



Societal Utilities for Cognitive Impairment in Schizophrenia: Developing a Preference-Based Scoring Algorithm Based on the Schizophrenia Cognition Rating Scale

Irene Santi · Andrew J. Lloyd · Claudia E. Hastedt ·
Matthijs M. Versteegh

Received: February 17, 2023 / Accepted: May 15, 2023 / Published online: July 13, 2023
© The Author(s) 2023

ABSTRACT

Introduction: Loss of cognitive function is a common feature in schizophrenia. However, generic measures of health-related quality of life favored by decision-makers, such as the EQ-5D, are not designed to detect changes in cognitive function. We report the valuation of the Schizophrenia Cognition Rating Scale (SCoRS), a schizophrenia-specific measure of cognitive impairment.

Methods: Expert opinion and psychometric analysis of the SCoRS from clinical trial data was undertaken to select 5 key items from the measure. These items were combined orthogonally to develop health-state vignettes. Vignettes were valued using composite time trade-off (cTTO) in one-on-one video calls. Several econometric models were fitted to the data to

estimate disutilities. Performance of EQ-5D- and SCoRS-based utilities were compared in the trial data.

Results: The SCoRS items selected for the valuation study represented attention, learning, processing speed, social cognition and memory. Four hundred respondents participated in the valuation study. The best observed health state was valued at 0.855 [standard deviation (SD) = 0.179] and the worst at 0.152 (SD = 0.575). At the most severe levels, ‘social cognition’ received the largest disutility followed by ‘learning’ and ‘memory’. The final model to estimate utilities had 15 parameters. SCoRS-based utilities were sensitive to change in cognition, but the EQ-5D was not.

Conclusion: It is feasible to value different dimensions of cognition separately using a validated instrument for proxy assessment. The resulting utilities indicate loss of quality of life due to reduced cognitive functioning.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12325-023-02553-7>.

Keywords: Schizophrenia; Utility; EQ-5D; Cognitive impairment; Cost-effectiveness

I. Santi (✉) · M. M. Versteegh
Institute for Medical Technology Assessment,
Erasmus University, Burgemeester Oudlaan 50, 3062
PA Rotterdam, The Netherlands
e-mail: santi@eshpm.eur.nl

A. J. Lloyd
Acaster Lloyd Consulting Ltd, London, UK

C. E. Hastedt
Boehringer Ingelheim International GmbH,
Ingelheim am Rhein, Germany

Key Summary Points

Common preference-based measures such as EQ-5D do not reflect cognition as a dimension of health.

Using data from a phase 2 trial comparing the glycine transporter-1 inhibitor BI 425809 compared to placebo, this study aimed to provide an alternative approach to value improvement in cognitive functioning to inform cost-effectiveness analysis of treatments for schizophrenia.

Using a schizophrenia-specific measure of cognitive functioning, the SCoRS, we derived utilities that demonstrated to be more sensitive to change in cognition than the EQ-5D.

It is feasible to describe different dimensions of cognition and value them separately using a validated, disease-specific instrument.

INTRODUCTION

The measurement and valuation of health is central to economic evaluations that involve the use of quality-adjusted life years (QALYs). Economic evaluations that estimate QALYs for improvements in cognition face the challenge that commonly used instruments, such as the EQ-5D, were not designed to measure and value cognitive function, and evidence suggests that the EQ-5D is poor at measuring differences in cognitive function [1]. Researchers may therefore resort to the Health Utilities Index Mark-3 (HUI-3) [2], but, unlike EQ-5D, in this instrument cognition is only characterized as the ability to remember things. In order to be able to use EQ-5D in a context where cognition is relevant, several initiatives have been deployed to value cognitive function separately, such as the development of a ‘bolt-on’ to EQ-5D for cognition [1, 3]. A ‘bolt-on’ is an additional

item added to the EQ-5D to address the measurement limitation. The bolt-on approach is a potentially useful solution, but it needs to be incorporated into the clinical trial design. Also, a single bolt-on dimension may not account for multidimensionality of cognition (e.g., memory, attention, and visuo-spatial skills). However, including all main elements of cognition as different bolt-ons would make the valuation task very complex for participants. In this study, we use an alternative approach to bolt-ons to place a value on improvements in cognitive impairment associated with schizophrenia (CIAS), which is based on an existing, disease-specific instrument and can be used alongside generic measures of health (such as EQ-5D). A valuation of an existing instrument has the benefit that utilities can be derived from studies that have already been conducted.

The multidimensionality of cognition is reflected in the Food and Drug Administration-approved MATRICS Consensus Cognitive Battery (MCCB), a selection of cognitive tests which evaluate cognitive capacity in seven dimensions: speed of processing, attention/vigilance, working memory, verbal learning, visional learning, reasoning and problem solving, and social cognition [4, 5]. A related instrument that assesses day-to-day cognitive functioning is the Schizophrenia Cognition Rating Scale (SCoRS) [6, 7], a 20-item interviewer-based rating scale developed to evaluate the impact of cognitive impairment on the patient’s daily functioning. The SCoRS is based on an interviewer assessment and refers to the patient’s actual cognitive functioning problems in daily life [6]. The SCoRS items correspond to the MATRICS cognitive domains, including memory, learning, attention, working memory, problem solving, processing speed, social cognition, and language. It has several advantages, including a brief administration time, association to real-world functioning, good test–retest reliability, and correlations with other measures of cognition and functioning, both performance-based and rater-assessed [14, 15]. The full 20-item version of the SCoRS provides a valuable source of information for health-state descriptions, but it has too many items for valuation purposes with the composite time trade-

off (cTTO) valuation, and it would yield incomprehensible health states.

Here, we show the results of a study designed to capture a time trade-off valuation of several dimensions of cognition as described by the SCoRS to inform cost-effectiveness analysis of therapies targeting cognitive function. We first describe a selection process to identify which items of the SCoRS should be included in a valuation study. We report on the results of the valuation study and apply several regression models to the data of the valuation study to estimate item level disutility. Subsequently, we demonstrate the performance of the resulting scoring algorithm on trial data relative to EQ-5D.

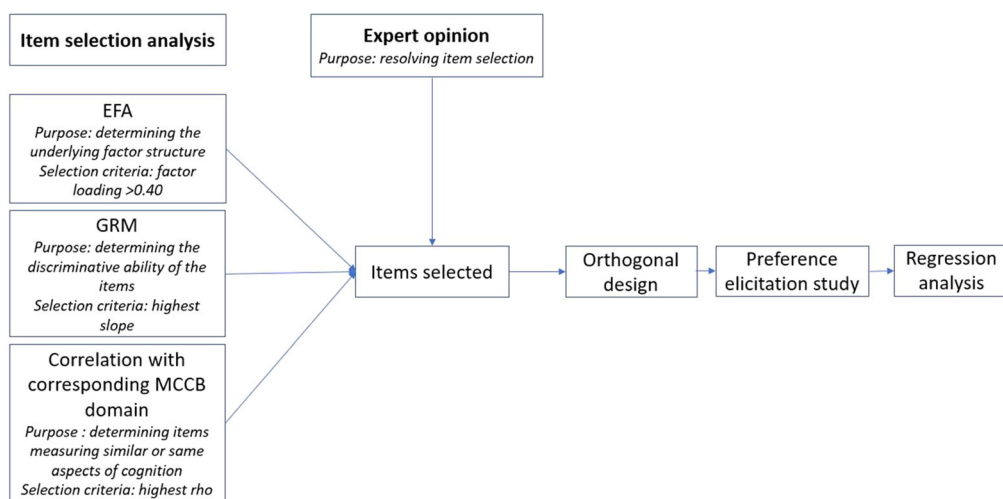
METHODS

The disease-specific approach was constructed in 3 stages: (1) the derivation of the classification system from the SCoRS; (2) the cTTO survey used to elicit values for a selection of SCoRS-based health states; and (3) the modeling of preference weights that can be used to generate utility values for the health states derived from the SCoRS. (Fig. 1).

Data from a multicenter phase 2 randomized clinical trial investigating the effect of the glycine transporter-1 inhibitor BI 425809 on

cognition in 509 patients with schizophrenia underwent secondary analysis. The sample is described in detail elsewhere [8]. In summary, the majority of patients were male (65%), mean age was 37.1 [standard deviation (SD)] = 7.7] and mean time since first diagnosis was 12.1 years (SD 7.8; range 0.2–36.0). Patients were psychiatrically stable and on stable treatment with standard of care antipsychotic medication. Baseline MCCB overall composite T-score was 31.5 (SD 13.2) and SCoRS interviewer-rated total score was 36.4 (8.9).

The objective of the psychometric analysis was to reduce the number of items in the SCoRS to roughly 5 or 6 in order to support the valuation task. These 5 or 6 items should reflect the range of the SCoRS measure, by including the best functioning items (from a psychometric perspective). All analyses were conducted on the full analysis set from the clinical trial using the Statistical Analysis System (SAS) v.9.4. The item content of the SCoRS covers a broad range of cognitive domains: attention, language, learning, memory/working memory, problem solving, processing speed, social cognition. Each of the 20 items is rated on a scale ranging from 1 to 4, with higher scores reflecting a greater degree of impairment (1 = none, 2 = mild, 3 = moderate, 4 = severe difficulties), resulting in a total sum score between 20 and 80.



EFA: exploratory factor analysis; GRM: graded response model; MCCB: MATRICS Consensus Cognitive Battery

Fig. 1 Item selection and preference study process

An exploratory factor analysis (EFA) was conducted on the baseline SCoRS ($n = 495$) to identify those items that most clearly represent the content domain of the underlying construct. As SCoRS items are not normally distributed, we used the unweighted least squares extraction method. As the aspects of cognitive function are likely to be correlated, we used oblique rotation as well as orthogonal. We applied the 0.40 criterion level in judging factor loadings as meaningful [9].

A graded response model (GRM), an item response theory (IRT) [10] model for ordinal data, was fitted to the 20 items to explore how many are needed for a precise measurement of cognitive deficits. The GRM model yields two item types of parameter estimates: the item thresholds and the item slope [11]. Item threshold parameters, β_{ik} , where i is the item and k is the response level, locate item response categories along the scale (i.e., the construct of interest). The item slope parameter, α , refers to the discriminative ability of the items, with higher slope values indicating a stronger relationship to the construct of interest. The SCoRS items have four response categories, therefore three item thresholds, k , were estimated. Measure precision (reliability) is conceptualized within IRT as “information”, which is inversely related to an individual’s standard error of measurement; the more information a measure provides in a certain trait range, the more precise an individual’s measurement in that trait range will be. Information was inspected by each “item information curve” (IIC), which indicated how informative, i.e., discriminating, each item was as a function of the latent trait, i.e., cognitive deficit.

Non-parametric correlation coefficients (Spearman’s Rho) were calculated between individual SCoRS items and the MCCB domains and composite scores. We expected a strong ($|r| \geq 0.60$) or at least moderate ($|r|$ values, 0.40–0.59) correlation between each MCCB domain score and the SCoRS items that are designed to measure similar or the same aspects of cognition.

The final selection of the most relevant items was based on both the results of these analyses

and expert input from the original developer of the SCoRS.

Design and Conduct of the Valuation Study

We used an orthogonal design to define health-state vignettes, which satisfied the criterion that all severity levels and all severity level combinations are equally prevalent and therefore balanced. [12] The design of five four-level factors specified 16 health states (Appendix 1 in supplementary material). The health states were randomly divided into two ‘blocks’ so that participants only had to value 8 health states. The health-state vignettes were assessed using the time trade-off (TTO) method.

Time Trade-off Method

The time trade-off (TTO) is a standardized interview method for valuing health states and was used to generate health utility weights for each health state. [13, 14] The method is designed to determine the point at which participants consider 10 years in the target health state to be equivalent to the prospect of X years in full health. Time in full health is varied until this point of indifference is reached. To minimize possible bias, we used the ping-pong method, which varies the amount of time in full health alternating between high and low values, changing by 6-month intervals. If a participant indicated that they believed that zero years in full health was preferable to any time living in a health state, then this implies that the participant considers the state to be worse than dead. At this point, the interviewer was asked to switch to a lead-time TTO exercise, which asks participants whether they would prefer to live for 10 years in full health followed by 10 years in a health state, or to live for 20 years in full health. This is referred to as the composite TTO (cTTO) [15].

Sample Size

A power analysis was conducted using simulation methods in R. Valuation data were simulated in 100 datasets with sample size n for the

16 health states of the orthogonal design specification. Sample size n could take on values between 100 and 500. Health-state values were simulated with disutilities for each worsening level. These disutilities were randomly drawn from a truncated normal distribution, with a mean of 0.05 and a standard deviation of 0.2, and multiplied by the level of severity. Subsequently, in each of the 100 datasets given a certain sample size, a regression equation was estimated, and the number of significant coefficients were stored. The specification was a 16-parameter random effects model with an intercept and dummy variables for the five dimensions included in the preference-based measure. The simulation results showed that, at a sample size of 400, a model with significant parameters could be estimated for nearly all of the 100 simulations, with increases in sample sizes not substantially improving results.

Recruitment

Members of the UK general population (aged > 18) were recruited by 12 field-based interviewers. The experienced TTO interviewers were all located in different regions of the UK. Recruitment was based on convenience sampling, including snowballing based on informal networks. The aim was to recruit a broadly representative sample of the UK general population in terms of age and sex using recruitment quotas. A total of 400 TTO interviews were conducted.

Compliance with Ethics Guidelines

The interview guide and survey instrument were approved by the Western Institutional Review Board. The study was qualified for an exemption because the research only includes interactions involving survey and interview procedures, and there were adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data. Eligible respondents provided informed consent to participate and to the scientific publication of the research while adhering to the General Data Protection Regulations (GDPR) and European Pharmaceutical Market Research Association regulations of data protection and privacy. Each

individual's consent to participate in the study was collected after an initial survey that verified eligibility and prior to the main interview. The survey terminated if individuals indicated that they were not willing to participate. Individuals who provided consent moved to the interview questions and their answers were saved in the survey database. Only individuals who completed the full interview received compensation for participation.

COVID-19 Adaptation of Design

In response to the UK COVID-19 lockdown at the start of the data collection (2021), all interviews were conducted under interviewer supervision using online video (Skype or Zoom), except for 5 face-to-face interviews. Participants were provided with information about the study and asked to complete a consent form along with a brief background questionnaire.

Interviewers requested participants to have health states either printed or presented on a separate screen. The online video interview was presented on a second screen so that the interviewer and participant were visible to one another at all times, enabling the visual analogue scale (VAS) or TTO board to be shown alongside the interviewer on the screen. The online interview encouraged eye contact and was designed to mimic a standard face-to-face interview as far as possible. Interviews all followed a standard script.

In the first part of the exercise participants rated each health state (plus a 'dead' state) on VAS, ranging from 0 (worst possible state) to 100 (full health). This was designed as a warmup task to familiarize participants with each description and the concept of rating states of health. Following the VAS exercise, participants completed two practice cTTO tasks. After that, they completed the cTTO exercise for eight health states.

Modeling Health State Preferences

To account for multiple observations per individuals, we modeled the cTTO data with random effects (RE) generalized linear regression, and random effects Tobit regression (accounting for left-censored at -1 data) [16] using

maximum likelihood estimation. Coefficients represent the utility decrement of moving from no impairment to mild impairment (level two), from no impairment to moderate impairment (level three), and from no impairment to severe impairment (level four).

The most appropriate model was selected based on the following criteria: the significance of the coefficients; logical consistency; goodness of fit; and accuracy. The model was considered to be logically consistent where worse health states had estimated values lower than better health states. The measurement of accuracy was compared using root mean square error (RMSE) and mean absolute error (MAE) with lower error preferred. Out of sample prediction accuracy was tested by holding out one health state and predicting its value within the 95% confidence interval through the regression analysis.

The final model was applied to the trial data to assess the responsiveness to change of the predicted utilities. We predicted that the mean changes in SCoRS utilities between baseline and end of treatment would reflect changes in MCCB cognitive score, while the EQ-5D would be insensitive to these changes. We used the effect size (ES: average score change/SD of initial score) [17, 18] and the standardized response mean (SRM: average score change/SD of score change) [19], in which the SCoRS utilities and the EQ-5D were compared with the MCCB. To examine the correlation of changes among the instruments, pairwise Spearman's correlation coefficients of the ES and the SRM were applied. All analyses were performed using R 3.6.1 (sample size calculation) STATA v.16 (estimation of regression equations) and SAS (trial data analysis).

RESULTS

Item Selection

The trial data included very few participants with severe (level 4) levels of cognitive deficits on the SCoRS. The exploratory factor analysis on the baseline dataset indicated that a single factor was the best structure, which was

consistent with a previous psychometric study on the SCoRS [7].

Table S1 (in supplementary material) provides the GRM estimates of the item discrimination (α_i) and threshold parameters (β_{i1} – β_{i3}), as well as estimates of their standard errors. Item discrimination parameter estimates (α) range from 0.90 to 1.73. The first threshold parameter represents the point on the latent trait continuum—i.e., cognitive impairment—where the probability of responding above category 1 is 50%, the second threshold parameter indexes the cognitive impairment level needed to reach the third response category from the second, and so on for the third threshold parameter. The results show that the threshold parameter estimates increased from the first threshold through the third, with, especially, a larger increase from the second to the third, indicating that only the most extremely impaired patients are likely to receive the fourth rating.

The inspection of each item IIC confirmed that items with larger slope parameters provide relatively more information. For example, within the memory domain, item 5 has slope 1.24 compared to item 4 with slope 0.90. Because information is a function of the square of the item slope, item 5 has nearly two times the information in the middle of the trait range (Figure S1—in supplementary material). Most of the SCoRS items were significantly correlated to the conceptually related MCCB domains, although correlations were poor (< -0.2) (Table S1). This has also been observed in other studies, and highlights the complex relationship between cognitive capacity (assessed by MCCB) and day-to-day cognitive functioning (assessed by SCoRS) with multiple mediating and moderating factors which may affect the ability of cognitive improvements to translate into better functioning [7, 20].

Item selection was based on item performance on EFA, IRT, and correlation with MCCB (Table 1). Expert advice was used to choose the items with the strongest correlation to cognition and the corresponding MCCB domain, and this criterion was then prioritized in cases where the other parameters did not indicate the same direction. Items were selected when, within a dimension, an item scored highest on EFA, IRT,

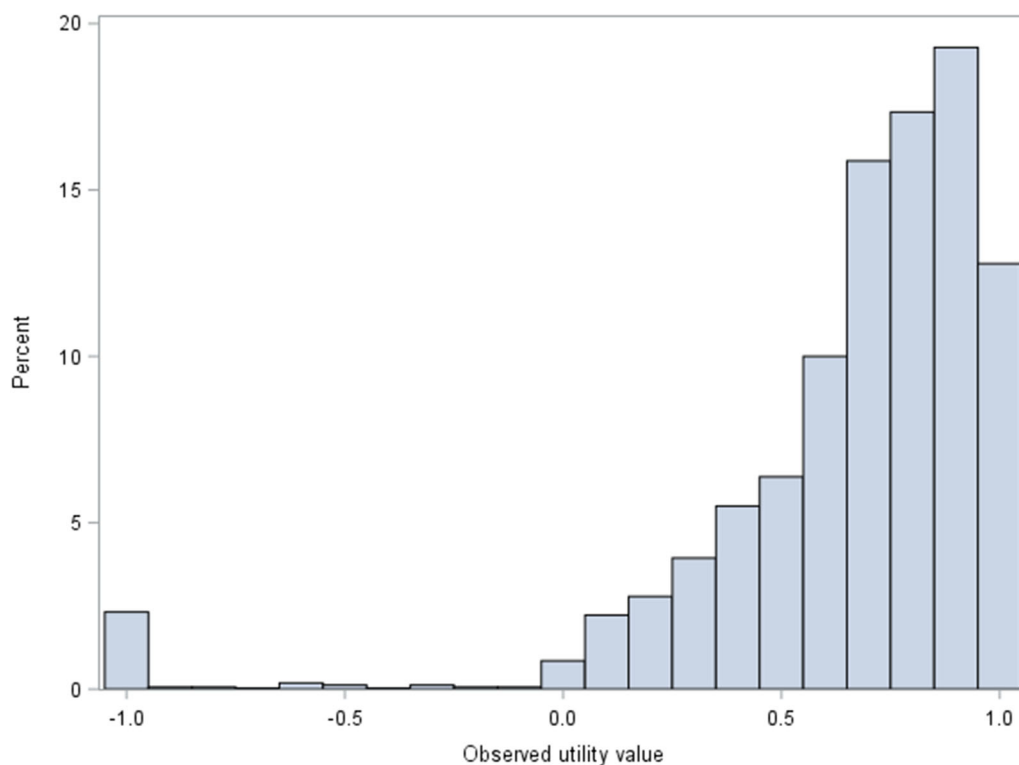


Fig. 2 Distribution of observed utility values (cTTO)

and MCCB correlation. This resulted in the selection of items 7, 13, and 14. In the dimensions' processing speed and language/social cognition, items with the highest MCCB correlation were not items with the best EFA or IRT score. Following expert opinion, preference was given to select items with the strongest correlation to cognition and the respective MCCB domain. The chosen items 15 and 18 had the highest correlations and the second-best slopes and factor loadings in their domain. The selected items were 7, 13, 14, 15, and 18, representing the memory, attention, learning, processing speed, and social cognition domain, respectively. In the problem-solving dimension, item scores on the three decision criteria were not consistent and too low. Given the a priori maximum of 5 items in the valuation study, no items were selected from this dimension.

Valuation Study

Sample characteristics were similar to the UK population in terms of gender and ethnicity. However, younger as well as higher-educated categories were slightly over-represented in our sample (Table S3 in supplementary material).

The final cTTO set includes 3,200 cTTO responses among 3,200 cTTO values, with 71 (2.2%) and 3,097 (96.7%) cTTO values considered "worse than death" and "better than death", respectively (Fig. 2). The mean cTTO value ranged from 0.152 for the health state "4444" to 0.855 for the state "4111" (Table S4 in supplementary material).

Estimation of Disutilities

Both the RE and the Tobit models reported insignificant results for level 2 (mild) and 3 (moderate difficulties) of attention, and for level 2 (mild difficulties) of processing speed and memory (Table 2, models 1 and 2). When

Table 1 Selection criteria for SCoRS items: EFA/IRT criteria satisfied = 1; correlation of SCoRS items with corresponding MCCB domains—highest correlations in *bold*

| Item n° | SCoRS domains/items | | | | | | | | | | |
|--|---------------------|--------|-----------------|------------------|-----------------------|--------|---------------------------|--------|--------|----------------|--------|
| | Attention | | Learning | | Memory/working memory | | | | | | |
| | 3 | 11 | 13 ^a | 6 | 14 ^a | 1 | 2 | 4 | 5 | 7 ^a | 8 |
| EFA – highest load | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| IRT – highest slope (by domain) | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Correlation with corresponding MCCB domain | – 0.04 | – 0.05 | – 0.12 | – 0.13 | – 0.17 | – 0.09 | – 0.11 | – 0.12 | – 0.09 | – 0.14 | – 0.09 |
| Item n° | SCoRS domains/items | | | | | | | | | | |
| | Problem solving | | | Processing speed | | | Language/social cognition | | | | |
| | 9 | 12 | 17 | 15 ^a | 16 | 10 | 18 ^a | 19 | 20 | 1 | 1 |
| EFA – highest load | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| IRT – highest slope (by domain) | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Correlation with corresponding MCCB domain | – 0.05 | 0.03 | – 0.03 | – 0.18 | – 0.14 | – 0.14 | – 0.14 | – 0.19 | – 0.11 | – 0.14 | – 0.14 |

EFA exploratory factor analysis, IRT item response theory, MCCB MATRICS Consensus Cognitive Battery

^aSelected

Table 2 Modeling results; statistically significant coefficient in *bold* (p value < 0.05)

| | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|-------|----------------|----------------|----------------|----------------|----------------|
| | | RE MLE | Tobit | RE MLE | Tobit | RE MLE |
| SCoRS dimension/item | Level | β | β | β | β | β |
| Attention | 2 | – 0.020 | – 0.020 | – 0.019 | – 0.019 | – 0.019 |
| 13. Staying focused | 3 | – 0.018 | – 0.018 | | | |
| | 4 | – 0.104 | – 0.106 | – 0.104 | – 0.106 | – 0.105 |
| Learning | 2 | – 0.042 | – 0.043 | – 0.041 | – 0.041 | – 0.043 |
| 14. Learning new things | 3 | – 0.064 | – 0.065 | – 0.063 | – 0.064 | – 0.065 |
| | 4 | – 0.181 | – 0.183 | – 0.180 | – 0.182 | – 0.187 |
| Processing speed | 2 | – 0.013 | – 0.012 | – 0.013 | – 0.013 | – 0.016 |
| 15. Speaking as fast as you would like | 3 | – 0.064 | – 0.065 | – 0.064 | – 0.064 | – 0.063 |
| | 4 | – 0.130 | – 0.132 | – 0.130 | – 0.132 | – 0.135 |
| Social cognition | 2 | – 0.047 | – 0.048 | – 0.046 | – 0.047 | – 0.047 |
| 18. Understanding what people mean when they are talking to you | 3 | – 0.097 | – 0.098 | – 0.097 | – 0.097 | – 0.097 |
| | 4 | – 0.236 | – 0.239 | – 0.236 | – 0.239 | – 0.245 |
| Memory | 2 | – 0.011 | – 0.011 | – 0.011 | – 0.011 | – 0.010 |
| 7. Remembering information and/or instructions recently given to you? | 3 | – 0.055 | – 0.055 | – 0.055 | – 0.055 | – 0.059 |
| | 4 | – 0.156 | – 0.158 | – 0.157 | – 0.158 | – 0.162 |
| Constant | | 0.959 | 0.961 | 0.959 | 0.960 | 0.966 |
| Number of participants | | 400 | 400 | 400 | 400 | 394 |
| Observations | | 3200 | 3200 | 3200 | 3200 | 3152 |
| RMSE | | 0.333 | 0.333 | 0.333 | 0.333 | |
| MAE | | 0.0523 | 0.0515 | 0.0516 | 0.0505 | |
| Illogical | | 0 | 0 | 0 | 0 | |
| Not significant | | 4 | 4 | 2 | 2 | |

Model 1 and model 2 include all levels of the covariates; model 3 and model 4 include level 2 and level 3 of attention combined; in model 5, answers that demonstrated a misunderstanding of the health states by rating a less severe health state (4111) worse than a more severe health state (4444) have been removed from the sample; participants' logical answers have been retained

SCoRS Schizophrenia Cognition Rating Scale, RE MLE random effect maximum likelihood estimation, RMSE root mean square error, MAE mean absolute error

combining level 2 and 3 of attention, the corresponding coefficient was statistically significant in both models (Table 2, models 3 and 4). The RMSE was equal for both the RE and the Tobit models. However, the MAE was lower for

the Tobit model, showing better accuracy than the RE model. The out of sample prediction analysis showed that, for 3 out of 16 health states, the predicted value was higher and for 2 it was lower than the observed 95% CI (Table S5

Table 3 Correlation of ES and SRM of MCCB, cTTO, and EQ-5D

| | MCCB overall composite score | |
|----------------------------|---------------------------------------|---------------------------------------|
| | ES, Spearman's rho (<i>p</i> -value) | SRM, Spearman's rho (<i>p</i> value) |
| EQ-5D (<i>n</i> = 443) | 0.03 (0.552) | 0.03 (0.535) |
| cTTO (<i>n</i> = 445) | − 0.22 (< .001) | − 0.21 (< .001) |

cTTO composite time trade-off, *ES* effect size, *SRM* standardized response mean, *MCCB* MATRICS Consensus Cognitive Battery

in supplementary material). A sensitivity analysis was conducted excluding six responders who valued state 41,111 equal or worse than state 4444. The results were consistent with the main analysis (Table 2, model 5).

Application of Cognition Utility Values to Trial Data

Table 3 shows the responsiveness to change measured as the correlation of the effect size and standardized response mean between the cTTO and MCCB overall composite score, and the EQ-5D and MCCB overall composite score. Correlations were small but statistically significant for the cTTO, while almost null and no statistically significant correlation was found for EQ-5D.

DISCUSSION

This study reports work to estimate utility weights for several subdimensions of cognition, based on items of the SCoRS measure. These weights allow us to estimate change in utility related to impairment of cognitive function as measured by the SCoRS. The EQ-5D is the standard method for estimating health utilities, but, as outlined above, there is clear evidence that it lacks validity when used to assess conditions in which cognitive function is a key element. We used cTTO to estimate the

relationship between length and quality of life for items of the SCoRS that reflect the cognition dimensions attention, learning, speed of processing, social cognition, and memory. At the most severe levels, items on ‘social cognition’ received the largest disutility followed by ‘learning’ and ‘memory’.

The aim of our study was to evaluate different aspect of cognitive function for which the EQ-5D was not designed, and therefore often does not reflect differences in cognition very well. We used the SCoRS because it was part of the outcomes of the clinical trial, with already demonstrated psychometric validity and correspondence with the MATRICS cognitive domains, and it is also an outcome measure accepted by regulatory bodies.

The SCoRS requires an average of about 15 min. [6] The MCCB is a long test battery that can usually be completed between 60 and 90 min of administration [4]. For health economic purposes, future studies for deriving utilities just need the SCoRS and not the MCCB. Where evidence suggests that the EQ-5D lacks content validity, some recent initiatives to overcome this issue have focused on the development of ‘bolt-ons’, whereby additional dimensions are added to the original 5 core dimensions as a single item. When EQ-5D is unable to measure a very important health issue (such as vision or hearing loss, dyspnoea, or itch) a common response would be to add those dimensions to the EQ-5D. There are, however, some challenges with the bolt-on approach in the present context.

First, bolt-on valuations are inherently limited to one or two items. Typically, the other five dimensions of EQ-5D have to be valued simultaneously with the bolt-on. Respondent information processing has to be taken into account, limiting the total number of items that can be included in a valuation study. Exploring the multi-faceted nature of cognition, and placing relative values on the different dimensions, would not be feasible if the EQ-5D dimensions have to be valued too. Also, as stated above, the use of a ‘bolt-on’ version of EQ-5D in a clinical trial necessitates that all of the development and validation work is completed on the new instrument prior to the start of the

trial. Clinical development programs often do not have scope or space in their timelines to support this.

This study presents a way to avoid these issues. First, to assess cognitive function, we have used a validated existing, physician-administered scale, widely used in clinical research in CIAS. The valuation work that we have undertaken has produced utilities specifically for impairment of cognitive functioning.

The current study has several limitations. First, the data underlying the selection of items for the valuation task were based on item performance in one clinical trial. While this trial was not small, the participating population of stable schizophrenia outpatients had a limited age range (18–50 years), and with mostly mild to moderate functional impairment. It is uncertain how this has affected the selection of items, but use of the utility values reported here should be limited to schizophrenia populations fairly similar to the one described in this study.

Second, the modeled data of the valuation study could not discriminate between levels 2 and 3 of attention. While it is not uncommon in the estimation of a utility tariff that some item levels fail to reach significance, its explanation is not straightforward. It is possible our sample was too small, but, regardless, we conclude that there is no loss of utility to avoid just mild difficulties with attention.

Third, deconstructing cognition in smaller dimensions and subsequently selecting one item to reflect these dimensions requires choices. Here we used expert opinion as criterion to prioritize for decision making when the other parameters indicated equal eligibility. It is possible that other experts could have preferred other criteria. Nonetheless, this ambiguity is difficult if not impossible to overcome in selecting items. The validation analyses suggested that out of sample predictive quality can be improved by including more health states.

Fourth, it has been described that the valuation of any health state can be sensitive to focusing effects [21], and arguably this applies even more when the dimension of health is presented in isolation, rather than presented alongside other dimensions of health, such as mobility or mood. This is a common issue in the

development of condition-specific preference-based measures, and similarly applies to the valuation of cognition as conducted in this study.

Last, and perhaps most importantly, severe limitations with cognition are uncommon in those who are able to participate as respondents in a valuation study. While, arguably, respondents may have experience with cognitive decline in others, witnessing the external effects of a loss in cognition is very different from experiencing it. Almost everyone will be familiar, for example, with intra-individual fluctuation in cognitive performance and some level of temporary impairment of memory and attention, but it is uncertain to what extent the general public may provide adequate insights into quality-of-life losses due to severe problems with cognition in patients with schizophrenia. This may be true in several instances, and is widely debated (e.g., Brazier et al. 2018, Versteegh et al., 2016) [22, 23]. The solutions proposed typically involve either having patients value dimensions of health themselves or informing respondents about the experience of a health state prior to a valuation task. With cognitive problems, however, neither might be feasible, and this current study does not provide a solution to those issues.

CONCLUSIONS

Cognition has been recognized as a dimension of health not reflected by common preference-based measures, such as EQ-5D. Here, we have shown that it is feasible to describe several dimensions of cognition and to value them separately using a validated, disease-specific instrument which is interview-based and rater-assessed. Our new method provides an alternative approach to value improvement in cognitive functioning that seems to be promising and more sensitive than EQ-5D. Exploring the results of this method in future studies is needed, especially in addition to established instruments like EQ-5D to further explore the relationship between quality of life and cognitive functioning.

ACKNOWLEDGEMENTS

We thank the participants of the study. We also thank Professor Richard S.E. Keefe, PhD (Duke University Medical Center) for his expert advice. We thank Saskia de Groot, PhD (iMTA) for earlier valuable work in this area.

Funding. This study and the journal's fee were funded by Boehringer Ingelheim International GmbH. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole and have given their approval for this version to be published.

Author Contributions. Irene Santi contributed to the conception and design of the analysis, analysis and interpretation of the data, and the drafting of the manuscript. Andrew Lloyd contributed to the conception and design of the analysis, analysis and interpretation of the data, and the drafting of the manuscript. Claudia Hastedt to the conception and design of the analysis, acquisition of the data, analysis and interpretation of the data, and the drafting of the manuscript. Matthijs Versteegh contributed to the conception and design of the analysis, analysis and interpretation of the data, and the drafting of the manuscript.

Prior Publication. This study was previously presented, in part, at the Virtual International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Europe 2021 (November 30–December 3, 2021).

Disclosures. Dr. Claudia Hastedt is an employee of Boehringer Ingelheim International GmbH. Dr. Andrew Lloyd is the Director and owner of Acaster Lloyd Consulting Lloyd

Consulting. At the time the study was conducted, Dr. Irene Santi and Dr. Matthijs Versteegh were employees of the institute of Medical Technology Assessment (iMTA), which, together with Dr. Andrew Lloyd, received funds from Boehringer Ingelheim International GmbH. Currently Dr. Irene Santi and Dr. Matthijs Versteegh are independent consultants.

Compliance with Ethics Guidelines. This study was approved by the Western Institutional Review Board (1–1,391,162-1). This study was performed in accordance with the Declaration of Helsinki of 1964 and its later amendments. Eligible respondents provided informed consent to participate. Each individual's consent to participate in the study was collected at the end of the informed consent and the responses were recorded in the survey database.

Data Availability. The datasets generated during and/or analyzed during the current study are not publicly available due to confidentiality agreement.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol.* 1999;52(4):293–301. [https://doi.org/10.1016/s0895-4356\(98\)00163-2](https://doi.org/10.1016/s0895-4356(98)00163-2).
2. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes.* 2003;1:54. <https://doi.org/10.1186/1477-7525-1-54>.
3. 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 Mak Int J Soc Med Decis Mak.* 2021;41(1):89–99. <https://doi.org/10.1177/0272989X20969686>.
4. Nuechterlein KH, et al. The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity. *Am J Psychiatry.* 2008;165(2):203–13. <https://doi.org/10.1176/appi.ajp.2007.07010042>.
5. Kern RS, et al. The MATRICS consensus cognitive battery, part 2: co-norming and standardization. *Am J Psychiatry.* 2008;165(2):214–20. <https://doi.org/10.1176/appi.ajp.2007.07010043>.
6. Keefe RSE, Poe M, Walker TM, Kang JW, Harvey PD. The schizophrenia cognition rating scale: an interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. *Am J Psychiatry.* 2006;163(3):426–32. <https://doi.org/10.1176/appi.ajp.163.3.426>.
7. Keefe RSE, et al. Reliability, validity and treatment sensitivity of the Schizophrenia cognition rating scale. *Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol.* 2015;25(2):176–84. <https://doi.org/10.1016/j.euroneuro.2014.06.009>.
8. Fleischhacker WW, et al. Efficacy and safety of the novel glycine transporter inhibitor BI 425809 once daily in patients with schizophrenia: a double-blind, randomised, placebo-controlled phase 2 study. *Lancet Psychiatry.* 2021;8(3):191–201. [https://doi.org/10.1016/S2215-0366\(20\)30513-7](https://doi.org/10.1016/S2215-0366(20)30513-7).
9. Ford JK, MacCALLUM RC, Tait M. The application of exploratory factor analysis in applied psychology: a critical review and analysis. *Pers Psychol.* 1986;39(2):291–314. <https://doi.org/10.1111/j.1744-6570.1986.tb00583.x>.
10. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika.* 1969;34(S1):1–97. <https://doi.org/10.1007/BF03372160>.
11. S. E. Embretson and S. P. Reise, *Item response theory for psychologists.* 2013. Accessed: Dec. 13, 2021. [Online]. Available: <https://www.taylorfrancis.com/books/e/9781410605269>
12. A. S. Hedayat and N. J. A. Sloane Stufken, John, “Orthogonal Arrays: Theory and Applications.”
13. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ.* 1986;5(1):1–30. [https://doi.org/10.1016/0167-6296\(86\)90020-2](https://doi.org/10.1016/0167-6296(86)90020-2).
14. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes.* 4th ed. Oxford: Oxford University Press; 2015.
15. B. M. F. Janssen, M. Oppe, M. M. Versteegh, and E. A. Stolk, “Introducing the composite time trade-off: a test of feasibility and face validity,” *Eur. J. Health Econ. HEPAC Health Econ. Prev. Care*, vol. 14 Suppl 1, pp. S5–13, Jul. 2013, <https://doi.org/10.1007/s10198-013-0503-2>.
16. Austin PC. “A comparison of methods for analyzing health-related quality-of-life measures”, *Value Health.* *J Int Soc Pharmacoeconomics Outcomes Res.* 2002;5(4):329–37. <https://doi.org/10.1046/j.1524-4733.2002.54128.x>.
17. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Statistics and strategies for evaluation. Control Clin Trials.* 1991;12(4 Suppl):142S-158S. [https://doi.org/10.1016/s0197-2456\(05\)80019-4](https://doi.org/10.1016/s0197-2456(05)80019-4).
18. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* 1989;27(3 Suppl):S178-189. <https://doi.org/10.1097/00005650-198903001-00015>.
19. C. B. Terwee, F. W. Dekker, W. M. Wiersinga, M. F. Prummel, and P. M. M. Bossuyt, “On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation,” *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.*, vol. 12, no. 4, pp. 349–362, Jun. 2003, doi: <https://doi.org/10.1023/a:1023499322593>.
20. S. Kharawala, C. Hastedt, J. Podhorna, H. Shukla, B. Kappelhoff, and P. D. Harvey, “The relationship between cognition and functioning in schizophrenia: A semi-systematic review,” *Schizophr. Res. Cogn.*, vol. 27, p. 100217, Mar. 2022, <https://doi.org/10.1016/j.scog.2021.100217>.
21. Dolan P. Thinking about it: thoughts about health and valuing QALYs. *Health Econ.* 2011;20(12):1407–16. <https://doi.org/10.1002/hec.1679>.

-
22. M. M. Versteegh, A. Leunis, C. A. Uyl-de Groot, and E. A. Stolk, "Condition-specific preference-based measures: benefit or burden?," *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.*, vol. 15, no. 3, pp. 504–513, May 2012, <https://doi.org/10.1016/j.jval.2011.12.003>.
 23. Brazier J, et al. "Identification, review, and use of health state utilities in cost-effectiveness models: an ISPOR good practices for outcomes research task force report", *Value Health. J Int Soc Pharmacoeconomics Outcomes Res.* 2019;22(3):267–75. <https://doi.org/10.1016/j.jval.2019.01.004>.