
Monograph:
Ara, R., Duarte, A., Harnan, S. et al. (4 more authors) (2015) Supporting the routine collection of patient reported outcome measures in the National Clinical Audits for assessing cost-effectiveness. Work Package 1. What patient reported outcome measures should be used in the 2013/14 National Clinical Audit Programme. Appendix I - Schizophrenia. Report. Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) (RR0035). Policy Research Unit in Economic Evaluation of Health and Care Interventions

---

EEPRU

Economic Evaluation of Health and Social Care Interventions Policy Research Unit

RESEARCH REPORT

Supporting the routine collection of patient reported outcome measures in the National Clinical Audits for assessing cost-effectiveness

Work Package 1

What patient reported outcome measures should be used in the 13 health conditions specified in the 2013/14 National Clinical Audit programme?

APPENDIX I, SCHIZOPHRENIA

Authors: Roberta Ara, Ana Duarte, Sue Harnan, Jo Leaviss, Steve Palmer, Mark Sculpher, John Brazier

Correspondence to: Roberta Ara, HEDS, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA. Email r.m.ara@sheffield.ac.uk

RR0035

March 2015

The University Of Sheffield.

UNIVERSITY of York

The Policy Research Unit in Economic Evaluation of Health and Care interventions is funded by the Department of Health Policy Research Programme. It is a collaboration between researchers from the University of Sheffield and the University of York.

The Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions is a 7 year programme of work that started in January 2011. The unit is led by Professor John Brazier (Director, University of Sheffield) and Professor Mark Sculpher (Deputy Director, University of York) with the aim of assisting policy makers in the Department of Health to improve the allocation of resources in health and social care.

INDEX

TABLES

FIGURES

Acronym	Definition
AE	Adverse events
AHRS	Auditory hallucinations ratings scale
AMSTAR	Assessing the quality of systematic reviews
BAI	Beck Anxiety Inventory
BPRS	Brief psychiatry ratings scale
BRAMES	Bech-Rafaelsen melancholia scale
BCVA	Best corrected visual activity
BDI	Becks depression index
BMI	Body mass index (kg/m^2)
CBT	Cognitive behavioural therapy
CG	Clinical guideline
CGI-S	Clinical global impression - severity
DH	Department of Health
EMA	European Medicines Agency
EQ-5D	EuroQol 5 dimensions
FR	Future research
GAF	Global assessment of functioning scale
GARF	Global assessment of relational functioning scale
GSDS	Groningen social disabilities scale
HADS	Hospital anxiety and depression scale
HoNOS	Health of the nation outcome scales
HRQoL	Health related quality of life
HS	Health states
HTA	Health technology assessment
HUI3	Health Utility Index mark 3
ICD-10	International statistical classification of diseases – 10^{th} revision
MBCT	Mindfulness-based cognitive therapy
MID	Minimally important clinical difference
NCA	National Clinical Audit
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PANSS	Positive and negative syndrome scale
PR	Potential recommendations
PREM	Patient Reported Experience Measure
PROM(s)	patient reported outcome measure(s)
Q-LES-Q	Quality of life enjoyment and satisfaction questionnaire
QoL	Quality of life
R&D	Research and development
RCT	Randomised controlled trial
SF-6D	Short form 6D
SF-12	Short Form 12 item
SF-36	Short form 36 item
SCL-90-R	Symptom checklist 90
SOFAS	Social and occupational functioning
SRM	Standardised response mean
STA	Single technology assessment
TTO	Time trade off

UK	United Kingdom
VAS	Visual analogue scale
WHOQOL-BREF	WHO quality of life - BREF
WP	Work package

1. BACKGROUND

EEPRU was approached by Jason Cox (R&D Division) to prepare a programme of research to support the appropriateness of, and use of, patient reported outcome measures (PROMs) collected for the National Clinical Audit (NCA). The EEPRU programme was informed by a R&D template prepared by Simon Bennett, Steve Fairman and Keith Willett at NHS England.

The purpose of introducing PROMs into the NCA programme is to be able to 1) compare performance between providers and commissioners in the National Health Service (NHS), 2) compare the cost-effectiveness of alternative providers in delivering the specific services (i.e. linking outcomes and resource use), and 3) assess the cost-effectiveness of alternative interventions and other changes in the NHS. The intention is to introduce PROMs across a range of conditions over the next 3 years commencing with 13 conditions in the 2014/15 NCA programme.

The agreed research programme consists of 3 concurrent work packages (WP) as described in the document submitted to the Department of Health (DH) (8th November 2013). The current document provides details on the objectives, methodology and results for Work Package 1 (WP1): to determine what PROMS should be used in the 13 health conditions specified in the 2014/15 NCA programme.

2. OVERVIEW

WP1 is split into three separate components consisting of:

- WP1.1 To examine whether the EuroQoL 5 dimensions (EQ-5D) is appropriate in the 13 health conditions specified in the 2013/14 NCA programme.
- WP1.2 To identify what measure could be used when the EQ-5D is not appropriate in the 13 health conditions, taking into account that the proposed measure would be used to generate preference-based utility measures (either directly through existing preference-based weights, or indirectly through existing mapping functions suitable for the proposed measure).
- WP1.3 To identify the evidence required to address questions of cost-effectiveness using the NCA data.

Each component consists of a series of reviews of the literature and the specific review objectives and methodologies are described in detail in the following sections.

3. METHOD

The full detailed methodology used is provided in Appendix A, including the search strategy, selection criteria for studies included, and data extraction etc. In summary, a review of the literature was undertaken to assess the appropriateness of the EQ-5D in terms of classic psychometric criteria (WP1.1); where the EQ-5D was not considered appropriate, additional searches were undertaken to identify alternative measures (WP1.2); and finally, existing health technology appriasials were reviewed and data requirements were compared with variables currently collected in the schizophrenia audit (WP1.3).

3.1 Psychometric properties (WP1.1)

Assessments reported in the included studies were categorised according to the following definitions:

Acceptability

Data relating to how acceptable the measure was to the person completing it, expressed as the proportion of completed surveys, or the proportion of missing data.

Reliability

There are two main definitions for reliability, a) the degree to which a measure reproduces the same results in an unchanged population and b) the degree to which a measure reproduces the same results when completed by different assessors (e.g. patient and proxy report). In both cases, reliability can be assessed by re-testing, and calculating the correlations or difference between tests. In case a) the comparison may be between the same populations separated by time, where no change in health state was observed (as compared to an alternative condition specific or generic measure). In case b) the measure may be completed by multiple people (proxies) on the patient's behalf and their responses compared with those of the patient. Where the outcome measure is specifically designed for self-report by patients, this test of reliability may be expected to produce less agreement.

Construct validity

This is an assessment of how well an instrument measures what it intends to measure. Two main definitions are used in this review.

a) Known group validity, where estimates for groups that are known to differ in a concept of interest are compared either qualitatively or statistically. The known groups may be defined using other measures, according to clinical categorisation.

b) Convergent validity assesses the extent to which a measure correlates with other measures of the same or similar concepts. Correlation coefficients were considered low if <0.3, moderate if between 0.3 and 0.5, and strong when >0.5.

Responsiveness

a) Change over time. This is an assessment of whether measurements using the instrument can detect a change over time, where a change is expected. This may be before and after an intervention, or through progression of a disease. Evidence was considered to be good where a t-test was significant, though weaker evidence to support responsiveness was considered where there was a change in the expected direction, but was not statistically significant or not tested. Effect size and standardised response mean were also acceptable assessments of responsiveness.

b) Ceiling and floor effects were also considered to be indicators of responsiveness. Assessments of ceiling effects include the proportion of patients who score full health within a group of patients with known health detriments. A ceiling or floor effect can affect the sensitivity of the measure in detecting changes over time in patients at the extremes of the measure (for example those with severe disease activity and those with just minor symptoms of the condition).

3.2 **Alternative measures** (WP1.2)

The main sources that were searched for information and recommendations relating to condition-specific or generic measures were:

- Recommendations made in the Oxford set of reviews (http://phi.uhce.ox.ac.uk/)
- The DH PROMS programme
- Recommendations of the Royal Colleges
- European medicines agency (EMA) research guidelines
- Research charity websites.

The recommendations made in these sources were presented and discussed narratively.

3.3 **Evidence required for economic evaluations** (WP1.3)

The existing HTAs were reviewed alongside the variables currently collected in the NCA to determine if clinical or PROM data routinely collected in the NCAs would suffice to address questions of cost-

effectiveness, and to identify any gaps in the evidence that would be required to compare providers, or the cost-effectiveness of interventions or policies.

4. RESULTS FOR SCHIZOPHRENIA

4.1 Evidence of appropriatness of EQ-5D in schizophrenia (WP1.1)

4.1.1 *Selection of systematic review*

Only one systematic review on schizophrenia was identified. (4)

4.1.2 *Structured abstract for Papaioannou et al 2011 (4)*

Purpose of review

The review aimed to investigate the construct validity and responsiveness of four generic health status measures, including two generic HRQL profile measures (Short-form 36 item (SF-36), Short-form 12 item (SF-12)), and two preference-based HRQL measures (SF-6D, EQ-5D) in schizophrenia.

Methods of review

Search and study selection: Ten databases were searched from inception: Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, NHS Economics and Evaluations Database, Health Technology Database, Database of Abstracts of Reviews of Effects, MEDLINE, PreMEDLINE, CINAHL, EMBASE, and Web of Science. Electronic searches were conducted in August 2009. Two sets of search strategies combined terms for each of the four health related quality of life (HRQoL) measures with terms for each of a number of health conditions, of which schizophrenia was one. Only results for schizophrenia were reported in this paper. The full search strategies were not reported.

Inclusion criteria: Studies were included in the review if they satisfied the following criteria: they contained HRQoL data as measured by one of four HRQoL instruments, namely SF-36, SF-12, SF-6D, or EQ-5D; adults ≥18 years old with schizophrenia or schizophrenia-related disorders, e.g. schizophreniform disorder or schizoaffective disorder. Data relating to HRQoL had to be from descriptive systems (their items or dimensions), health state utility values generated by the EQ-5D or SF-6D, or the EQ-5D visual analogue scale (VAS). Studies had to contain data from the HRQoL instrument that allowed measurement of construct validity (convergent or known groups), or responsiveness (effect sizes, standardised response means, or correlation with change scores on symptom measures).

Exclusion criteria: Studies were excluded from the review if the study population were primarily individuals with alcohol and/or drug dependency with comorbid schizophrenia or schizophrenia-

related disorders. Studies that only contained data relating to other psychometric properties were excluded, e.g. reliability, face validity, and content validity.

Data extraction and synthesis: Data were extracted by one reviewer using a newly developed form, designed for specific use in the review. Due to heterogeneity between studies, a narrative synthesis was performed and data tabulated according to the psychometric quality assessed, namely construct validity and responsiveness. Papaioannou et al. used definitions of validity and responsiveness as follows: construct validity, the degree to which an instrument measures the construct it is designed to measure and in the settings it is designed to measure. This can be measured by one of two methods. Known or extreme groups: where two groups who differ in a trait or behaviour, one group is expected to score significantly higher or lower on a particular measure compared with the other group (definition from Streiner 2003); Convergent validity: where the relationship between two instruments measuring the same construct is assessed by Pearson's product movement correlation or Spearman's rank correlation.(7) The review used the following categories for evidence of correlation: >0.6 very strong; ≥0.5 to >0.6 strong; ≥0.3 to <0.5 moderate; <0.3 weak. Secondly, responsiveness was defined as the extent to which an instrument can detect a clinically significant or practically important change over time (definition from Walters 2009).(8)

Authors' conclusions

The authors concluded that the current available evidence for use of the EQ-5D in patients with schizophrenia was mixed, and that there was not enough evidence to recommend the use of EQ-5D in this population to National Institute for Health and Care Excellence (NICE) or other such agencies.

4.1.3 Assessment of the review in relation to objectives of work package 1.1

Relevance of review question: The aim of Papaioannou et al 2011 was concordant with the aims of WP1.1.

Assessment of review quality: Assessment of the quality of the review was conducted using a modified version of the Assessing the methodological quality of systematic reviews (AMSTAR) tool and also by considering the strength and quantity of the evidence.(9) The adequacy of the reported data in the context of WP1.1 was also assessed. A summary of the quality assessment is shown in the Appendix.

Pappaioannou et al. (2011) scored well against most of the relevant AMSTAR criteria. Inclusion/exclusion criteria were clearly defined. Quality assessment of the included studies was conducted and whilst no formal method for assessing the quality of this type of study has been previously validated, methods published elsewhere were followed.(10) However, there was no reference to a published protocol to evidence an a priori design, therefore increasing the risk of reporting bias in terms of changes to the analysis plan in response to the results found. Study selection was carried out by only one reviewer, and double data extraction or data-checking was not conducted, leaving the study at higher risk of errors.

Acceptability of the search: Comprehensive search strategies have been fully reported in the review. A clear description of the iterative approach to the search was applied. Four different iterations were applied: keyword searching; broader terms combined with quality of life (QoL) terms searching; and Quasi QoL terms. The search is considered comprehensive.

Acceptability of study selection: Study selection criteria were clearly defined and concordant with the inclusion criteria for WP1.1

Adequacy of available data and synthesis: The review only provided a small amount of data relating to each study, however this was adequate for the requirements of WP1.1.

In conclusion, the methods employed in the review were generally of an acceptable quality and design to meet the requirements of WP1.1. However, the review concluded that there was not enough evidence to recommend the EQ-5D for use by NICE or other such agencies in this patient population. The searches for the review were conducted in August 2009. As such, an update of this review was conducted, to identify any additional evidence relating to the assessment of the EQ-5D in schizophrenia, and a narrative synthesis combining the studies from Papaioannou et al 2011 (4) with newly identified studies is provided below.

4.1.4 Results of the update and reanalysis of Papaioannou et al 2011(4)

A total of 33 studies were included in Papaioannou et al 2011.(4) Eight of these studies evaluated the construct validity or responsiveness of the EQ-5D. The remaining studies focused on the other generic measures (SF-36, SF-12, SF-6D) and therefore did not meet the inclusion criteria of WP 1.1 and were not included in this synthesis.

Update searches were conducted by EEPRU in May 2014 and retrieved 87 unique titles (search terms given in Appendix B). Of these, the full text of 37 were obtained and considered for inclusion in the review. Two studies met the inclusion criteria and were not already included in Papaioannou et al.(3;6) A total of 10 studies are therefore included in this review.

Of the ten included studies, two studies used the UK EQ-5D tariff.(1;2) Pitkanen et al. used Finnish weights.(3) No further details of which EQ-5D tariff was used for the remaining 7 studies are provided in Papaioannou et al.(4) Only two studies were conducted in the UK.(4-6) The remaining studies were conducted in other European countries, two in Germany,(1;11) two in the Netherlands,(12;13) one in Spain,(14) one in Italy,(15) one in Finland(3) and one not reported.(16)

Patient characteristics were fairly similar across studies. Mean ages ranged from 28.9 years(4) to 41.5 years.(12) Four studies included only patients with schizophrenia, as classified by the ICD-10,(14) Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview,(12) or unreported method of classification.(13;16) Three studies included patients with schizophrenia, schizotypal or delusional disorders as classified by the ICD-10(1;3;11), and one included patients with schizophrenia or schizophreniform disorder.(15) One study included patients with a diagnosis of non-affective psychosis, which included schizophrenia, schizoaffective disorder, bipolar disorder, and psychotic depression),(4) whilst another included young adults (aged 14 to 35) with first-episode psychosis.(6) This diagnosis included schizophrenia, schizophreniform disorder, schizo-affective disorder, biopolar disorders type 1 and 2, delusional disorder and major depressive disorder. The number of withdrawals and study designs were not reported consistently by Papaioannou et al.(17)

A range of measures were used to assess the construct validity and/or responsiveness of the EQ-5D. The majority of measures used were designed for use in mental health conditions, or to capture mental health symptoms: Positive and Negative Syndrome Scale (PANSS), Beck anxiety inventory (BAI), Beck depression inventory (BDI), or Beck Hopelessness Scale (BHS);(4;5;13;15) symptom checklist 90 (SCL-90-R) or clinical global impression-severity (CGI-S);(1;11;14;15) Bech–Rafaelsen melancholia scale (BRAMES);(1) brief psychiatry rating scale (BPRS);(12) Auditory hallucinations rating scale (AHRS).(13) Most also included a measure of functioning: Global Assessment of Functioning Scale (GAF), Social and Occupational Functioning (SOFAS), global assessment of relational functioning scale (GARF), Health of the nation outcome scales (HoNOS);(1;3-5;11;14) or Groningen social disabilities schedule (GSDS).(13) A few studies compared the EQ-5D with another generic quality of life measure: WHO quality of life BREF (WHOQOL-BREF),(1;11;13) Quality of Life

Scale (QLS);(5) or a disease specific quality of life measure: Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q).(3) One study did not conduct a comparison, and instead performed a Rasch analysis.(6)

Acceptability: The review did not report any data for acceptability, and neither did either study identified in the update search.

Construct validity (known group): Two studies reported data on construct validity using the known-groups method.(3;5) Using a minimally important clinical difference (MID) of >0.03, Barton et al. found a significant difference in mean EQ-5D index scores between those with milder and those with more severe symptoms and functioning.(5) Pitkanen et al. dichotomised patients using a cut-off of 50≤GAF or GAF>50, and found a statistically significant difference ($p<0.001$ in both cases) between the means of the two groups for both the EQ-5D and the Q-LES-Q (which measures enjoyment and satisfaction). When comparing across diagnosis sub-groups (ICD-10: F20 vs. F21-F29) no significant difference was reported for either the EQ-5D ($p=0.350$) or the Q-LES-Q ($p=0.354$).(3) Pitkanen et al. also reported that both the EQ-5D and the Q-LES-Q, met the criteria for principle component analysis with all dimensions on both measures scoring over 0.5 (0.4 used as cut-off) and all except mobility in the EQ-5D dimensions scoring above 0.65.(3)

Construct validity (convergent): Seven studies(1;3;5;11;12;14;15) tested the convergent validity of the EQ-5D compared to a variety of other measures such as the PANSS, GAF, Hamilton depression ratings scale, QLS, SOFAS, CGI-S, or BPRS (Appendix). The statistical significance (p-values) were generally not reported in Papaioannou et al. Correlations between the EQ-5D and symptom and functional measures showed differences between studies in the strength of these relationships, ranging from non-existent to strong. Barton explored the relationship between the EQ-5D and three symptoms measures (BAI, BDI, BHS), with resulting correlations ranging from moderate to very strong (r: 0.360 to 0.656). They also found a weak but significant relationship with a measure of functioning (GAF, $r=0.263$). However correlations with the PANSS, were not significant.(5) Scalone 2008 also reported the relationship between the EQ-5D and the symptom measure PANSS, was non-existent or mostly weak.(15) Prieto 2004 showed moderate to strong correlations with both symptom and functional measures (CGI-S and GAF),(14) and McCrone found a moderate correlation ($r=0.343$ p not reported) with the BPRS at baseline.(12) Konig 2007 examined effect sizes for the mean values of symptom and functioning measures between individuals who answered either yes (extreme problems and moderate problems were collapsed into a 'yes' category by Konig due to

small numbers of extreme problems) or no problems on each of the dimensions of the EQ-5D. Effect sizes were moderate to large ranging from 0.37 to 1.29 for symptom measures, and from 0.24 to 1.4 for functioning measures, although the effect sizes for EQ-5D pain/discomfort responses were smaller (data not reported).(11)

For other quality of life measures, Konig 2007 found moderate to strong correlations between EQ-5D and the HoNOS, but weak to moderate correlations with the GARF.(11) Barton 2009 found a weak, non-significant relationship between EQ-5D and both the QLS and SOFAS,(5) and Pitkanen et al. found weak to moderate correlations to the Q-LES-Q items, with correlations ranging from 0.28 to 0.47 (all except mobility greater than 0.30), and a moderate overall correlation between the two measures ($r=0.455$, $p<0.001$).(3)

Responsiveness (changes over time): Four studies assessed responsiveness by examining changes over time (Appendix). Results for responsiveness were again mixed. Positive result were found by Barton 2009 and Badia 1999, with Badia showing large effect sizes (1.13) for olazapine treated patients, and moderate to large effect sizes for patients treated with other antipsychotics (0.78 to 0.96) (NB, assuming olazapine is more effect than the others with no adverse effect on HRQoL).(16) A significant difference in improvement in mean EQ-5D scores between improvers and non-improvers was found by Barton 2009, where improvement was classified as MID (0.03).(5) However, van de Willige (2005) found a lack of responsiveness to change for most of the symptom and functioning measures.(13) Correlations between change scores on the EQ-5D and other clinical measures showed few significant correlations, one for a PANSS subscale (positive), $r= 0.53$, $p<0.001$), and one for a subscale of AHRS (distress, $r= 0.25$, $p<0.01$). Analyses of the relationships between EQ-5D and social function (GSDS) showed a range of correlations (from 0.29 to 0.39, p ranges $p<0.005$ to $p<0.05$). For generic health measures, correlations between EQ-5D and WHOQoL-Bref ranged from 0.25 to 0.58.(13)

Rasch analysis: One study assessed the psychometric properties of the EQ-5D using Rasch analysis in young adults with first episode of psychosis (mean PANSS score =62.7).(6) The authors concluded that the EQ-5D is valid in this population but there was a possible bias for the health dimensions anxiety/depression and usual activities across ethnic groups (white vs. non-white), and was potentially more suitable for comparing mean values across groups of patients (e.g. major depressive disorder vs. schizophrenia), or over time, than for individuals.

4.1.5 Conclusion of appropriateness of EQ-5D in schizophrenia

The evidence base assessing the psychometric properties of the EQ-5D in patients with schizophrenia was relatively large (n-10) and at least nine of these used the UK EQ-5D index. However, not all psychometric properties were reported and overall the results were mixed. The construct evidence (known groups) was good with two studies reporting that the EQ-5D detected differences in the expected direction for known groups, characterised by function (GAF), severity (PANNS) and condition (ICD).(3;5) However, the evidence for both responsiveness and convergent validity was mixed. For responsiveness, while there was some evidence that the EQ-5D was responsive to change, this evidence was limited to the PANSS positive subscale, the Groningen social disabilities schedule (GSDS) and the auditory hallucinations rating scale (AHRS), and no association was found when changes in the BPRS were small (<25%). Finally, despite one study reporting moderate to large effect sizes (ES) for both symptom and function measures, the relationship between the EQ-5D and symptoms (function) was reported as poor in three (three) studies.

In conclusion, there is sufficient evidence to raise doubts about the appropriateness of the EQ-5D in patients with schizophrenia (Table 1).

Table 1: Summary of evidence on EQ-5D for schizophrenia

Measure (N)	**Acceptability**	**Reliability**	**Construct (KGV; Convergent)**	**Responsiveness (Change over time; Ceiling effects)**
Adults				
EQ-5D (11)	Not reported	Not reported	Good; Mixed	Mixed; Not reported
Sufficient evidence to raise doubts about the appropriateness of EQ-5D in schizophrenia				

4.2 Alternative measures in schizophrenia (WP1.2)

The evidence relating to the appropriateness of the EQ-5D in schizophrenia raised sufficient doubts to warrant consideration of other measures in schizophrenia.

Searches identified three reports of relevance to WP1.2 (Appendix)(18-20) and one set of presentation slides from the Royal College of Psychiatrists, which describes the National Audit of Schizophrenia (NAS). The latter is considered in detail in WP1.3, and will not be described here.

Whilst none of the three reports described their methods clearly (at least two were the product of expert panels/working groups, but the process is not described), and none necessarily considered psychometric properties, all three recommend the same measures, namely the BPRS and the positive and negative syndrome scale (PNASS). The EMA report states that these are reliable and validated measures, but does not provide evidence to support this statement.(20) Both measures are clinician-completed, which may be a necessity in this patient group. However, the National Audit of Schizophrenia takes a wider view of outcomes, and includes questions to be completed by the clinician, the patient and carers.

In addition to the literature found through searches, it is worth noting that in the recent introduction of Payment by Results in the area of mental health, providers are mandated to collect a Patient Reported Experience Measure (PREM), Clinician Reported Outcome Measure (CROM) and a PROM.(21) The recommended CROM and PREM are HoNOS and Friends and Family Test respectively (DH 2013). In terms of the PROM, the Warwick Edinburgh Mental Well Being Scale (WEMWBS) or the short version (s-WEMWBS) is currently recommended and both are being tested by the Care Pathways and Packages Project.(22) The WEMWBS was developed for use in the general population and is currently being validated in the area of mental health.

In the meantime, there is recognition that a new measure is needed that would be suitable across the wide spectrum of psychotic and non-psychotic mental health conditions.(23) As already described in section 9.2, the Recovering Quality of Life (ReQoL) instrument, a new preference-based measure in the area of mental health, is being developed by the Policy Research Unit in Economic Evaluation of Health and Care Interventions and is due to be available around July 2015. The ReQoL will have both a long and a short version and will be suitable for use across the psychotic (which will include anxiety and depression) and non-psychotic (which will include schizophrenia) conditions. Once the measure is available and has been validated in people with schizophrenia, the ReQoL may become a candidate measure for inclusion in the NCA.

4.3 **Evidence for economic evaluations in schizophrenia** (WP1.3)

4.3.1 Cost-effectiveness modelling approach used in recent HTAs in schizophrenia

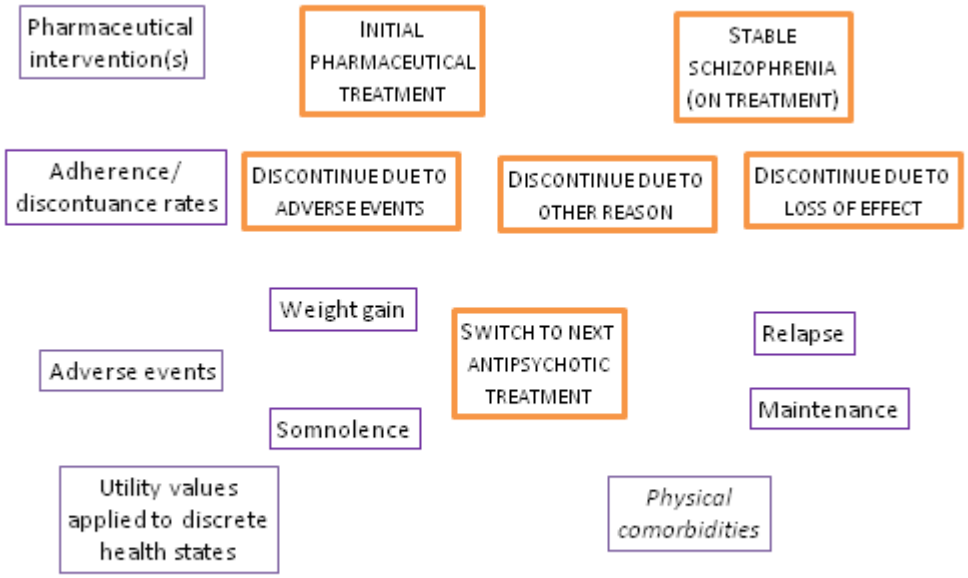
Just one single technology appraisal (STA) relating to schizophrenia was identified from the searches.(24) The evaluation compared pharmacological treatments for schizophrenia in adolescents

(15-17 years). A Markov model with discrete health states based on maintenance and relapse (plus death) was used after an initial decision tree for treatment switching was used to compare the cost-effectiveness of three lines of treatment (Table 2, Figure 1). As treatment related adverse events such as substantial weight gain and somnolence are prevalent, and frequently lead to discontinuation of treatment, these were also captured within the model framework.

Clinical trial data were used to inform the treatment specific probabilities of weight gain (>7%), somnolence and discontinuation due to: lack of efficacy, adverse events or other reasons. The long-term risk of relapse was modelled using relative risks obtained from the literature. Evidence from adults was used due to lack of more suitable data in adolescents.

Quality adjusted life years (QALYs) were obtained by assigning mean utility values to the discrete health states. Again, due to lack of more suitable data in adolescents, EQ-5D data collected from adults were used. The results of the searches conducted to inform the model parameters suggest the volume of EQ-5D data in patients with schizophrenia is very limited and none were available in adolescents.

Figure 1: Modelling approach used in the schizophrenia HTA



Legend: Orange framed boxes with uppercase text describe the health states used in the schizophrenia TA model while the purple framed boxes with lower case (plain) text describe the evidence used. Italised text indicative of additional variables which would be informative for future economic evaluations in psychological therapies.

Table 2: Summary of existing models used in schizophrenia HTAs

	Model method, clinical effect	Method used to model utilities
STA (TA213): Schizophrenia – aripiprazole (in adolescents); 2011(24)		
	Decision tree followed by Markov model Discrete health states: maintenance, relapse, death Effectiveness: probabilities (weight gain, somnolence, discontinuation) Source: RCTs used for clinical effect	Utility: EQ-5D; mean values assigned to discrete HS Source: published literature AEs: disutilities due to weight gain (>7%) and somnolence included

HS: health states; AE: Adverse Events; STA: Single Technology Appraisal; RCT: randomised controlled trial

In summary, the following evidence would be required to compare providers or the cost-effectiveness of interventions for patients with schizophrenia:

- Pharmaceutical intervention(s)
- Compliance to intervention

- Treatment related adverse events (such as weight gain, somnolence)
- Discontinuation rates (and reason)
- Recurrence/relapse rates (with dates)
- Utility values

The majority of this evidence would need to be dated and linked through timings of collection.

4.3.2 Fields collected in the schizophrenia NCA

The NCA for schizophrenia comprises of data collected from eligible Trusts and Health Boards in England and Wales, and collects information on patients diagnosed with schizophrenia (ICD10: F20.0-F20.9) or schizoaffective disorder (ICD-10: F25.0-F25.9) who are treated in the community, including nursing homes and residential care but not inpatients. The fields in the schizophrenia NCA are collected via three questionnaires; an audit tool completed by the psychiatrist accountable for the patient's care,(25) a service user's questionnaire, and a carer's questionnaire. Patients are randomly selected from within the Trust/Health Board from patients who meet the criteria for the National audit of schizophrenia (NAS). The fields are provided in the Appendix.

The mandatory fields provide information on patient characteristics (age, gender, ethnicity); current health status (full remission, partial remission with minimal symptoms, partial remission with substantial symptoms, not in remission) current antipsychotic medications, comorbidity (history of cardiovascular disease, diabetes, hypertension or dyslipideamia), and psychological therapies (cognitive behavioural therapy, family therapy). The optional fields provide additional information on current and historical use of antipsychotic medications for patients not currently in remission; physical health status (smoker, alcohol intake, current measures of: body mass index, blood pressure, lipids, glucose) and interventions offered for these; and whether psychological therapies were taken up by the patient if offered (Appendix).

The service-user and carer questionnaires provide information on experience of, and satisfaction with, the health services provided (Appendix). There are no HRqoL data currently collected in the audit. Both questionnaires are completed anonymously hence cannot be matched to patient records.

4.3.3 Comparing fields in schizophrenia NCA with variables used in existing HTAs

The mandatory fields in the schizophrenia NCA do not provide sufficient detail to model the individual treatment effects (maintenance, relapse, weight gain etc) as applied in the existing HTA cost-effectiveness evaluations in schizophrenia. Although remission (full or partial) is a mandatory field, it is believed these records are subjective clinical decisions. If this is the case, this evidence could be improved through the use of a clinical instrument with clearly defined criteria for remission. The rate of patients in remission could then be used to inform economic models comparing interventions or providers. Some of the optional fields in the clinical audit tool, such as the use of antipsychotic medications and history of medications in patients not in remission, could supply some of the additional evidence required to model the cost-effectiveness of different interventions and policies.

Both the service user and carer questionnaires are principally formed around qualitative questions relating to experiences of the health care services and information provided by the clinicians. Although this is valuable information, it would not be incorporated within an economic modelling which requires information on costs and clinical benefits (i.e. QALYs). There are currently no data collected in the schizophrenia NCA which could be used to inform the HRQoL associated with the condition or the interventions prescribed. As far as we are aware, there are no scheduled plans for imminent inclusion of any generic or condition specific PROM. A mechanism to link the service-user and patient responses to the clinical audits, together with the inclusion of a variable which could be used to generate preference-based utilities would greatly enhance the dataset.

4.4 Recommendations for schizophrenia

Based on the evidence reviewed, the EQ-5D is not thought to be appropriate for patients with schizophrenia. It is not believed that there are data in the schizophrenia NCA which could be used to inform the HRQoL associated with the condition, either directly through a preference-based measure, or indirectly through an alternative measure. In addition, it is not believed that the other variables collected in the audit will suffice to compare providers or conduct robust economic evaluations. Potential recommendations (PR) and areas for future research (FR) are discussed below. All suggested future research areas are indicative and would require a discussion and detailed proposal if required.

It is recommended that both the HoNOS (a clinician-completed measure) and the WEMWBS are collected in the NCA (PR.1). As the WEMWBS is not a preference-based measure, it cannot be used to generate QALYs in economic evaluations. It is therefore recommended that the ReQOL is collected in future rounds of the audit once it has been validated in schizophrenia.(PR.2) It is also recommended that the psychometric properties of the ReQOL is assessed in detail using the data collected in the schizophrenia NCA (FR.1).

The schizophrenia audit does not currently collect suffient detailed information to compare providers or perform economic evaluations. Additional mandatory fields to capture the information required would increase the felexibility of the secondary use of the data (PR.3). Formal recommendations of which fields to include would require a detailed inspection of the exact data collected in the current schizophrenia audit (FR.2).

Table 3: Recommendations and associated future research for schizophrenia

PR.1	*Include both the HoNOS (a clinician-completed measure) and WEMWBS in the schizophrenia NCA*
PR.2	*Include the ReQOL in the NCA once available and validated in patients with schizophrenia*
FR.1	*Assess the psychometric properties of the ReQOL using the data collected in the NCA*
PR.3	*Increase the mandatory fields in the NCA to facilitate future economic evaluations*
FR.2	*Inspect the fields collected in the NCA with a view to making recommendations on the information required to compare providers and conduct economic evaluations*

5. SUMMARY

5.1 Summary of evidence used to inform the conclusions for WP1.1 and WP1.2

An existing review was updated. A total of ten primary research studies were identified. Evidence was mixed (Table 4). Construct validity by known group (defined by severity, diagnosis subgroup and function) was good, but both construct validity by convergent methods and responsiveness by change over time were mixed. The EQ-5D was responsive to change over time in two studies, but the correlations between change scores in another study were only significant between the EQ-5D and an affect subscale (PANSS positive subscale), a social function scale (GSDS) and an auditory hallucination scale (AHRS). Small changes were not reflected in the EQ-5D scores. There are sufficient concerns with the EQ-5D in this population to prevent its recommendation. Three guidelines relating to other measures were identified and all recommended using the BPRS and the PANSS. Within the recently introduced payment by results initiative, the patient-reported measures used are WEMWBS and s-WEMWBS, and a clinician-reported measure, HoNOS, is also used. It is recommended that the same measures are used for the NCA. In this population, it may be useful to include a clinician-reported measure alongside patient-reported measures. ReQoL, could be considered as an alternative once available.

Table 4: Summary of evidence currently available for recommended measure(s)

Condition	N	Acceptability	Reliability	Construct		Responsiveness		Overall
				KGV	Convergent	Change over time	Ceiling Effect	
EQ-5D	10	NR	NR	Good	Mixed	Mixed	NR	Not appropriate
HoNOS (clinician-completed)		The recommendation is based on those in PBR [DH2013] and the psychometric properties of this measure have not been reviewed in the current report						
WEMWBS		The recommendation is based on those in PBR [DH2013] and the psychometric properties of this measure are currently under review elsewhere [DH3013]						
ReQOL		This measure is currently in development and will be available in 2015						

N= number of studies used to inform conclusions, KGV: known group validity; NR, the existing review did not review this psychometric property.

5.2 Summary of evidencerequired for use in economic evaluations (WP1.3)

Although the audit includes a service user and carer questionnaire, these do not currently include a PROM, concentrating of experience of and satisfaction with the health services provided. These data will be useful when comparing providers, but cannot be used to inform economic evaluations. It is not believed that the mandatory fields in this audit provide sufficient detail to model individual

treatment effects (maintenance, relapse, compliance, weight gain etc), but some of the optional fields could provide some evidence on antipsychotic medications and history of medications in patients not in remission. While there is a mandatory field relating to 'relapse' it is believed this is a subjective clinical decision, thus it may not be possible to use this in economic models.

APPENDIX: SCHIZOPHRENIA

The tables in this Appendix provide additional information for the reviews (WP1.1, 1.2 and 1.3) conducted for Schizophrenia.

Table A1: Quality assessment of Papaioannou et al 2011 systematic review of schizophrenia.(4)

Quality assessment criteria	**Compliance with criteria**
AMSTAR	
Was an a priori design provided?	Yes
Was there duplicate study selection and data extraction?	No
Were the methods used to combine the findings of the studies appropriate?	Yes, narrative synthesis due to heterogeneity
Was the scientific quality of the included studies assessed and documented?	Yes, using method described in Fitzsimmons et al.(26)
Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes
Overall judgement of quality of review	Good but only 1 reviewer
Quality of the searches	Acceptable
Strength of the evidence	
Were the conclusions robust and conclusive?	No, evidence was mixed and limited
Quantity of the evidence	
Was there enough data to be confident that any additional data published subsequently would be very unlikely to change the conclusions drawn?	No because evidence was mixed
Adequacy of data reported	
Did the review provide sufficient data to allow integration of an update/assessment of the methods used?	Yes
Did the review assess EQ-5D in a way compatible with the aims of work package 1.1?	Yes, construct validity (known groups or convergent) or responsiveness (effect sizes, standardised response means, or correlation with change scores on symptom measures).

Table A2: Characteristics of primary studies for Schizophrenia. Partly adapted from Pappaionnou et al 2011.(4)

Author, year	Study design	Condition	Study information	Male/female	Mean (SD) age at baseline in years
Badia, 1999, country not reported(16)	N/R	Schizophrenia (classification N/R)	N=approx 2949 n=2128 olanzapine n=821 risperiodone or haloperidol; small numbers on other antipsychotics	N/R	N/R
Barton, 2009, UK(5)	N/R	Current diagnosis of affective or non affective psychosis including schizophrenia, schizoaffective disorder, bipolar and psychotic depression in relative remission (PANSS≤4)	N=77	55/22	28.9 range 18-52
Konig, 2007, Germany(1)	N/R	ICD-10: schizophrenia, schizotypal or delusional disorders	N=166	97/67	40.5 (11.1) range 21-80
Konig, 2009, Germany(27)	N/R	ICD-10: schizophrenia, schizotypal or delusional disorders	N=143	83/60	40.4 (11.6)
McCrone, 2009, Netherlands(12)	N/R	SCAN interview diagnosed schizophrenia (classification scheme not specified)	N=409	245/164	41.5 (11.5)
Prieto, 2004, Spain(14)	N/R	ICD-10 schizophrenia	N=2657 n=2128 olanzapine, n=417 risperidone, n=112 haloperidol	1691/966	35.32 (11.57)
Scalone, 2008, Italy(15)	N/R	Schizophrenia, schizophreniform disorder	N=637 n=551 schizophrenia, n=86 schizopherniform disorder	414/223	N/R, range 18-40
Van de Willige, 2005, Netherlands(13)	N/R	DSM-IV schizophrenia (chronic sample). Auditory hallucinations > 2 years, use of at least 2 antipsychotic drugs	N=76	42/34	36 (11.2)
Studies from update search					
Pitkänen, 2011, Finland(3)	Cross-sectional	Hospitalised patients with diagnosis of schizophrenia, schizotypal disorders	N=311 41% schizophrenia 29% non-organic psychotic	183/128	38 (13) Range 18-65

Author, year	Study design	Condition	Study information	Male/female	Mean (SD) age at baseline in years
		or delusional disorders (ICD10: F20–29)	disorder 16% schizoaffective disorder 6% acute and transient psychotic disorder 5% persistent delusional disorder 2% schizotypal disorder 1% other non-organic psychcotic disorder		
Stochl, 2013, UK(6)	Cross-sectional	Young adults (aged 14 to 35) with first-episode psychosis	N=1,027 recruited, 714 with complete data and known diagnosis. 22% Schizophrenia 4% Schizophreniform disorder 6% schizo-affective disorder 2% bipolar disorder 1 4% Bipolar disorder 2 4% delusional disorder 7% major depressive disorder 29% psychosis, not otherwise specified 21% diagnosis not available	709/318	Median 22 (IQR 19 to 26) Range 14 to 35

PANSS: Positive and Negative Syndrome scale

,

Table A3: Characteristics of primary studies for schizophrenia. Partly adapted from Pappaionnou et al 2011. (4)

Author, Year, Location	EQ-5D	Comparison measure	Psychometric properties assessed	Assessment of psychometric properties
Badia, 1999, country not reported(16)	EQ-5D Index and EQ-5D VAS	No measures reported	Responsiveness	Effect sizes for EQ-VAS and EQ-5D index (no further details)
Barton, 2009, UK(5)	EQ-5D Index	Patient completed: i)Severity of mental health symptoms: BAI, BDI, BHS, GAF Clinician-completed: i)Severity of mental health symptoms: PANNS ii)Functioning: QLS, SOFAS	Construct validity – known-groups	Differences in EQ-5D scores (% reporting problems on EQ-5D health dimensions) when sub-grouped by severity: BAI (≤18 vs. ≥19); BDI (≤19 vs. ≥20); BHS (≤8 vs. ≥9); GAF (≤60 vs. ≥61); SOFAS (≤60 vs. ≥61); for all above and for PANNS and QLS also compared sub-groups of equal size (when ranked by severity)
			Convergent validity	Correlation between EQ-5D and all other measures
			Responsiveness	Mean EQ-5D scores for sub-groups who improved (post intervention) assessed by clinical measures (BAI, BDI, NHS, QLS, PANSS, GAF, SOFAS)
Konig, 2007, Germany(11)	EQ-5D Index and EQ-5D VAS	Clinician-completed: i)Symptoms PANSS, SCL-90R & CGI-S ii)Functional GAF, GARF, SOFAS & HoNOS Patient-completed: i)Quality of life-generic TTO direct utility & WHOQOL-BREF	Convergent validity	Effect sizes calculated using the mean values of symptom and functioning measures between individuals who answered 'yes' or 'no' for each EQ-5D dimension Correlations between EQ-5D VAS and index and symptom/functioning measures
Konig, 2009, Germany(1)	EQ-5D Index (UK and German)	Clinician-completed: i)Symptoms PANSS, SCL-90R, CGI-S, and BRAMES ii)Functional GAF, GARF, SOFAS, and HoNOS Patient-completed: i)Quality of life-generic TTO direct utility & WHO-QOL-BREF	Convergent validity	Correlations between EQ-5D and symptom and functional measures
McCrone, 2009, Netherlands(12)	EQ-5D Index	Clinician-completed: i)Symptoms BPRS	Convergent validity	Correlations between EQ-5D Index and symptom measure at baseline and change after treatment
			Responsiveness	SRM
Prieto, 2004, Spain(14)	EQ-5D Index	Clinician-completed: i)Symptom CGI-S	Convergent validity	Correlations between EQ-5D and symptom and functional measures

Author, Year, Location	EQ-5D	Comparison measure	Psychometric properties assessed	Assessment of psychometric properties
		ii)Functional GAF		
Scalone, 2008, Italy(15)	EQ-5D	Clinician-completed: i)Symptom PANSS, CGI-S ii)Functional GAF	Convergent validity	Correlations between EQ-5D and symptom and functional measures
Van de Willige, 2005, Netherlands(13)	EQ-5D Index	Clinician-completed: i)Symptom PANSS, AHRS ii)Functional GSDS iii)Quality of life-generic WHOQOL-BREF	Responsiveness	Differences in EQ-5D descriptive system scores between baseline and follow-up. Correlations between changes in EQ-5D and symptom and functional measures
Studies from update search				
Pitkänen, 2012, Finland(3)	EQ-5D (Finnish weights)	Q-LES-Q (general activities) Nurse completed: GAF	Internal consistency	Correlation between EQ-5D health dimensions and 14 items on Q-LES-Q Sub-groups defined by GAF scores Spearman's correlation of overall scores.
			Construct validity	Principle component analysis Mann Whitney used to test for differences in EQ-5D index and Q-LES-Q total sub-group by GAF ≤50 vs. GAF>50, and diagnosis group (Dg) F20 vs F21-29
Stochl, 2013(6)	EQ-5D (UK)	Diagnostic subtypes	Rasch analysis	Item response modelling Relationship between EQ-5D and diagnosis Differential item functioning

BAI: Beck Anxiety Inventory; BDI: Beck Depression Inventory; BHS: Beck Hopelessness Scale; PANSS: Positive and Negative Syndrome Scale; GAF: Global Assessment of Functioning Scale; QLS: Quality of Life Scale; SOFAS: Social and Occupational Functioning; S-QOL, schizophrenia quality of life scale; Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire; Db, diagnosis group; BRAMES, Bech-Rafaelsen Melancholia Scale;

Table A4: Convergent validity results for schizophrenia, adapted from Pappaionnou et al 2011 (4)

Author, year	Method of measuring convergence (e.g. Spearman rank correlation, statistical significance)	Convergent validity results (NB any subjective decisions of 'strength' are as reported in Pappaionnou)
McCrone, 2009(12)	Correlations between EQ-5D Index and symptom measures: Clinician-completed: i) Symptoms BPRS	EQ-5D and BPRS: r=0.343 (moderate)
Konig, 2007(11)	Effect sizes calculated using the mean values of symptom and functioning measures between individuals who answered 'yes' or 'no' for each EQ-5D dimension * Correlations between EQ-5D VAS and index and symptom/functioning measures: Clinician-completed: i)Symptoms PANSS, SCL-90R & CGI-S ii)Functional GAF, GARF, SOFAS & HoNOS Patient-completed: i)Quality of life-generic TTO direct utility & WHOQOL-BREF	Effect sizes mostly moderate to large for symptom measures (0.37-1.29) and functioning measures (0.24-1.4) Effect sizes for pain/discomfort were smaller Moderate correlations between EQ-5D VAS and index and symptom measures (0.34-0.73), functioning measures (0.20-0.65), and generic QoL measures (0.47-0.57)
Konig, 2009(1)	Correlations between EQ-5D and symptom and functional measures: Clinician-completed: i)Symptoms PANSS, SCL-90R, CGI-S, and BRAMES ii)Functional GAF, GARF, SOFAS, and HoNOS Patient-completed: i)Quality of life-generic TTO direct utility & WHO-QOL-BREF	Correlation with the TTO direct elicitation of utility values and the EQ-5D VAS and EQ-5D index (UK and German) were weak in correlation (0.25). However, the TTO method did not correlate well with a number of theoretically related measures.
Prieto, 2004(14)	Correlations between EQ-5D and symptom and functional measures: Clinician-completed: i)Symptom CGI-S	Moderate to strong correlation with one symptom (CGI-S) and one functional measure (GAF) (range 0.34-0.54, p<0.001)

	ii)Functional GAF	
Barton, 2009(5)	Correlations between EQ-5D and symptom measures: Patient completed: i)Severity of mental health symptoms: BAI, BDI, BHS, GAF Clinician-completed: i)Severity of mental health symptoms: PANNS ii)Functioning: QLS, SOFAS	Correlations between EQ-5D index and three symptom measures (BAI, BDI, BHS) were moderate to very strong (r=0.360-0.656) Correlation between EQ-5D and GAF weak but significant (0.263). Non significant and weak correlations with PANSS, QLS, and SOFAS
Scalone, 2008(15)	Correlations between EQ-5D and symptom and functional measures: Clinician-completed: i)Symptom PANSS, CGI-S ii)Functional GAF	Weak to moderate correlations between EQ-5D and symptom measures (PANSS and CGI-S), range 0.189-0.393
Studies from update search		
Pitkänen, 2011(3)	Correlations between EQ-5D index, and EQ-5D health dimensions and Q-LES-Q items	Item correlations ranged from 0.28 to 0.47 (all except mobility greater than 0.30)
	Correlation between overall index scores	Spearman's correlation co-efficient shows moderate correlation, r=0.455; p<0.001

* Note, this is as reported in Papaioannou et al.,[Papaioannou 2011] and as categorised in the original study[Konig 2007] but might also be considered to be responsiveness (effect sizes).

Table A5: Construct validity results (known groups) for schizophrenia, adapted from Pappaionnou et al 2011. [Pappaioannou 2011]

Author, year	Method of measuring construct validity (e.g. known groups)	Construct validity results (NB any subjective decisions of 'strength' are as reported in Pappaionnou)
Barton, 2009(5)	Known groups: clinically significant differences in EQ-5D index scores (defined as >0.03) according to disease severity groups measured by: Patient completed: i)Severity of mental health symptoms: BAI, BDI, BHS, GAF Clinician-completed: i)Severity of mental health symptoms: PANNS ii)Functioning: QLS, SOFAS	Minimally important clinical difference (MID) (>0.03) between those with milder and more severe scores on seven of the symptom and functioning measures, including PANSS, Hamilton depression rating scale, and GAF.
Studies from update search		
Pitkänen, 2012(3)	Known groups: a)split by 50 ≤GAF or GAF>50 Known groups: b)split by ICD-10 F20 vs. F21-F29 Mean EQ-5D index compared to mean total Q-LES-Q	EQ-5D: a) 0.64 vs. 0.89 (p<0.001), b) 0.81 vs. 0.75 (p=0.350) Q-LES-Q: a) 35 vs. 45 (p<0.001), b) 43 vs. 42 (p=0.354)
	Principal component for both EQ-5D and Q-LES-Q	EQ-5D met criteria (Kaiser-Meyer-Olkin =0.701; Bartlett's test (χ^2= 166, df=10, P<0.001) for component analysis. Q-LES-Q met criteria (Kaiser-Meyer-Olkin =0.889; Bartlett's test (χ^2= 1117, df=105, P<0.001) for component analysis.

Table A6: Responsiveness results for schizophrenia, adapted from Pappaionnou et al 2011. (4)

Author, year	Method of measuring responsiveness	Responsiveness results (NB any subjective decisions of 'strength' are as reported in Pappaionnou)
McCrone, 2009(12)	EQ-5D SRMs for BPRS response groups (>25% versus <25%).	Where improvement on BPRS was at least 25%, EQ-5D SRM was small in size (0.39). Where deterioration on BPRS was at least 25% or improvement on BPRS <25%, EQ-5D SRMs were very small (0.17 to 0.05 respectively).
Van der Willige, 2005(13)	Differences in EQ-5D descriptive system scores between baseline and follow-up: Correlations between change scores for EQ-5D vs. other measures: Clinician-completed: i)Symptom PANSS, AHRS ii)Functional GSDS iii)Quality of life-generic WHOQOL-BREF	Statistically significant differences for daily functioning (Z=1.79, p>0.05 <0.10), and anxiety and depression (Z=3.53, p<0.001) Moderate correlations between changes on EQ-5D VAS and changes in PANSS (0.34-0.47, p>0.01 and p<0.0005). Correlations between changes on EQ-5D index and changes in PANSS existed only on PANSS positive symptoms subscale (0.53, p<0.0001). Moderate to strong correlations with 3 of 4 AHRS subscales and the EQ-5D VAS (0.46-0.50, p<0.001)
Barton, 2009(5)	Mean difference in EQ-5D score for improvers vs non improvers (based on improvement on 6 of 7 symptom or functioning measures). MID 0.03. Patient completed: i)Severity of mental health symptoms: BAI, BDI, BHS, GAF Clinician-completed: i)Severity of mental health symptoms: PANNS ii)Functioning: QLS, SOFAS	Difference in mean EQ-5D scores between improvers and non-improvers (categorised as improvement on 6 out of 7 clinical measures) was equal to or greater than the MID (0.03)
Badia, 1999(16)	Effects sizes by treatment groups for EQ-VAS and EQ-5D index. Olanzapine versus other antipsychotics: risperodone, haloperidol or other).	Large effect sizes for olanzapine-treated patients pre- and post-treatment EQ-5D index: 1.13 Moderate to large effect sizes for other antipsychotics: EQ-5D index: 0.78 to 0.96)

Table A7: Rasch analysis results for schizophrenia

Author, year	Method of measuring validity	Results (NB any subjective decisions of 'strength' are as reported in Pappaionnou)
Stochl 2013(6)	Item response modelling by PCM and GPCM, using maximum likelihood estimation.	All items discriminate equally well. Akaike information criterion slightly preferred GPCM (AIC= 5,265) over PCM (AIC= 5,269). Bayesian information criterion supported PCM (BIC=5,321) over GPCM (BIC=5,336). BIC is preferred in this instance as it penalizes more for model complexity and is more suitable for descriptive purposes.
	RSM	Likelihood ratio test between RSM and OCM was non-significant (likelihood difference = 3.8, df=4, p=0.43). Global fit indices favoured RSM (PCM: AIC = 5,249.3, BIC = 5,292.1; RSM: AIC = 5,248.9, BIC = 5,272.7). Some evidence of over-fit, indicating items overlap, but no remedial action recommended.
	Item-person map of RSM	Items are well spaced along the intended latent measurement continuum. This confirms that EQ-5D can discriminate at all levels of HRQoL, regardless of the status of the patient.
	Rasch estimates of HRQoL and corresponding standard errors	Standard errors of measurement are quite large for reliable placement of any individual on the HRQoL latent continuum. Even less information (and thus less measurement precision) is available for people reporting lack of problems with any of the EQ-5D dimensions and for people with nearly maximal sum scores. EQ-5D is therefore more suitable for group comparisons or profiling of improvement over time for groups of patients rather than for individual psychiatric assessment of HRQoL.
	Relationship between EQ-5D and diagnosis (expect no relationship as EQ-5D is not a diagnostic measure)	4-class model (latent class analysis) best fits data. The chi-squared shows a significant (p = 0.03) relationship between diagnosis and EQ-5D latent class summaries of HRQoL profiles; a higher proportion of patients suffering from major depressive disorder and schizoaffective disorder belonged to class 3—the most [HRQoL] impaired class—in comparison with other latent classes and diagnostic categories. This was due to the anxiety/depression item, which has a direct relationship to one of the main symptoms of schizophrenia. Removal of anxiety/depression item saw the chi-squared become non-significant (p=0.27)
	Differential item functioning	No DIF for gender or diagnosis. DIF for ethnicity, indicating EQ-5D may vary across patients with different ethnic minority group status. Statistically significant for white patients in anxiety/depression and usual activities: more likely to score some or severe problems. This should be controlled for in analyses.

PCM, partial credit model; GPCM, generalise partial credit model; AIC, Akaike information criterion; BIC, Bayesian information criterion; RSM, rating scale model;

Table A8: Mandatory fields collected in the schizophrenia NCA

TRUST/HEALTH BOARD AND TEAM[a]
NAS patient ID, Initial of data collector/clinician, Clinical team responsible for the patient's care (Assertive outreach team, community mental health team, crisis resolution team, Early intervention in psychosis team, other)
PATIENT DETAILS[a]
Year of birth, Sex, Ethnicity, Current ICD-10 mental health dimensions (F20, F25), How long ago was this diagnosis first made (between 1-2 years, up to 4 years, up to 10 years, more than 10 years)
PATIENT'S CURRENT MENTAL HEALTH[a]
Use knowledge of the patient to rate patient's current mental health (on clozapine/not on clozapine: full remission, partial remission with minimal symptoms and disability, partial remission with substantial symptoms and disability, not in remission)
ANTIPSYCHOTIC MEDICATIONS[a]
Is the patient currently prescribed antipsychotic medications (Y/N), Was the patient provided with written information (or an appropriate alternative) about the most recent antipsychotic prescribed (Y/N, don't know, patient never prescribed an antipsychotic), For the most recently prescribed antipsychotic, was the patient involved in deciding which antipsychotic they were prescribed
HISTORY OF PRESCRIBING FOR PATIENTS NOT IN REMISSION AND NOT CURRENTLY PRESCRIBED CLOZAPINET[a]
Not mandatory
HISTORY OF PRESCRIBING FOR PATIENTS NOT IN REMISSION AND CURRENTLY BEING PRESCRIBED CLOZAPINE
Not mandatory
PHYSICAL HEALTH MONITORING[a]
Does the patient currently have any of the following significant physical health problems (and is it recorded in your case record or the GP's records): Cardiovascular disease, diabetes, hypertension, dyslipidaemia
PHYSICAL HEALTH RECORDS
Not mandatory
PSYCHOLOGICAL THERAPIES
Has cognitive behavioural therapy EVER been offered to the patient, Has family intervention (where patient is in contact with family) EVER been offered to the patient
CARE PLAN
Not mandatory

[a] collected via the NAS audit of practice form

Table A9: Optional fields collected in the schizophrenia NCA (WP1.3)

TRUST/HEALTH BOARD AND TEAM[a]
No additional non-mandatory fields
PATIENT DETAILS[a]
No additional non-mandatory fields
PATIENT'S CURRENT MENTAL HEALTH[a]
No additional non-mandatory fields
ANTIPSYCHOTIC MEDICATIONS[a]
Provide current dose of all antipsychotics currently being prescribed for the patient (list), If the current antipsychotic dose is known to be above the BNF recommended dose, has a rationale for this been documented in the patient's records, If the patient is currently being prescribed two or more antipsychotic drugs at the same time, has a rationale for this been documented in the patient's records, How long has the patient been on clozapine, What was the reason for starting the patient on clozapine (Treatment resistant/poor response, adverse effects from previous antipsychotic medication, both of the above, not known, other, How many antipsychotic medications was the patient prescribed before clozapine.
HISTORY OF PRESCRIBING FOR PATIENTS NOT IN REMISSION AND NOT CURRENTLY PRESCRIBED CLOZAPINET[a]
Why is this patient not currently prescribed clozapine (not yet had adequate trial of two other antipsychotices, clozapine contratindicated for this patient, clozapine tried but patient did not respon adequately, clozapine offered but patient refused, ongoing anxiety and depression but not psychotic symptoms, Trust restrictions on use of clozapine, waiting for an inpatient bed, lack of facility for community initiation, Is the current antipsychotic the first antipsychotic medication prescribed for the patient, How many other antipsychotics did the patient receive before the current one, How long has the patient been on the current antipsychotic medication, In the past 12 months, has medication adherence been investigated as a potential cause of inadequate response to antipsychotic medications, In the past 12 months, has alcohol or substance misuse been investigated as a potential cause if inadequate response to antipsychotic medications
HISTORY OF PRESCRIBING FOR PATIENTS NOT IN REMISSION AND CURRENTLY BEING PRESCRIBED CLOZAPINE
Before starting clozapine was the patient trialled on at least two second generation antipsychotics, If the patient is currently prescribed clozapine plus another antipsychotic, has the patient been trialled on this combination for at least 8 weeks at optimal dose, In the past 12 months, has medication adherence been investigated as a potential cause of inadequate response to antipsychotic medications, In the past 12 months, has alcohol or substance misuse been investigated as a potential cause if inadequate response to antipsychotic medications
PHYSICAL HEALTH MONITORING[a]
No additional non-mandatory fields
PHYSICAL HEALTH RECORDS
Smoking status (current smoker, non smoker), Current alcohol intake, Current substance misuse, Current most recent BMI, Waist circumference, Current/most recent blood pressure, Current/most recent glucose, Current most recent cholesterol, Family history of cardiovascular disease, diabetes, hypertension, dyslipidaemia, Has any interventions been offered, or a referral made, within last 12 months for any of the following: advice about diet/exercise, treatment for cardiovascular disease, diabetes, dyslipidaemia or hypertension, Help with: smoking cessation, reducing alcohol consumption or reducing substance misuse. What was the source of the information used to answer questions (case record and psychiatrist knowledge, case record, psychiatric knowledge and GP, GP only, other)
PSYCHOLOGICAL THERAPIES
If psychological therapy offered, was this taken up by the patient (CBT or family intervention: yes, no patient refuse, no reason not recorded, do not know, other)
CARE PLAN
Does the patient have a current care plan

[a] collected via the NAS audit of practice form, CBT: Cognitive based therapy

Table A10: Non-mandatory fields collected in the schizophrenia NCA (2) (WP1.3)

SERVICE USER: questions relate to experiences within the previous 12 months (Anonymous) *THE QUALITY OF CARE YOU RECEIVED*[a]
Are you satisfied with the times and places of your appointments, Are you satisfied with the amount of time available for talking with members of the service about your problems, Do you feel confident that members of the service are competent in dealing with your problems, Taking everything into consideration, are you pleased with the care you have received for the service so far, How satisfied were you with the service you received at your GP surgery during the last 12 months.
GETTING HELP FROM PEOPLE YOU KNOW WHEN YOU NEED IT[a]
Do you have a key worker or care coordinator, Do you know how to contact your key worker or care coordinator, How satisfied are you with your access to your key worker or care coordinator, Has there been a change in your key worker or care coordinator in the last year, Has there been a change in your psychiatrist in the last year, Do you know how to get help for your mental health if there is a crisis or emergency and you need help right away, Do you have a care plan that provides you and other people with information about what your main mental health issues are and what help you are getting with these, Do you have an advance directive that provides you and other people with information about what you would like to happen should you become unwell
MEDICATION[a]
Was the purpose of the current medication for your mental health explained to you, including what could happen if you stopped taking it, Were the side effects of the medication discussed with you, Were your views taken into account when deciding which medication to prescribe, Were you given written or online information about your medication
YOUR PHYSICAL HEALTH[a]
Has your weight been checked by a nurse or doctor in the last 12 months, Has your blood pressure been checked by a nurse or doctor in the last 12 months, Has your blood pressure been checked by a nurse or doctor in the last 12 months, Have you had blood tests carried out in the last 12 months, In relation to smoking cigarettes (I smoke and am getting help to stop; I smoke and am not getting help to stop; I smoke and do not want help to stop; I do not smoke)
OTHER TYPES OF TREATMENT AND HELP[a]
In relation to work and employment (I do not have a job but am getting help to find one, I do not have a job and am not getting help to find one, I do not have a job and am not looking for one, I have job), In relation to other activities: I am involved in activities during my day (e.g. education, drop-in group) etc, In relation to Cognitive behavioural therapy (I have had or am receiving this treatment, I have not had this treatment, I do not want this treatment), In relation to family intervention/therapy (I have had or am receiving this treatment, I have not had this treatment, I do not want this treatment)
OVERALL
To what extent have services helped you to achieve good mental health in last year (helped a lot, helped a little, made no difference, made me worse)
CARER SURVEY (questions relate to experiences within the previous 12 months (Anonymous)[b]
Information and advice for carers: Enough information about the condition/illness of the person you care for to enable you to feel confident in caring for them, enough information about how their condition/illness is likely to develop in the longer term? that you can get whatever information you need when you need it (e.g. through your doctor or on your own)? with how easy it is to understand the information you have? with the amount of advice available to you (e.g. from healthcare workers or other carers) that you are clear about who to go to for the information and advice you need? that you are clear about who to contact if there is an emergency and you need help right away? that you are clear about who to call if you have a routine enquiry, Your involvement in treatment and care planning: In general, how satisfied were you with: important decisions (e.g. medication or hospitalisation, Ability to influence important decisions. Support from medical and/or care staff: In general, how satisfied were you with: how easy it was to get help and support from staff for the person you care for (e.g. to prevent relapse)? how easy it was to get help and support from staff for yourself (e.g. advice on how to deal with certain behaviours)? the quality of help and support from staff for the person you care for? Your relationships with key staff who support the person you care for, How well the staff are communicating with each other (i.e. that they share important information), How seriously staff take what you say to them, the level of understanding staff have of what it must be like to be in your situation.

CARER SURVEY (Background information)

Year of birth, Gender, Ethnic background, Employment status, In what year did you first start caring for someone with a mental health problem, Estimate how many hours you spent in the last week looking after someone with a mental health problem, Was this: more hours than usual, about the same number of hours as usual, fewer hours than usual, Who is the person whom you care for in relation to you (son/daughter, partner, brother/sister, parent, friend, other, Do you live with each other at the moment, If no, where are they currently living, Which of the following statements best describes your role as a carer at the moment: I am the only caregiver, I share caring responsibilities with others, but a I am the main caregiver, I share caring responsibilities with others, I share caring responsibilities, but someone else is the main caregiver, Other, How many people with mental health problems do you currently care for

Table A11. Reports of other measures in schizophrenia (WP1.2)

Source, date	Population	Method used to reach recommendation			Measures recommended	Implementation issues for large scale use?
		General methods	**Psychometric properties considered?**	**Measures considered**		
Royal College of Psychiatrists guideline: adults 2012(18)	Adult psychiatry (section on psychotic disorders)	Methods not described.	Unclear	Unclear	BPRS – 24 symptom constructs to be rated on 7-point scale PANSS – interview (45-50 mins), clinician rated KGV – interview covering 14 symptoms, clinician rated SSPI – 20 item	Most appear to be clinician-rated, and involve an interview.
Royal College of Psychiatrists guideline: older adults(19)	Older adults	Authored by a working group	Unclear	Unclear	BPRS – 24 symptom constructs to be rated on 7-point scale PANSS – interview (45-50 mins), clinician rated	Clinician-rated, and involve an interview.
EMA research guideline 2012(20)	Adults, paediatrics, adolescents	Expert panel and stakeholder consultation: Efficacy working party of committee for medicinal products for human use (CHMP)	Unclear	Unclear	PANSS and BPRS considered to be reliable and validated. Other recent measures acceptable if referenced to PANSS and BPRS. Secondary outcomes should be recorded, and use CGI	All measures are clinician-rated and involve an interview

BPRS, brief psychiatric rating scale; PANSS, positive and negative syndrome scale; KGV, Krawiecka, Goldberg and Vaughan Scale; SSPI, signs and symptoms of psychotic illness; CGI, clinical global impression

Reference List

(1) Konig HH, Gunther OH, Angermeyer MC, Roick C. Utility assessment in patients with mental disorders: validity and discriminative ability of the time trade-off method. Pharmacoeconomics 2009;27(5):405-19.

(2) Punekar YS, Sunderland T, Hawkins N, Lindsay J. Cost-effectiveness of scheduled maintenance treatment with infliximab for pediatric Crohn's disease. Value Health 2010 Mar;13(2):188-95.

(3) Pitkanen A, Valimaki M, Endicott J, Katajisto J, Luukkaala T, Koivunen M, et al. Assessing quality of life in patients with schizophrenia in an acute psychiatric setting: reliability, validity and feasibility of the EQ-5D and the Q-LES-Q. Nord J Psychiatry 2012 Feb;66(1):19-25.

(4) 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 Sep;14(6):907-20.

(5) Barton GR, Hodgekins J, Mugford M, Jones PB, Croudace T, Fowler D. Measuring the benefits of treatment for psychosis: validity and responsiveness of the EQ-5D. Br J Psychiatry 2009 Aug;195(2):170-7.

(6) Stochl J, Croudace T, Perez J, Birchwood M, Lester H, Marshall M, et al. Usefulness of EQ-5D for evaluation of health-related quality of life in young adults with first-episode psychosis. Qual Life Res 2013 Jun;22(5):1055-63.

(7) Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. USA: Oxford University Press; 2003.

(8) Walters SJ. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practice Guide to analysis and interpretation. Wiley Blackwell; 2009.

(9) Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.

(10) Fitzpatrick R, Gibbons E, Mackintosh A. An overview of patient-reported outcome measures for people with anxiety and depression. Report to the department of health, 2009. Patient-reported Outcome Measurement Group 2009.

(11) Konig HH, Roick C, Angermeyer MC. Validity of the EQ-5D in assessing and valuing health status in patients with schizophrenic, schizotypal or delusional disorders. Eur Psychiatry 2007 Apr;22(3):177-87.

(12) McCrone P, Patel A, Knapp M, Schene A, Koeter M, Amaddeo F, et al. A comparison of SF-6D and EQ-5D utility scores in a study of patients with schizophrenia. J Ment Health Policy Econ 2009 Mar;12(1):27-31.

(13) Van de Willige G, Wiersma D, Nienhuis FJ, Jenner JA. Changes in quality of life in chronic psychiatric patients: A comparison between EuroQol (EQ-5D) and WHOQoL. Qual Life Res 2005;14:441-51.

(14) Prieto L, Sacristan JA, Hormaechea JA, Casado A, Badia X, Gomez JC. Psychometric validation of a generic health-related quality of life measure (EQ-5D) in a sample of schizophrenic patients. Curr Med Res Opin 2004 Jun;20(6):827-35.

(15) Scalone L, Pirfo E, Mencacci C, et al. Relationship between clinical outcomes and patients' reported outcomes in schizophrenia: the contribution of the EQ-5D. Value Health 2008;11:A592.

(16) Badia X, Casado A, Sacristan JA, et al. Antipsychotic treatment and changes in health related quality of life in patients with schizophrenia using EQ-5D. 1999.

(17) 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 Sep;14(6):907-20.

(18) Royal College of Psychiatrists. Outcomes Measures Recommended for use in Adult Psychiatry. 2011. Report No.: OP78.

(19) Royal College of Psychiatrists Faculty of the Psychiatry of Old Age. Individual Patient Outcome Measures Recommended for Use in Older People's Mental Health. 2012. Report No.: OP86.

(20) European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on clinical investigation of medicinal products, including depot preparations in the treatment of schizophrenia. 2012.

(21) Department of Health. Mental Health Payment by Results Guidance for 2013-14. Department of Health; 2014.

(22) Department of Health. Closing the Gap: Prioritis for essential change in mental health. Department of Health; 2014.

(23) Department of Health. The NHS Outcomes Framework 2014/15. Department of Health; 2013.

(24) National Institute for Health and Care Excellence. Aripiprazole (Abilify®) for the treatment of schizophrenia in adolescents (15-17 years) TA213. 2010.

(25) Royal College of Psychiatrists. National audit of schizophrenia. Royal College of Psychiatry 2014Available from: URL: www.rcpsych.ac.uk/quality/nas

(26) Fitzsimmons D, Gilbert J, Howse F, Young T, Arrarras JI, Br+®dart A, et al. A systematic review of the use and validation of health-related quality of life instruments in older cancer patients. Eur J Cancer 2009;45(1):19-32.

(27) Konig HH, Born A, Heider D, Matschinger H, Heinrich S, Riedel-Heller SG, et al. Cost-effectiveness of a primary care model for anxiety disorders. Br J Psychiatry 2009 Oct;195(4):308-17.