated for excessive daytime sleepiness. J Sleep Res. 2021;30(3):e13210.
22. Geraerds A, Bonsel GJ, Janssen MF, Finch AP, Polinder S, Haagsma JA. Methods Used to Identify, Test, and Assess Impact on Preferences of Bolt-Ons: A Systematic Review. Value Health. 2021;24(6):901-16.
23. Finch AP, Brazier JE, Mukuria C. Selecting Bolt-On Dimensions for the EQ-5D: Examining Their Contribution to Health-Related Quality of Life. Value Health. 2019;22(1):50-61.

24. Hoogendoorn M, Jowett S, Dickens AP, Jordan R, Enocson A, Adab P, et al. Performance of the EQ-5D-5L Plus Respiratory Bolt-On in the Birmingham Chronic Obstructive Pulmonary Disease Cohort Study. Value Health. 2021;24(11):1667-75.
25. Luo N, Wang X, Ang M, Finkelstein EA, Aung T, Wong TY, et al. A Vision "Bolt-On" Item Could Increase the Discriminatory Power of the EQ-5D Index Score. Value Health. 2015;18(8):1037-42.
26. Spronk I, Bonsel GJ, Polinder S, van Baar ME, Janssen MF, Haagsma JA. The added value of extending the EQ-5D-5L with an itching item for the assessment of health-related quality of life of burn patients: an explorative study. Burns. 2021;47(4):873-9.
27. Swinburn P, Lloyd A, Boye KS, Edson-Heredia E, Bowman L, Janssen B. Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: lessons learned from a feasibility study in the UK. Value Health. 2013;16(8):1156-62.
28. Finch AP, Brazier J, Mukuria C. Selecting Bolt-on Dimensions for the EQ-5D: Testing the Impact of Hearing, Sleep, Cognition, Energy, and Relationships on Preferences Using Pairwise Choices. Med Decis Making. 2021;41(1):89-99.
29. Perneger TV, Courvoisier DS. Exploration of health dimensions to be included in multi-attribute health-utility assessment. Int J Qual Health Care. 2011;23(1):52-9.
30. Yang Y, Brazier J, Tsuchiya A. Effect of adding a sleep dimension to the EQ-5D descriptive system: a "bolt-on" experiment. Med Decis Making. 2014;34(1):42-53.
31. Kim SH, Jo MW, Ock M, Lee SI. Exploratory Study of Dimensions of Health-related Quality of Life in the General Population of South Korea. J Prev Med Public Health. 2017;50(6):361-8.
32. Mulhern BJ, Sampson C, Haywood P, Addo R, Page K, Mott D, et al. Criteria for developing, assessing and selecting candidate EQ-5D bolt-ons. Quality of Life Research. 2022;31(10):3041-8.
33. Rencz F, Janssen MF. Testing the Psychometric Properties of 9 Bolt-Ons for the EQ-5D-5L in a General Population Sample. Value Health. 2024;27(7):943-54.
34. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-57.
35. Chua AP, Soh ZY, Rahman SA, Luo N, Shorey S. Lived experience of patients with sleep apnea: a systematic synthesis of qualitative evidence. Qual Life Res. 2023;32(5):1447-67.
36. What is the WHO definition of health? : World Health Organization; 2021 [cited 2024 June 16]. Available from: https://www.who.int/about/who-we-are/frequently-asked-questions.
37. Flemons WW, Reimer MA. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. Am J Respir Crit Care Med. 1998;158(2):494-503.
38. Weaver TE, Laizner AM, Evans LK, Maislin G, Chugh DK, Lyon K, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep. 1997;20(10):835-43.
39. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213.
40. de VetHCW, Terwee CB, Mokkink LB, et al. Field-testing: itemreduction and data structure. In: Terwee CB, Knol DL, de Vet HCW, et al., eds. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press;2011:65–95.
41. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
42. Juniper, E.F., Guyatt, G.H. and Jaeschke, R. (1996) How to develop and validate a new health-related quality of life instrument. In: Spilker, B., Ed., Quality of Life and Pharmacoeconomics in Clinical Trials, 2nd Edition, Lippincott-Raven Publishers, Philadelphia, 49-56.
43. Bikov A, Frent S, Deleanu O, Meszaros M, Birza MR, Popa AM, et al. Time Spent with Saturation below 80% versus 90% in Patients with Obstructive Sleep Apnoea. Journal of Clinical Medicine [Internet]. 2023; 12(13).

44. Temirbekov D, Güneş S, Yazıcı ZM, Sayın İ. The Ignored Parameter in the Diagnosis of Obstructive Sleep Apnea Syndrome: The Oxygen Desaturation Index. Turk Arch Otorhinolaryngol. 2018;56(1):1-6.
45. Flemons WW, Reimer MA. Measurement properties of the calgary sleep apnea quality of life index. Am J Respir Crit Care Med. 2002;165(2):159-64.
46. Sampaio RS, Pereira MG, Winck JC. Adaptation of the sleep apnea quality of life index (SAQLI) to Portuguese obstructive sleep apnea syndrome patients. Pulmonology. 2012;18(4):166-74.
47. Billings ME, Rosen CL, Auckley D, Benca R, Foldvary-Schaefer N, Iber C, et al. Psychometric performance and responsiveness of the functional outcomes of sleep questionnaire and sleep apnea quality of life instrument in a randomized trial: the HomePAP study. Sleep. 2014;37(12):2017-24.
48. Shergis JL, Ni X, Jackson ML, Zhang AL, Guo X, Li Y, et al. A systematic review of acupuncture for sleep quality in people with insomnia. Complement Ther Med. 2016;26:11-20.
49. Kim SH, Jeong JH, Lim JH, Kim BK. Acupuncture using pattern-identification for the treatment of insomnia disorder: a systematic review and meta-analysis of randomized controlled trials. Integr Med Res. 2019;8(3):216-26.
50. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994;6(4):284-90.
51. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
52. Finch AP, Brazier JE, Mukuria C, Bjorner JB. An Exploratory Study on Using Principal-Component Analysis and Confirmatory FactorAnalysis to Identify Bolt-On Dimensions: The EQ-5D Case Study. Value in Health. 2017;20(10):1362-75.
53. Finch AP. An investigation of methods for identifying and selecting bolt-on dimensions: the EQ-5D-5L case study: University of Sheffield.; 2017.
54. Kharroubi SA, Beyh YS, Brazier J, Yang Y. Modelling a preference-based index for EQ-5D-3L and EQ-5D-3L + Sleep using a Bayesian framework. Qual Life Res. 2020;29(6):1495-507.
55. Chen G, Olsen JA. Filling the psycho-social gap in the EQ-5D: the empirical support for four bolt-on dimensions. Qual Life Res. 2020;29(11):3119-29.
56. Efthymiadou O, Mossman J, Kanavos P. Health related quality of life aspects not captured by EQ-5D-5L: Results from an international survey of patients. Health Policy. 2019;123(2):159-65.
57. Jelsma J, Maart S. Should additional domains be added to the EQ-5D health-related quality of life instrument for community-based studies? An analytical descriptive study. Population Health Metrics. 2015;13.
58. Sprangers MAG, Bonnetain F. EORTC QLQ-C30. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer Netherlands; 2014. p. 1933-5.