



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/143564/>

Version: Published Version

Article:

Franklin, M., Mukuria, C., Mulhern, B. et al. (2019) Measuring the burden of schizophrenia using clinician and patient-reported measures: an exploratory analysis of construct validity. *The Patient - Patient-Centered Outcomes Research*, 12 (4). pp. 405-417. ISSN: 1178-1653

<https://doi.org/10.1007/s40271-019-00358-x>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Measuring the Burden of Schizophrenia Using Clinician and Patient-Reported Measures: An Exploratory Analysis of Construct Validity

Matthew Franklin¹ · Clara Mukuria¹ · Brendan Mulhern^{1,2} · Irwin Tran³ · John Brazier¹ · Stuart Watson^{4,5}

© The Author(s) 2019

Abstract

Background Various self-reported or clinician-reported (as a proxy) measures exist to quantify the burden of schizophrenia on patients. Evidence of the psychometric relationship between these measures to inform their practical use is limited.

Objectives Our objective was to conduct an exploratory analysis of the construct validity of patient-reported (EQ-5D, SF-6D, WEMWBS, SQLS subscales of Psychosocial, Motivation, Symptoms) versus clinician-reported measures (PANSS, CGI-SCH, NSA-4, HoNOS-PbR) to inform future use of patient-reported measures for burden-of-illness assessment and/or economic evaluation.

Methods In an adult patient population with schizophrenia, construct validity was assessed in relation to convergent and known-group validity. Convergent validity was assessed using Spearman's rank absolute correlation strength (ACS: weak ≤ 0.3 , moderate = $0.3 < 0.5$, strong ≥ 0.5) and graphically using locally weighted scatterplot smoothing (LOWESS) techniques. Known-group validity was assessed using Cohen's *d* absolute effect size (AES: small ≤ 0.5 , moderate = $0.5 < 0.8$, large ≥ 0.8). Floor and ceiling effects were assessed as a proxy of sensitivity in this cross-sectional study. Statistical significance was assessed at the 5% threshold level ($p < 0.05$). Across head-to-head assessments, the frequency of producing the strongest ACS, largest AES, and statistically significant results determined the best overall construct validity.

Results Overall, 304 patients consented to the study. In relation to statistically significant results, the SF-6D most frequently exhibited the strongest ACS and largest AES against the clinician-reported measure scores (ACS range 0.084–0.436; AES range 0.043–0.746), and the SQLS Motivation subscale most frequently exhibited the weakest/smallest values (ACS range 0.009–0.157; AES range 0.002–0.397), although these results were mixed according to the clinician-reported measure used for comparative analysis (ACS range 0.009–0.529; AES range 0.002–0.934).

Conclusion The SF-6D indicated the best (mostly moderate) construct validity but still missed the negative symptoms of the condition. Although further evidence is required to confirm or refute these exploratory results, compared with the EQ-5D, the SF-6D can be self-reported to better capture generic health-related quality-of-life aspects of schizophrenia for the purpose of economic evaluation. The lack of construct validity for SQLS Motivation and Symptoms subscales were hypothesized post-hoc to be representative of the complementary information elicited by the subscales not captured by the clinician-reported measures. Therefore, the SQLS can be self-reported to capture complementary (i.e., additional) information relative to clinician-reported measures.

Key Points for Decision Makers

Compared with the EQ-5D, the SF-6D can be self-reported to better capture health-related quality-of-life aspects of schizophrenia for the purpose of economic evaluation.

The SQLS can be self-reported to capture complementary information relative to clinician-reported measures.

A more mental health-focused preference-based measure that can capture the negative symptoms of schizophrenia and that can be used for the purpose of economic analysis is still desirable for healthcare-related decision making.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40271-019-00358-x>) contains supplementary material, which is available to authorized users.

✉ Matthew Franklin
matt.franklin@sheffield.ac.uk

Extended author information available on the last page of the article

1 Background

Schizophrenia is a mental health disorder characterized by a range of different psychological impacts, including changes in thinking and behavior. The health-related quality of life (HRQoL) and social burden of schizophrenia is large, affecting both patients and their caregivers, for example, their social and financial situation [1, 2]. Many outcome measures have been designed to assess the burden of schizophrenia. These measures may have different conceptual perspectives that can be condition specific, such as the Schizophrenia Quality-of-Life Scale (SQLS) [3], designed to capture schizophrenia-specific QoL aspects, or generic, such as the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) [4, 5], designed to capture broader outcomes and be applicable to more than one mental health condition. The WEMWBS has been assessed in individuals with schizophrenia but only as part of a mixed-diagnosis group [6], so it is unclear whether it is an appropriate measure for this population. Measures may also be categorised according to whether or not they are preference based; examples of preference-based measures include the three-level EuroQoL Five-Dimension (EQ-5D) [7, 8] and the Short-Form Six-Dimension (SF-6D) [9]. These preference-based measures are used to form a profile score that is converted into a preference-based index score (usually based on societal preferences) and thus allow economic evaluation of interventions using cost-utility analysis (CUA) to inform the allocation of resources by healthcare-governing agencies such as the UK National Institute for Health and Care Excellence (NICE) [10]. In CUA, QoL measured on a preference-based scale anchored at 0 (dead) to 1 (full health) is combined with length of life to generate quality-adjusted life-years (QALYs), allowing comparisons between interventions that affect quantity of life and/or QoL. However, while there can be “some confidence” in the use of generic measures (e.g., the SF-6D and EQ-5D) in patients with mood and anxiety disorders because of their demonstrated psychometric validity and responsiveness [11], this is not the case with schizophrenia. In patients with schizophrenia, the data are less conclusive [11–13], and it has been argued that preference-based measures focused on the impact of the mental disorder (rather than generic measures covering both physical and mental health) should instead be considered [11].

The SQLS, WEMWBS, EQ-5D, and SF-6D are all self-reported measures. Evidence suggests that, in people with schizophrenia, self-report should be interpreted with caution as condition-related aspects such as cognitive impairment may impair reliability [14]. Measures completed by observers (e.g., clinicians), such as the Positive and Negative Symptom Scale (PANSS) [15] and Clinical Global Impression-Schizophrenia Change (CGI-SCH) [16] have been developed to aid assessment of the burden of schizophrenia.

Each measure differs in its features and constructs, and at times there is value in using more than one measure when the choice of measure(s) is based on the purpose of assessment [17, 18]. When using more than one measure, the relationship between the measures should be considered, for instance, including a clinician-reported measure alongside patient-reported measures [19].

The overall aim of this analysis was to perform an exploratory assessment of the construct validity of patient-reported measures in the context of patients with schizophrenia to assess their potential utility when reflecting the burden of schizophrenia alongside clinician-reported measures.

2 Methods

2.1 Data

Data were drawn from a multicenter, non-interventional, cross-sectional survey designed to assess the burden of illness of people with persistent symptoms of schizophrenia [20] (ClinicalTrials.gov identifier: NCT01634542). The dataset included adult patients in the UK who had persistent schizophrenia symptoms despite receiving adequately dosed antipsychotic treatment and who had not had an acute exacerbation in the last 3 months. Respondents and clinicians completed questionnaires at their usual clinic visits.

2.2 Outcome Measures

2.2.1 Patient-Reported Measures

Patients completed the SQLS, a 30-item self-reported measure of schizophrenia-related QoL on a five-point item scale from 0 (best state) to 4 (worst state), with three subscales: (1) Psychosocial (15 items), (2) Motivation (and energy; seven items); and (3) Symptoms (and side effects; eight items). Each subscale has a range from 0 (best possible state) to 100 (worst possible health state), whereby the scoring and rescaling for each subscale follows the methods described by Wilkinson et al. [3]: the scale score (SS) equals the total of raw scores of each item in the scale (RS_{tot}) divided by the maximum possible score of all the items in the scale (RS_{max}), all multiplied by 100; i.e., $SS = (RS_{tot}/RS_{max}) \times 100$. They also completed the WEMWBS, a 14-item scale of positive well-being and psychological functioning [21] with scores ranging from 14 (worst state) to 70 (best state). Two generic HRQoL preference-based measures were included: EQ-5D and SF-6D (the latter derived from the SF-12 version 2 [9]). The EQ-5D has five dimensions with three severity levels, describing 243 health states [7], and a preference-based score ranging from 1 (best state) to -0.594 (i.e., 0 is equivalent to a state of dead, and negative values are equivalent to

states valued as worse than dead) [8]. The SF-6D has six dimensions, with between four and six response levels, and preference-based score ranging from 1 (perfect health) to 0.345 (worst state) [9].

2.2.2 Clinician-Reported Measures

Clinician-reported measures included the PANSS; a 30-item, 7-point schizophrenia syndrome scale from 1 (absent) to 7 (extreme), which can be summed across or within three subscales: positive (seven items, score range 7–49); negative (seven items, score range 7–49); general psychopathology (16 items, score range 16–112); total score range 30–210 [15]. The CGI-SCH was used to assess the severity of schizophrenia (five items) in positive, negative, cognition, depression, and overall symptoms [16]. All items are rated on a 7-point scale, from 1 (normal/not ill) to 7 (among the most severely ill), with no overall score produced [16]. Negative symptoms associated with schizophrenia were assessed using the Negative Symptoms Assessment (NSA-4) [22] 7-point global rating scale (negative symptoms over the last 7 days). Finally, clinicians also completed the Health of the Nation Outcome Scales for Payment by Results (HoNOS-PbR) designed by UK psychiatrists for use in routine clinical practice as a record of a patient's progress across 12 items [23]: behavior (three items), function (two items), symptoms (three items), and social (four items); a total score can be derived via summation of the 12 items.

2.3 Exploratory Validity Analysis

This analysis is considered exploratory as no a priori assumptions were made about the relationship between the clinician- and patient-reported measures; rather, the relationships identified will inform a discussion about the level of association between these measures, their uses in measuring the burden of schizophrenia, and areas for future research in a patient population with schizophrenia. Therefore, to explore the relationship of the patient-reported HRQoL measures in patients with schizophrenia versus clinician-reported measures, floor and ceiling effects and construct validity were explored.

Floor and ceiling effects are important to assess as, within cross-sectional studies, they can be used as a proxy of sensitivity in relation to how well a measure can detect changes in HRQoL. For example, if a large proportion of the sample is at the floor (a score representing the highest level of symptoms or poor functioning) or ceiling (a score representing no problems), this impairs the ability (sensitivity) of the measure to pick up decreases or increases in HRQoL, respectively. The presence of floor and ceiling effects and data distribution (histograms are presented in Appendix S1 in the Electronic Supplementary Material [ESM]) were used

to select the appropriate statistical tests to assess construct validity; that is, if the data were non-normally distributed and affected by floor and ceiling effects, nonparametric tests were preferred to parametric tests.

Construct validity refers to the extent to which an instrument measures what it is intended to measure (e.g., HRQoL) compared with other indicators. To do this, a “gold standard” is required against which to assess the measures of interest. However, because no gold standard measure covering the full complex nature of “mental health” or schizophrenia exists, we used a range of clinician-reported measures as indicators. Two related types of construct validity were undertaken: convergent and known-group.

Convergent validity was used to assess the relationship strength between measures. This can be done using correlation analysis and locally weighted scatterplot smoothing (LOWESS) techniques. Here, correlation analysis indicates the degree to which the instruments measure related factors at the overall, dimension, or item level. Correlation associations were considered weak if scores were ≤ 0.3 , moderate if $0.3 < 0.5$, and strong if ≥ 0.5 [24]; statistical significance was defined at the 5% threshold level. Spearman's rank correlation coefficient was used as a nonparametric test based on the score distribution across the measures (see Sects. 3.1 and 3.2). LOWESS is a form of nonparametric regression that plots a line of central tendency between two variables on a scatterplot, thereby visualizing the relationship across the possible score ranges. LOWESS captures general patterns in the relationship between two measures without making assumptions about their actual relationship.

Known-group validity assesses the extent to which scores on an instrument differ across groups where they are expected to differ (e.g., clinical severity indicators). This can be measured by calculating effect sizes (calculated as the difference in mean scores between two adjacent severity subgroups divided by the standard deviation of scores for the milder of the two subgroups) between groups that provide a standard indicator of the size difference. Cohen's *d* was used to calculate standardized effect sizes and the *p* value calculated from the *F* statistic. Effect sizes ≤ 0.5 were considered small, $0.5 < 0.8$ moderate, and ≥ 0.8 large [24]; statistical significance was defined at the 5% threshold level. For this assessment, a focused literature search was performed to identify already established clinically meaningful severity cut-off points for the clinical measures. If cut-off points could not be identified, ad-hoc cut-offs were established based on the score format and distribution of the measure to inform this exploratory analysis; that is, the authors assessed the score format (e.g., if based on a continuous or categorical scale) then identified what proportion of patients were distributed across this score range, whereby an ad-hoc cut-off was based on establishing an equal proportion of people between two or more groups such that

enough variation existed in each group that a meaningful known-group effect size could be identified if one existed. Ad-hoc cut-offs were used for the PANSS, CGI-SCH, and the NSA-4, whereas established cut-offs from the literature were used for the HoNOS-PbR [25]. It should be noted that established cut-offs for the PANSS were identified from the literature, but were based on a percentage change in scores over time (i.e., requiring data to be collected from at least two time points) and therefore could not be used for this cross-sectional study [26]. These cut-offs are reported in the results, and the implications of using ad-hoc cut-offs are included in the discussion.

We assumed clinician-reported measures represent the “true state” of patients in the absence of any inherent “gold standard” for this assessment. Therefore, a moderate to strong/large correlation and known-group effect size between the clinician-reported and patient-reported measures suggests construct validity of the patient-reported measures. Convergent validity between different patient-reported measures may be used as an estimate of the coherence and consistency of patient report and thus the potential impact of invalid or random responses (e.g., due to schizophrenia-related cognitive impairment [14]) on the reliability of the overall construct validity assessment of the measures in this population.

Across the head-to-head assessments, the frequency of producing the strongest absolute correlation strength (ACS), largest absolute effect size (AES), and statistically significant results will be used to determine the best overall construct validity between patient-reported measures. Evidence to suggest the existence of floor-and-ceiling effects, or invalid or random responses (i.e. evidence of poor convergent validity between patient-reported measures), will be used to inform the suggested reliability of these construct validity results.

3 Results

Overall, 304 patients consented to the study; however, as the WEMWBS was included in the protocol later in the process, it was included for only 297 patients. A summary of patient and condition characteristics is provided in Table 1; measure scores and completion rates are presented in Tables 2 and 3. Based on the clinician-reported scores, these patients were defined as having a “mild” NSA-4 global negative rating score and “mildly” severe CGI-SCH symptoms.

3.1 Floor and Ceiling Effects

Floor and ceiling effect statistics are presented in Tables 4 and 5. In general, all of the clinical measures had evidence of ceiling effects; this was less apparent with the PANSS and

HoNOS-PbR total, but not the subscale scores. The EQ-5D had issues with ceiling effects; this was less apparent for other patient-reported measures.

3.2 Convergent Validity

All correlation coefficients are presented in Tables 6 and 7. In summary, amongst the patient-reported measures, the SF-6D most frequently (six times) exhibited the strongest correlations against the clinician-reported measure scores, with the SQLS Psychosocial, WEMWBS, and EQ-5D exhibiting the strongest strength correlation three times each. The SQLS Symptoms subscale did not strongly correlate with any of the clinician-reported measures.

Between patient-reported measures, all correlations were of moderate to strong strength and statistically significant (the exception being any correlations with the SQLS Motivation subscale, which were of weak strength and had one

Table 1 Summary of patient and condition characteristics ($N=304$)

Variable	Statistic
Age: all patients, n (range)	42 (20–70)
Age: groups	
20–29	39 (12.8)
30–39	83 (27.3)
40–49	96 (31.6)
50–59	69 (22.7)
60–70	17 (5.6)
Male sex	237 (78)
Ethnicity	
White	246 (80.9)
Mixed	6 (2.0)
Asian or Asian British	22 (7.3)
Black or Black British	30 (9.9)
Characteristics of patient’s condition for $N=166^a$	
Time since diagnosis of schizophrenia (years) ^b	12.1 (0–46)
Time in treatment since diagnosis (years) ^b	12.8 (0–46)
Duration of negative symptoms (years)	11.2 (0–46)
Duration of positive symptoms (years)	12.4 (0–46)

Data are presented as n (%) or mean (range) unless otherwise indicated

NICE National Institute for Health and Care Excellence

^aCharacteristics about the patient’s condition were only collected for a subgroup of the study cohort (i.e., 166 people; 55% of the study cohort)

^bThe mean time since diagnosis for this subgroup was 12.1 years, but the mean time in treatment since diagnosis was 12.8 years—based on the mean times, treatment for this patient group may have started before diagnosis. NICE’s recommendation to begin assessment “without delay” may have resulted in this patient group receiving treatment before a full diagnosis was confirmed [33], which possibly explains this difference in time between diagnosis and treatment

Table 2 Summary and descriptive statistics of patient reported-outcomes measure scores

Short name	No. completed, N (%) ^a	Mean	Median	SD	Minimum	Maximum
EQ-5D tariff score	304 (100)	0.69	0.74	0.27	- 0.18	1
SF-6D tariff score	302 (99.3)	0.67	0.66	0.12	0.35	1
SQLS Psychosocial	303 (99.7)	43.09	43.33	15.73	8	85
SQLS Motivation	304 (100)	55.73	53.57	10.58	25	96
SQLS Symptoms	303 (99.7)	34.02	34.38	15.08	3	78
WEMWBS	251 (84.5)	43.78	44	9.58	14	67

EQ-5D EuroQoL Five-Dimension, SD standard deviation, SF-6D Short-Form Six-Dimension, SQLS Schizophrenia Quality-of-Life Scale, WEMWBS Warwick-Edinburgh Mental Wellbeing Scale

^aPatient- and clinician-reported measures could have been completed for N=304, apart from for the WEMWBS measure, which was for N=297

Table 3 Summary and descriptive statistics of clinician-reported outcomes measure scores

Short name	No. completed, N (%) ^a	Mean	Median	SD	Minimum	Maximum
PANSS total	303 (99.7)	65.43	66	17.16	31	122
PANSS Positive	303 (99.7)	15.69	16	5.24	7	36
PANSS Negative	303 (99.7)	17.19	16	6.02	7	35
PANSS Psychopathology	303 (99.7)	32.54	33	8.86	16	66
CGI-SCH Positive	303 (99.7)	2.97	3	1.18	1	6
CGI-SCH Negative	303 (99.7)	2.70	3	1.12	1	6
CGI-SCH Depressive	303 (99.7)	2.40	2	1.12	1	6
CGI-SCH Cognitive	303 (99.7)	2.46	2	1.06	1	5
CGI-SCH Overall	303 (99.7)	3.15	3	1.03	1	6
NSA-4 rating	303 (99.7)	3.01	3	1.29	1	6
HoNOS-PbR total	274 (90.1)	9.38	9	4.72	0	28
HoNOS-PbR Behavior	297 (97.7)	0.91	1	1.14	0	5
HoNOS-PbR Functional	301 (99.0)	1.81	2	1.31	0	6
HoNOS-PbR Symptoms	298 (98.0)	3.52	3.5	2.08	0	10
HoNOS-PbR Social	284 (93.4)	3.01	3	2.21	0	12

CGI-SCH Clinical Global Impression-Schizophrenia Change, HoNOS-PbR Health of the Nation Outcome Scales for Payment by Results, NSA-4 Negative Symptoms Assessment, PANSS Positive and Negative Symptom Scale, SD standard deviation

^aPatient- and clinician-reported measures could have been completed for N=304, apart from for the WEMWBS measure which was for N=297

Table 4 Floor and ceiling effect assessment for patient-reported outcomes measures

Short name	No. completed, N (%) ^a	Floor score/worst state	Ceiling score/best state	N (%) floor/worst state	N (%) ceiling/best state
EQ-5D tariff score	304 (100)	- 0.594	1	0 (0)	49 (16.1)
SF-6D tariff score	302 (99.3)	0.345	1	0 (0)	4 (1.3)
SQLS Psychosocial	303 (99.7)	100	0	0 (0)	0 (0)
SQLS Motivation	304 (100)	100	0	0 (0)	0 (0)
SQLS Symptoms	303 (99.7)	100	0	0 (0)	0 (0)
WEMWBS	251 (84.5)	14	70	1 (0.4)	0 (0)

EQ-5D EuroQoL Five-Dimension, SF-6D Short-Form Six-Dimension, SQLS Schizophrenia Quality-of-Life Scale, WEMWBS Warwick-Edinburgh Mental Wellbeing Scale

^aPatient- and clinician-reported measures could have been completed for N=304, apart from for the WEMWBS measure which was for N=297

Table 5 Floor and ceiling effect assessment for clinician-reported outcome measure

Short name	No. completed, N (%)	Floor score/worst state	Ceiling score/best state	N (%) Floor/worst state	N (%) Ceiling/best state
PANSS total	303 (99.7)	210	30	0 (0)	0 (0)
PANSS Positive	303 (99.7)	49	7	0 (0)	14 (4.6)
PANSS Negative	303 (99.7)	49	7	0 (0)	8 (2.6)
PANSS Psychopathology	303 (99.7)	112	16	0 (0)	2 (0.7)
CGI-SCH Positive	303 (99.7)	7	1	0 (0)	37 (12.2)
CGI-SCH Negative	303 (99.7)	7	1	0 (0)	47 (15.5)
CGI-SCH Depressive	303 (99.7)	7	1	0 (0)	75 (24.8)
CGI-SCH Cognitive	303 (99.7)	7	1	0 (0)	61 (20.1)
CGI-SCH Overall	303 (99.7)	7	1	0 (0)	18 (5.8)
NSA-4 rating	303 (99.7)	7	1	0 (0)	30 (9.9)
HoNOS-PbR total	274 (90.1)	48	0	0 (0)	2 (0.7)
HoNOS-PbR Behavior	297 (97.7)	12	0	0 (0)	138 (46.5)
HoNOS-PbR Functional	301 (99.0)	8	0	0 (0)	55 (18.3)
HoNOS-PbR Symptoms	298 (98.0)	12	0	0 (0)	25 (8.4)
HoNOS-PbR Social	284 (93.4)	16	0	0 (0)	34 (12.0)

CGI-SCH Clinical Global Impression-Schizophrenia Change, HoNOS-PbR Health of the Nation Outcome Scales for Payment by Results, NSA-4 Negative Symptoms Assessment, PANSS Positive and Negative Symptom Scale, SD standard deviation

Table 6 Correlation coefficient matrix between patient and clinician-reported outcomes measures

Measure reporter	Measure	Patient-reported measures					
		SQLS Psycho-social	SQLS Motivation	SQLS Symptoms	WEMWBS	EQ-5D tariff	SF-6D tariff
Clinician-reported measures	PANSS total	0.214*	0.093	0.133*	-0.238*	-0.229*	-0.274*
	PANSS Positive	0.225*	0.128*	0.131*	-0.224*	-0.212*	-0.277*
	PANSS Negative	0.017	-0.037	-0.014	-0.158*	-0.034	-0.084
	PANSS Psychopathology	0.243*	0.104	0.174*	-0.220*	-0.264*	-0.285*
	CGI-SCH Positive	0.265*	0.072	0.210*	-0.270*	-0.297*	-0.327*
	CGI-SCH Negative	0.047	-0.047	0.015	-0.172*	-0.142*	-0.121*
	CGI-SCH Depressive	0.413*	0.122*	0.170*	-0.302*	-0.298*	-0.368*
	CGI-SCH Cognitive	0.020	-0.027	0.051	-0.033	-0.179*	-0.110
	CGI-SCH Overall	0.300*	0.047	0.231*	-0.297*	-0.312*	-0.355*
	NSA-4 rating	0.191*	0.083	0.095	-0.272*	-0.175*	-0.282*
	HoNOS-PbR total	0.426*	0.127*	0.323*	-0.374*	-0.432*	-0.421*
	HoNOS-PbR Behavior	<i>0.210*</i>	0.084	0.143*	-0.197*	-0.199*	-0.200*
	HoNOS-PbR Functional	0.092	-0.009	0.256*	-0.139*	-0.327*	-0.256*
	HoNOS-PbR Symptoms	0.529*	0.157*	0.316*	-0.396*	-0.414*	-0.436*
HoNOS-PbR Social	0.285*	0.053	0.195*	-0.303*	-0.245*	-0.255*	

Cohen's cut-offs: < 0.3 = weak; 0.3 < 0.5 = moderate; ≥ 0.5 = strong; bold formatting indicates moderate or strong, and italics indicates the strongest strength correlation across patient-reported or caregiver-reported measure

CGI-SCH Clinical Global Impression-Schizophrenia Change, EQ-5D EuroQoL Five-Dimension, HoNOS-PbR Health of the Nation Outcome Scales for Payment by Results, NSA-4 Negative Symptoms Assessment, PANSS Positive and Negative Symptom Scale, SF-6D Short-Form Six-Dimension, SQLS Schizophrenia Quality-of-Life Scale, WEMWBS Warwick-Edinburgh Mental Wellbeing Scale

* $p < 0.05$

Table 7 Correlation coefficient matrix between patient-reported outcome measures

Measure reporter	Measures	Patient-reported measures					
		SQLS Psychosocial	SQLS Motivation	SQLS Symptoms	WEMWBS	EQ-5D tariff	SF-6D tariff
Patient-reported measures	SQLS Psychosocial	1	0.268*	0.574*	− 0.631*	− 0.480*	− 0.670*
	SQLS Motivation	<i>0.268*</i>	1	0.082	− 0.148*	− 0.203*	− 0.252*
	SQLS Symptoms	0.574*	0.082	1	− 0.386*	0.458*	− 0.450*
	WEMWBS	− 0.631*	− 0.148*	− 0.386*	1	0.406*	0.542*
	EQ-5D tariff	− 0.480*	− 0.203*	0.458*	0.406*	1	0.637*
	SF-6D tariff	− 0.670*	− 0.252*	− 0.450*	0.542*	0.637*	1

Cohen's cutoffs: < 0.3=weak, 0.3 < 0.5=moderate, ≥ 0.5=strong; bold formatting indicates moderate or strong, and italics indicates the strongest strength correlation across patient-reported or caregiver-reported measure

EQ-5D EuroQoL Five-Dimension, SF-6D Short-Form Six-Dimension, SQLS Schizophrenia Quality-of-Life Scale, WEMWBS Warwick-Edinburgh Mental Wellbeing Scale

* $p < 0.05$

statistically non-significant correlation). This provides some evidence of coherent and consistent self-reporting of outcomes across these measures. The scatter plots and LOW-ESS lines provided further support of convergent validity between clinician-reported and patient-reported measures, the results and figures for which are described and presented in Appendix S2 in the ESM.

3.3 Known-Group Validity

All effect sizes are presented in Tables 8 and 9. Although the overall majority of effect sizes were small, the SQLS Psychosocial, WEMBS, EQ-5D, and SF-6D each indicated some medium to large effect sizes across the clinician-reported measures (but not with the PANSS); all medium to large effect sizes were statistically significant. The WEMWBS indicated a medium effect size between the NSA-4 “mild” and “moderate to severe” cut-offs and was the only patient-reported measure to indicate anything but a small effect size between the NSA-4 groups. Across clinician-reported measures, the WEMWBS, EQ-5D, and SF-6D more often indicated a larger effect size than the SQLS Psychosocial, Motivation, or Symptoms subscales.

4 Discussion

A total of 304 patients with persistent symptoms of schizophrenia were recruited to a UK-based cross-sectional survey. These exploratory results suggest that, when patient-reported measures (EQ-5D, SF-6D, WEMWBS, SQLS subscales of Psychosocial, Motivation, and Symptoms) were assessed against clinician-reported measures (PANSS, CGI-SCH, NSA-4, HoNOS-PbR), the patient-reported EQ-5D, SF-6D, WEMWBS, and SQLS Psychosocial subscale had

moderate construct validity in patients with schizophrenia. There was less support for the construct validity of the SQLS Symptoms subscale and nearly no support for the SQLS Motivation subscale. There was also evidence of consistent reporting of outcomes between the patient-reported measures, which suggests that those with schizophrenia in this patient sample could report their HRQoL consistently across measures.

4.1 Floor and Ceiling Effects

The EQ-5D had some issues with ceiling effects, which is a common finding across different conditions [27]; these ceiling effects were less apparent for the other patient-reported measures. The high ceiling effect suggests that, for a proportion of patients, the dimensions of EQ-5D are not sensitive to their schizophrenia-specific ill health. The five-level EQ-5D-5L was developed in an attempt to improve sensitivity to changes in health and address the ceiling effect issues associated with the EQ-5D [28].

4.2 Construct Validity

Our results suggest that only the condition-specific SQLS Psychosocial subscale in some cases indicated better construct validity than the generic measures, depending on the clinician-reported measure used for analysis; however, the SQLS Motivation and Symptom subscales always indicated weak construct validity, and the SF-6D more often indicated the better construct validity than any other patient-reported measure. For the EQ-5D and SF-6D, these results were similar to previous studies in this patient population [29]. The identified statistically significant correlations between patient-reported measures suggested reasonably consistent reporting of HRQoL across self-reported measures,

Table 8 Testing the known-group validity between the patient-reported and PANSS and CGI measures

	Groups (score range)	SQLS psychosocial		SQLS motivation		SQLS symptoms		WEMWBS		EQ-5D tariff		SF-6D tariff	
		Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)
PANSS total	Abs–Min (30–60)	40.7 ± 16.0	– 0.258 (0.029)*	55.0 ± 10.1	– 0.119 (0.311)	32.6 ± 15.2	– 0.153 (0.193)	45.7 ± 10.4	0.333 (0.011)*	0.727 ± 0.267	0.252 (0.033)*	0.694 ± 0.119	0.352 (0.003)*
	Mild–Ext (> 60)	44.7 ± 15.4		56.2 ± 10.9		34.9 ± 14.9		42.6 ± 8.9		0.659 ± 0.270		0.653 ± 0.113	
PANSS Pos.	Abs–Min (7–14)	40.9 ± 17.0	– 0.246 (0.035)*	54.4 ± 11.0	– 0.219 (0.060)	32.8 ± 15.7	– 0.143 (0.221)	45.6 ± 10.9	0.314 (0.015)*	0.729 ± 0.251	0.278 (0.017)*	0.698 ± 0.123	0.423 (< 0.001)*
	Mild–Ext (> 14)	44.7 ± 14.6		56.7 ± 10.2		34.9 ± 14.6		42.6 ± 8.4		0.655 ± 0.280		0.649 ± 0.108	
PANSS Neg.	Abs–Min (7–14)	43.9 ± 14.8	0.082 (0.496)	55.8 ± 9.3	0.015 (0.903)	35.1 ± 13.8	0.111 (0.356)	45.2 ± 9.8	0.233 (0.078)	0.691 ± 0.268	0.029 (0.812)	0.672 ± 0.116	0.043 (0.722)
	Mild–Ext (> 14)	42.6 ± 16.2		55.7 ± 11.2		33.4 ± 15.8		43 ± 9.4		0.683 ± 0.272		0.667 ± 0.118	
PANSS Psy.	Abs–Min (16–32)	40.2 ± 15.5	– 0.361 (0.002)*	54.8 ± 10.5	– 0.172 (0.135)	32.2 ± 15.0	– 0.238 (0.039)*	45.2 ± 10.0	0.283 (0.026)*	0.733 ± 0.264	0.344 (0.003)*	0.695 ± 0.117	0.427 (< 0.001)*
	Mild–Ext (> 32)	45.8 ± 15.6		56.6 ± 10.6		35.8 ± 15.0		42.5 ± 9.1		0.641 ± 0.270		0.646 ± 0.112	
CGI Pos.	Norm–Min (1–2)	38.5 ± 15.1	– 0.481 (< 0.001)*	54.0 ± 10.0	– 0.265 (0.026)*	31.1 ± 15.9	– 0.314 (0.009)*	46 ± 10.2	0.382 (0.003)*	0.774 ± 0.212	0.536 (< 0.001)*	0.713 ± 0.110	0.622 (< 0.001)*
	Mild–Sev (> 2)	45.8 ± 15.5		56.8 ± 10.8		35.8 ± 14.3		42.4 ± 8.9		0.634 ± 0.288		0.643 ± 0.113	
CGI Neg.	Norm–Min (1–2)	43.3 ± 15.8	0.024 (0.835)	55.7 ± 9.9	0.002 (0.985)	34.6 ± 15.6	0.076 (0.512)	45.2 ± 9.3	0.271 (0.033)*	0.714 ± 0.245	0.192 (0.097)	0.680 ± 0.115	0.163 (0.159)
	Mild–Sev (> 2)	42.9 ± 15.7		55.7 ± 11.1		33.5 ± 14.7		42.6 ± 9.7		0.662 ± 0.288		0.661 ± 0.118	
CGI Dep.	Norm–Min (1–2)	38.4 ± 14.6	– 0.713 (< 0.001)*	55.0 ± 10.1	– 0.165 (0.153)	32.0 ± 15.3	– 0.297 (0.011)*	45.5 ± 9.4	0.417 (0.001)*	0.750 ± 0.228	0.550 (< 0.001)*	0.701 ± 0.115	0.632 (< 0.001)*
	Mild–Sev (> 2)	49.0 ± 15.2		56.7 ± 11.1		36.5 ± 14.5		41.6 ± 9.4		0.606 ± 0.297		0.630 ± 0.108	
CGI Cog.	Norm–Min (1–2)	43.4 ± 15.9	0.048 (0.679)	56.4 ± 10.3	0.137 (0.236)	33.9 ± 15.5	– 0.018 (0.879)	43.9 ± 9.8	0.024 (0.853)	0.708 ± 0.271	0.184 (0.111)	0.673 ± 0.122	0.077 (0.504)
	Mild–Sev (> 2)	42.7 ± 15.6		54.9 ± 10.9		34.2 ± 14.6		43.7 ± 9.4		0.659 ± 0.267		0.664 ± 0.111	
CGI Ove.	Norm–Min (1–2)	39.4 ± 14.6	– 0.314 (0.021)*	54.4 ± 10.0	– 0.170 (0.210)	30.7 ± 15.2	– 0.288 (0.033)*	46.6 ± 10.1	0.389 (0.010)*	0.786 ± 0.217	0.498 (< 0.001)*	0.711 ± 0.111	0.479 (< 0.001)*
	Mild–Sev (> 2)	44.3 ± 15.9		56.2 ± 10.7		35.1 ± 14.9		42.9 ± 9.3		0.655 ± 0.278		0.656 ± 0.116	

Cohen’s cut-off: < 0.5 = small, 0.5 < 0.8 = medium, ≥ 0.8 = large; bold formatting indicates medium or large effect sizes, and italics indicates the largest effect size across patient-reported measures. *p* values were calculated using the *F* statistic

Abs absent, *CGI* Clinical Global Impression, *Cog* cognition, *Dep* depression, *EQ-5D* EuroQoL Five-Dimension, *ES* effect size (Cohen’s *d*), *Ext* extreme, *Min* minimum, *Neg* negative, *Norm* normal, *PANSS* Positive and Negative Symptom Scale, *Pos* positive, *Psy* psychopathology, *Sev* severe, *SF-6D* Short-Form Six-Dimension, *SQLS* Schizophrenia Quality-of-Life Scale, *WEMWBS* Warwick-Edinburgh Mental Wellbeing Scale

**p* < 0.05

Table 9 Testing the known-group validity between the patient-reported and NSA-4 and HoNOS-PbR measures

	Groups (score range)	SQLS psychosocial		SQLS motivation		SQLS symptoms		WEMWBS		EQ-5D tariff		SF-6D tariff	
		Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)	Mean ± SD	ES (<i>p</i> value)
NSA-4 rating	No (1)	44.2 ± 15.1	–	57.6 ± 9.4	–	36.5 ± 17.4	–	45.2 ± 10.0	–	0.711 ± 0.269	–	0.683 ± 0.127	–
	Min (2)	39.5 ± 13.8	0.332	53.9 ± 9.4	0.397	31.9 ± 14.7	0.299	46.7 ± 8.2	– 0.170	0.731 ± 0.245	– 0.082	0.702 ± 0.107	– 0.172
	Mild (3)	41.7 ± 15.6	– 0.146	56.6 ± 10.1	– 0.281	32.6 ± 14.2	– 0.052	44.5 ± 8.8	0.258	0.713 ± 0.229	0.078	0.669 ± 0.108	0.307
	Mod–Sev (> 3)	47.3 ± 16.9	– 0.345	56.0 ± 12.2	0.057	36.6 ± 15.2	– 0.272	39.4 ± 10.2	0.543	0.612 ± 0.313	0.366	0.635 ± 0.123	0.289
	–	–	(0.006)*	–	(0.231)	–	(0.102)	–	(<0.001)*	–	(0.012)*	–	(0.001)*
HoNOS-PbR total	No Prob–Mild (0–10)	39.8 ± 14.9	– 0.495 (<0.001)*	55.3 ± 10.5	– 0.101 (0.383)	31.4 ± 14.9	– 0.411 (<0.001)*	45.8 ± 9.0	0.508 (<0.001)*	0.744 ± 0.236	0.517 (<0.001)*	0.694 ± 0.118	0.516 (<0.001)*
	Mod–Sev (> 10)	47.4 ± 15.7	–	56.3 ± 10.6	–	37.5 ± 14.7	–	41.1 ± 9.7	–	0.608 ± 0.294	–	0.636 ± 0.106	–
HoNOS-PbR Beh.	No Prob (0)	39.5 ± 15.1	– 0.420 (<0.001)*	54.8 ± 10.1	– 0.159 (0.167)	32.0 ± 15.3	– 0.254 (0.028)*	45.6 ± 9.7	0.340 (0.008)*	0.731 ± 0.257	0.309 (0.008)*	0.697 ± 0.127	0.447 (<0.001)*
	Prob (> 0)	46.0 ± 15.6	–	56.5 ± 10.9	–	35.8 ± 14.7	–	42.4 ± 9.2	–	0.648 ± 0.276	–	0.646 ± 0.103	–
HoNOS-PbR Fun.	No–Min (0–1)	41.3 ± 15.8	– 0.189 (0.104)	55.1 ± 10.2	– 0.098 (0.397)	30.0 ± 14.3	– 0.480 (<0.001)*	45.6 ± 9.4	0.344 (0.007)*	0.772 ± 0.225	0.581 (<0.001)*	0.699 ± 0.120	0.455 (<0.001)*
	Min–Sev (> 1)	44.3 ± 15.6	–	56.2 ± 10.9	–	37.0 ± 15.0	–	42.3 ± 9.5	–	0.621 ± 0.283	–	0.647 ± 0.110	–
HoNOS-PbR Sym.	No–Min (0–3)	36.3 ± 13.6	– 0.934 (<0.001)*	54.5 ± 10.4	– 0.238	30.4 ± 14.0	– 0.484 (<0.001)*	47.1 ± 8.0	0.700 (<0.001)*	0.763 ± 0.240	0.579 (<0.001)*	0.711 ± 0.115	0.746 (<0.001)*
	Mild–Sev (> 3)	49.6 ± 14.8	–	57.0 ± 10.7	(0.039)*	37.5 ± 15.3	–	40.8 ± 9.9	–	0.612 ± 0.278	–	0.629 ± 0.104	–
HoNOS-PbR Soc.	No–Min (0–2)	38.8 ± 14.7	– 0.508 (<0.001)*	55.1 ± 9.8	– 0.103 (0.373)	31.1 ± 14.6	– 0.361 (0.002)*	46.6 ± 8.9	0.558 (<0.001)*	0.748 ± 0.248	0.428 (<0.001)*	0.700 ± 0.120	0.483 (<0.001)*
	Minor–Sev (> 2)	46.5 ± 15.7	–	56.2 ± 11.2	–	36.4 ± 15.1	–	41.5 ± 9.5	–	0.635 ± 0.278	–	0.644 ± 0.109	–

Cohen's cutoff: < 0.5 = small, 0.5 < 0.8 = medium, ≥ 0.8 = large; bold formatting indicates medium or large effect sizes, and italics indicates the largest effect size across patient-reported measures. *p* values were calculated using the *F* statistic

Beh behaviour, *EQ-5D* EuroQoL Five-Dimensions, *ES* effect size (Cohen's *d*), *Fun* function, *HoNOS-PbR* Health of the Nation Outcome Scale Payment by Results, *Mod* moderate, *Min* minimum, *NSA-4* Negative Symptoms Assessment, *Prob* problem, *SD* standard deviation, *Sev* severe, *SF-6D* Short-Form Six-Dimension, *Soc* social, *SQLS* Schizophrenia Quality-of-Life Scale, *Sym* symptoms, *WEMWBS* Warwick-Edinburgh Mental Wellbeing Scale

**p* < 0.05

providing some reliability to the exploratory construct validity assessment.

A consistent result across these patient-reported measures was a lack of construct validity with schizophrenia-related negative symptoms (i.e., PANSS and CGI negative subscales), suggesting these measures may not be appropriate for assessing negative aspects of schizophrenia, a result echoing those of a previous study [17]. The WEMWBS was relatively more useful in assessing negative symptoms (i.e., CGI-SCH negative subscale and NSA-4), albeit this was not unexpected given it focuses on psychological wellbeing.

4.3 Implications for Clinical Research and Economic Evaluations

Although this study provides some exploratory evidence to suggest moderate construct validity for the generic measures used as part of the analysis, there is limited evidence of strong construct validity and so a new measure focused around mental health is still desirable to improve on the validity performance of these current measures. In particular, patient-reported measures compared with clinician-reported measures appear less able to capture the impact of negative symptoms.

Despite the use of a schizophrenia-specific measure (i.e., SQLS) to capture condition-specific aspects of schizophrenia, the construct validity results suggested the SQLS provided little benefit over the generic HRQoL measures included in the analysis (particularly the SF-6D). This would therefore suggest a need to develop patient-reported measures that better elucidate and quantify aspects of schizophrenia, such as negative symptoms [29]. However, regarding this previous point, the assumption that patient-reported measures are valid only insofar as they approximate to clinician-reported measures is a convenient assumption often used for statistical validation exercises but is flawed and highlights a rather patriarchal approach. Clinician- and self-reported measures are two different perspectives on an individual's experience and can offer complementary (i.e., asking the clinician will produce different but complementary information to that produced by asking the patient) rather than substitutive or equivalent information (i.e., asking the clinician should produce similar information to that produced by asking the patient). It would be unusual for patient-reported and clinician-reported measure scores to show perfect agreement. This difference between patient- and clinician-reported outcomes has been attributed to aspects such as schizophrenia-related cognitive impairment, which impairs the patient's ability to comprehend and report on their own condition [14]. However, individual experiences are not rendered more accurate through some ill-defined hermeneutic process of interpretation, and the subjective-objective discrepancy in measurement may not be considered as

evidence of a failure of the subjective measure. The practical utility of the SQLS Motivation and Symptoms subscales, given the poor construct validity results, should be interpreted with this in mind. Classifying and quantifying the complexities of an individual's experience is a significant challenge, especially with a poorly understood, ill-defined mental health disorder such as schizophrenia. This is further compounded by myriad internal and external factors that influence assessment, attribution, and communication by the patient, not least of which is the person's "mood" at the time of responding to subjective measurement tools, which should be interpreted in this light [18].

4.4 Limitations

Patients in the sample were mainly those with mild symptoms, which negates the generalizability of these results to outcomes measurement in a more severe population. The lack of clinically meaningful cut-offs identified in the literature meant we relied on ad-hoc cut-offs to assess known-group validity. Therefore, these results may not be generalisable to other similar studies, and the cut-offs may have had an effect on the interpretation of known-group validity. However, for this exploratory analysis, the use of ad-hoc cut-offs provided informative results about the ability of the patient-reported measures to detect statistically significant effect sizes between groups, which can be compared with the results of future studies when and if clinically meaningful cut-offs are established for the PANSS, CGI-SCH, and NSA-4. Ceiling effects within particular measures (e.g., EQ-5D) may have affected correlation analysis, but the known-group validity was used to confirm results. The data were also cross-sectional, so change over time could not be assessed, which is an important aspect to assess in the context of economic evaluation.

As stated in the introduction, evidence suggests that, in people with schizophrenia, self-report should be interpreted with caution, as condition-related aspects such as cognitive impairment may impair reliability [14]. This is an inherent concern when using self-reported data for outcomes assessment in patients with mental health conditions that could impair their ability to provide reliable responses; however, we explored the convergent validity between the different patient-reported measures as an estimate of the coherence and consistency of patient self-report and thus the potential impact of invalid or random responses (e.g., due to schizophrenia-related cognitive impairment [14]) on the reliability of the overall construct validity assessment of the measures in this population. This approach offered a practical solution to assessing the consistency of reporting between similar patient-reported measures when it was not possible to use other practical methods of assessing reliability in reporting within measures, such as assessing test-retest reliability,

which assesses intra-observer reliability within measures by asking the person to complete the same measure twice at different (but chronologically close together) time periods [19]. Our analysis suggested that people in our study cohort had reasonably consistent reporting of outcomes between patient-reported measures, as represented by the moderate to strong correlation strength and statistically significant results; however, as we could not specifically control for cognitive impairment (e.g., using regression analysis controlling for a measure of cognitive impairment, such as the Mini Mental State Examination [30]), some aspects of cognitive impairment could have affected our overall assessment, which should be noted when interpreting our exploratory results.

Because of the large number of outcome measures collected, multiple statistical tests were performed, which can increase the likelihood of erroneous inferences. Statistical methods exist to control for issues associated with multiple testing (e.g., the Bonferroni correction [31]); however, because of the stricter statistical significance threshold associated with these methods, a large sample size is required to identify a statistically significant result. Trials tend not to be powered for multiple testing (as most trials are powered based on a single primary outcome), so, for this exploratory analysis using data from a cross-sectional survey study, 300 people were considered an adequate sample size within which corrections for multiple testing were not accounted, which is a limitation. Therefore, future studies may want to recruit larger sample sizes and apply statistical methods to control for multiple tests to confirm or refute the results identified within the current study.

4.5 Considerations for Future Research

Although this study provides exploratory evidence to suggest moderate construct validity for the generic measures used as part of this analysis, there is limited evidence of strong convergent validity, so a new measure focused around mental health is still desirable to improve on the validity performance of these current measures. Related to this point, preference-based measures that can be used for the purpose of clinical assessment and economic evaluation in the field of mental health research are also lacking. Two new measures (which were unavailable at the time of the study that produced the data for this analysis) have undergone initial validation and have been developed to assess QoL in people with different mental health conditions [32]: Recovering Quality of Life (ReQoL) measure with 10 (ReQoL-10; there are plans to make this version preference-based) or 20 (ReQoL-20) items. The intention is that these measures will “plug a gap” in capturing aspects of QoL important to people with mental health conditions (including those with schizophrenia) for clinical assessment and economic

evaluations. Given that the results from this and previous studies have provided mixed evidence for the appropriateness of using the EQ-5D and SF-6D for economic evaluations [11–13, 29], using an alternative preference-based measure in patients with schizophrenia (such as the ReQoL-10) should be explored as part of future research. Understanding how the ReQoL measures compare with existing non-preference-based measures in this patient population (e.g., the SQLS and WEMWBS) will inform the measure’s use for clinical outcome assessment.

5 Conclusion

The exploratory results from this study suggest that the patient-reported measures showed moderate construct validity when assessed against clinician-reported measures for some aspects of schizophrenia severity but showed weak construct validity for the negative symptoms of the condition. In particular, the SF-6D had the best overall construct validity but showed a weak relationship with clinician-rated measures for negative symptoms. Compared with the EQ-5D, the SF-6D may better capture HRQoL aspects of schizophrenia for the purpose of economic evaluation. However, a new measure to assess the burden of schizophrenia is still desirable to improve on the psychometric properties of existing measures for this patient population. There was evidence of consistent reporting of outcomes between patient-reported measures, which provides exploratory evidence that patients with schizophrenia can self-report their HRQoL. There is a suggestion that the SQLS can be self-reported to capture complementary information relative to clinician-reported measures, which is desirable when quantifying the wider burden of schizophrenia.

Author Contributions All authors contributed to the idea about the content of the manuscript and have provided written contributions to the paper, including edits to draft versions. MF led the analysis and writing of the overall manuscript, including final editing and formatting. CM, BM, IT, and JB provided expert oversight to the writing and methodology used. SW led the original UK research study team that collected the data used for the purpose of analysis in this manuscript and provided expert clinical oversight for the writing of the manuscript. All authors act as guarantors for the content of the manuscript.

Compliance with Ethical Standards

Data Availability The full dataset is available from Roche Products Ltd. on request; please contact medinfo.uk@roche.com.

Conflict of interest IT was an employee of Roche Products Ltd. when the work was carried out; he is now an employee of Genentech Inc. BM was an employee of, and currently holds an honorary contract with, ScHARR (University of Sheffield); he is now an employee of the Centre for Health Economics Research and Evaluation (University of Technology Sydney). The authors thank Jennifer Rowell (Roche)

and Stephen Bradley (Roche) for their aid and input during the editing stages of the manuscript before submission for publication. JB developed the SF-6D. JB and BM are EuroQol Group members. MF, CM, and SW have no conflicts of interest that are directly relevant to the content of this article.

Funding This study (NCT01634542) was sponsored by Roche Products Ltd. The writing of this manuscript was part-funded by Roche Products Ltd and the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber (NIHR CLAHRC YH), <http://www.clahrc-yh.nir.ac.uk>. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. All authors were involved in the conception and design, analysis, and interpretation of the data; the drafting of the paper; revising it critically for intellectual content; and the final approval of the version to be published. All authors agree to be accountable for all aspects of the work.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Awad A, Voruganti L. The burden of schizophrenia on caregivers: a review. *Pharmacoeconomics*. 2008;26(2):149–62.
- Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. *Schizophr Bull*. 2004;30(2):279.
- Wilkinson G, Hesdon B, Wild D, Cookson R, Farina C, Sharma V, et al. Self-report quality of life measure for people with schizophrenia: the SQLS. *Br J Psychiatry*. 2000;177(1):42–6.
- Gage H, Family H, Murphy F, Williams P, Sutton J, Taylor D. Comparison of sole nurse and team delivered community clozapine services for people with treatment resistant schizophrenia. *J Adv Nurs*. 2015;71(3):547–58.
- Stewart-Brown SL, Platt S, Tennant A, Maheswaran H, Parkinson J, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. *J Epidemiol Community Health*. 2011;65(Suppl 2):A38–9.
- Maheswaran H, Weich S, Powell J, Stewart-Brown S. Evaluating the responsiveness of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): Group and individual level analysis. *Health Qual Life Outcomes*. 2012;10(1):156.
- Brooks R, Group E. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
- Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–108.
- Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–9.
- National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: NICE; 2013.
- Mulhern B, Mukuria C, Barkham M, Knapp M, Byford S, Brazier J. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. *Br J Psychiatry*. 2014;205(3):236–43.
- Papaioannou D, Brazier J, Parry G. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review. *Value Health*. 2011;14(6):907–20.
- Payakachat N, Ali MM, Tilford JM. Can the EQ-5D detect meaningful change? A systematic review. *Pharmacoeconomics*. 2015;33(11):1137–54.
- Takeuchi H, Fervaha G, Remington G. Reliability of a patient-reported outcome measure in schizophrenia: results from back-to-back self-ratings. *Psychiatry Res*. 2016;244:415–9.
- Kay SR, Opler LA, Lindenmayer J-P. Reliability and validity of the positive and negative syndrome scale for schizophrenics. *Psychiatry Res*. 1988;23(1):99–110.
- Haro J, Kamath S, Ochoa S, Novick D, Rele K, Fargas A, et al. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatr Scand*. 2003;107(s416):16–23.
- Chino B, Nemoto T, Fujii C, Mizuno M. Subjective assessments of the quality of life, well-being and self-efficacy in patients with schizophrenia. *Psychiatry Clin Neurosci*. 2009;63(4):521–8.
- McCabe R, Saidi M, Priebe S. Patient-reported outcomes in schizophrenia. *Br J Psychiatry*. 2007;191(50):s21–8.
- Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2015.
- Nordstroem A-L, Talbot D, Bernasconi C, Berardo CG, Lalonde J. Burden of illness of people with persistent symptoms of schizophrenia: a multinational cross-sectional study. *Int J Soc Psychiatry*. 2017;63(2):139–50.
- Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. *Health Qual life Outcomes*. 2007;5(1):63.
- Alphs L, Morlock R, Coon C, Cazorla P, Szegedi A, Panagides J. Validation of a 4-item Negative Symptom Assessment (NSA-4): a short, practical clinical tool for the assessment of negative symptoms in schizophrenia. *Int J Methods Psychiatr Res*. 2011;20(2):e31.
- Jacobs R. Investigating patient outcome measures in mental health. York: University of York; 2009.
- Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155.
- Parabiaghi A, Kortrijk HE, Mulder CL. Defining multiple criteria for meaningful outcome in routine outcome measurement using the Health of the Nation Outcome Scales. *Soc Psychiatry Psychiatr Epidemiol*. 2014;49(2):291–305.
- Obermeier M. The Positive and Negative Syndrome Scale for schizophrenia: an established rating instrument in need of clarification. 2012.
- Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873–84.
- Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.
- Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess (Winchester, England)*. 2014;18(34):vii.
- Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc*. 1992;40(9):922–35.
- Cabin RJ, Mitchell RJ. To Bonferroni or not to Bonferroni: when and how are the questions. *Bull Ecol Soc Am*. 2000;81(3):246–8.
- Keetharuth AD, Brazier J, Connell J, Bjorner JB, Carlton J, Buck ET, et al. Recovering Quality of Life (ReQoL): a new

generic self-reported outcome measure for use with people experiencing mental health difficulties. *Br J Psychiatry*. 2018;212(1):42–9.

33. NICE. Psychosis and schizophrenia in children and young people. National Institute for Health and Care Excellence: Guidelines; 2013. <https://www.nice.org.uk/guidance/cg155>.

Affiliations

Matthew Franklin¹  · **Clara Mukuria**¹  · **Brendan Mulhern**^{1,2}  · **Irwin Tran**³ · **John Brazier**¹  · **Stuart Watson**^{4,5} 

¹ Health Economics and Decision Science (HEDS), School of Health and Related Research (ScHARR), University of Sheffield, West Court, 1 Mappin Street, Sheffield S1 4DT, UK

² Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW, Australia

³ Health Economics and Strategic Pricing, Roche Products Limited, Welwyn Garden City, UK

⁴ Institute of Neurosciences, Newcastle University, Newcastle, UK

⁵ Northumberland Tyne and Wear NHS Foundation Trust, Newcastle, UK