This is a repository copy of *Comparing Generic and Condition-Specific Preference-Based Measures in Epilepsy: EQ-5D-3L and NEWQOL-6D*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/100623/

Version: Accepted Version

Article:
Mulhern, B., Pink, J., Rowen, D. orcid.org/0000-0003-3018-5109 et al. (5 more authors) (2017) Comparing Generic and Condition-Specific Preference-Based Measures in Epilepsy: EQ-5D-3L and NEWQOL-6D. Value in Health, 20 (4). pp. 687-693. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2016.03.1860

Article available under the terms of the CC-BY-NC-ND licence
(https://creativecommons.org/licenses/by-nc-nd/4.0/)

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. 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.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Comparing generic and condition specific preference based measures in epilepsy: EQ-5D-3L and NEWQOL-6D

Brendan Mulhern (MRes)[1,2], Joshua Pink (PhD)[3], Donna Rowen (PhD)[2], Simon Borghs (MSc)[4], Thomas Butt (PhD)[5], Dyfrig Hughes (PhD)[6], Tony Marson (MD)[7], John Brazier (PhD)[2]

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

2 School of Health and Related Research, University of Sheffield, Sheffield, UK

3 Warwick Medical School, University of Warwick, UK

4 UCB Pharma, Slough, UK

5 UCB Biopharma, Brussels, Belgium

6 Centre for Health Economics & Medicines Evaluation, Bangor University, UK

7 Department of Molecular and Clinical Pharmacology, University of Liverpool, UK

Corresponding author: Brendan Mulhern, Centre for Health Economics Research and Evaluation, University of Technology Sydney, 1-59 Quay St, NSW 2000; Australia.

e-mail: Brendan.mulhern@chere.uts.edu.au

Tel no: +61 2 9514 4725

Fax no: +61 2 9514 4730

Funding: This study was funded by Epilepsy UK and UCB Biopharma

Key words: EQ-5D, NEWQOL-6D, preference based measures, psychometrics, epilepsy

Running head: Psychometrics of utility measures in epilepsy

Abstract

Background: There is debate about the psychometric characteristics of the EQ-5D-3L for use in epilepsy. In response to these concerns, an epilepsy-specific preference-based measure (NEWQOL-6D) was developed. However the psychometric characteristics of NEWQOL-6D have not been assessed. The aim of this study was to investigate the validity and responsiveness of EQ-5D-3L and NEWQOL-6D for use in the assessment of treatments for newly diagnosed focal epilepsy.

Methods: The analysis used data from the Standard And New Antiepileptic Drugs (SANAD) trial including patients with focal epilepsy. We assessed convergent validity using correlations, and known group validity across different epilepsy and general health severity indicators using ANOVA and effect sizes. The responsiveness of the measures to change over time was assessed using standard response means (SRM). We also assessed agreement between the measures.

Results: There was some level of convergence and agreement between the measures in terms of utility score, but divergence in the concepts measured by the descriptive systems. Both instruments displayed known group validity, with significant differences between severity groups, and generally slightly larger effect sizes for NEWQOL-6D across the epilepsy specific indicators. Evidence for responsiveness was less clear, with small to moderate SRMs demonstrating different levels of change across different indicators.

Discussion: There was an overall tendency for the NEWQOL-6D to better reflect differences across groups, but this does not translate into large absolute utility differences. Both the EQ-5D-3L and NEWQOL-6D show some evidence of validity for providing utility values for economic evaluations in newly diagnosed focal epilepsy.

Introduction

In the economic evaluation of health technologies, the Quality-Adjusted Life-Year (QALY) is often the preferred measure of health outcome. The QALY combines a period of time spent in a health state and the associated quality of life into a single figure to assess the overall effectiveness of treatments. The quality aspect of the QALY is known as health utility, and can be derived from preference-based measures of health such as the EQ-5D-3L, which has a utility value set based on the preferences of the general population, and is anchored at 1 (equivalent to full health) and 0 (equivalent to dead).

Epilepsy is a neurological disorder characterised by unprovoked, recurring seizures. There are different types of seizures, with focal (partial onset) seizures originating within specific areas on the brain [1]. It is a prevalent condition, and is present in around 1% of the world's population [2]. Patients newly diagnosed with epilepsy are usually initiated on antiepileptic drug treatment, with dose and subsequent regimens adapted according to treatment response and tolerability. For those whose seizures are difficult to control, multiple concurrent therapies can be used to maximise seizure control [3,4]. Given the nature of the condition, epilepsy and antiepileptic drug treatment both have a range of impacts on Health Related Quality of Life (HRQL) including on mental health (for example anxiety about the onset of seizures), and cognitive and social functioning.

For the economic evaluation of treatments for epilepsy, it is important to measure the HRQL (and utilities) of the patient groups receiving treatment, and the EQ-5D-3L is often used for this purpose. However, although EQ-5D-3L is a generic measure, and is used to compare health status across conditions, there is evidence suggesting that it may lack some level of validity for use in epilepsy. Stavem and colleagues [5] found that a number of the EQ-5D-3L dimensions did not discriminate well between patients using antiepileptic drugs and patients with neurologic comorbidities. Furthermore, the EQ-5D-3L may not capture all of the HRQL dimensions relevant to epilepsy patients [6], and may not be sensitive to seizure control [7]. In terms of responsiveness, EQ-5D-3L may be insensitive to health status at the upper end due to a ceiling effect (with many patients reporting the 'best' health state, with no problems, on the measure). Furthermore, the recall period of one day may impact the sensitivity of the instrument when used for episodic conditions such as epilepsy [8].

In response to these findings, an epilepsy-specific preference based measure was developed for use in the assessment of HRQL (NEWQOL-6D) [9]. NEWQOL-6D was developed using baseline data from the Standard and New Anti-Epileptic Drugs (SANAD) study [10]. SANAD included patients with predominantly newly developed focal epilepsy for whom one treatment (carbamazepine) was considered the standard treatment, and who were randomly assigned to receive treatment as usual or one of four other treatments (gabapentin, lamotrigine, oxcarbazepine, or topiramate). Condition specific preference based measures may improve health care decision making in areas where the psychometric validity of generic measures is lacking by providing more precise utility values that are based on the dimensions of HRQL impacted by the condition. However it is currently unclear whether NEWQOL-6D provides an improvement in the validity of utility values generated for the calculation of QALYs, and therefore psychometric evidence comparing generic and condition specific measures is required.

The aim of this study was to investigate the psychometric performance of the EQ-5D-3L and NEWQOL-6D to provide evidence for the use of the utility values generated in the assessment of epilepsy treatments in newly diagnosed focal epilepsy patients using an existing dataset. This includes the following objectives:

1. To assess the psychometric validity of the EQ-5D-3L and NEWQOL-6D
2. To assess the responsiveness (i.e. change in utility) of EQ-5D-3L and NEWQOL-6D to clinical outcomes in newly diagnosed focal epilepsy.

Methods

Measures - EQ-5D-3L

The EQ-5D-3L [11] is the most widely used generic preference based measure internationally. It is recommended by the National Institute for Health and Care Excellence (NICE) in the UK for use in the economic evaluation of interventions [12], and is accepted by other reimbursement agencies around the world [13,14]. EQ-5D-3L measures health across five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) each with three response levels (none, some and extreme/unable), and therefore describes 243 (3^5) health states. A selection of health states were

valued by a representative sample of the UK general population using the Time Trade-off (TTO) preference elicitation technique [15]. This produced a utility scale with a range from 1 for the best state, 11111, to -0.594 for the worst state, 33333, where negative values are equivalent to states valued as worse than dead [15]. The visual analogue scale (EQ-VAS, which reports health on a 0 (worst imaginable) to 100 (best imaginable) scale) was also collected and used to provide an indication of the overall health of the sample.

Measures - NEWQOL-6D

The NEWQOL-6D [9] was developed from the NEWQOL instrument [16] using the baseline data from a large randomised controlled trial of anti-epileptic drugs, (the SANAD study, see dataset section). NEWQOL-6D assesses epilepsy specific HRQL across six dimensions (worry about attacks (seizures); depression; memory; concentration; control; stigma) each with four severity response levels, therefore describing 4096 (4^6) possible health states. To produce the utility scale, 50 health states were valued by a representative sample of the UK general population using TTO (to promote comparability with the UK EQ-5D-3L tariff [15]). This produced a value set with a range from 0.954 (for the best state 111111) to 0.341 for the worst state, 444444,. Further work with the instrument has shown that general population and patient valuations of NEWQOL-6D health states have limited differences [17].

Dataset

The SANAD data [10] from three time points (baseline, year 1 and year 2) was used for the analysis,. The data for each time point includes the EQ-5D-3L, the NEWQOL battery [16] which allows the NEWQOL-6D to be calculated, and a range of clinical and other health indicators, including the number of seizures, self-reported health change over the last year, cognitive problems, healthcare services received. The Hospital Anxiety and Depression Scale (HADS) [18], a widely used measure of anxiety and depression that provides cut offs for possible and probable caseness, was also used as a comparative measure. Patients with complete data for each of the measures across each time point were included.

Analysis

A range of psychometric analyses were carried out to assess the validity and responsiveness of the EQ-5D-3L and NEWQOL-6D.

Descriptive analysis

The means, standard deviations, medians, ranges and completion rates of the utility scores for each of the measures at each time point were assessed.

Validity

Psychometric analysis of validity assesses the extent to which an instrument measures what it is intended to measure (in this case HRQL). It should be noted that validity of patient-reported outcome measures such as EQ-5D-3L and NEWQOL-6D is difficult to prove, as in many health conditions there is rarely a 'gold standard' against which to compare. This is the case in epilepsy, where many of the measures lack in psychometric evidence, and means that validity is compared across measures using various well established tests and guidelines about the magnitude of the relationship.

Concurrent validity

Concurrent validity assesses the strength of the relationship between measures of the same concept using Pearson correlations. Strong correlations indicate that the preference-based measures are assessing related constructs. Correlations are considered weak if scores are <0.3, moderate if scores are ≥0.3 and <0.7, and strong if scores are ≥0.7 [19]. Correlations between the EQ-5D-3L and NEWQOL-6D utility scores and dimensions at each time point (baseline, year 1, year 2) were assessed.

Agreement between the utility values was assessed using Bland Altman plots, which are used to visualise the relationship between measures scored on the same scale. This was done by plotting the mean of the EQ-5D-3L and NEWQOL-6D values against the difference (NEWQOL-6D minus EQ-5D-3L) and assessing the number of plots outside the agreement range (+/- 2 SD from the mean) [20].

Known group validity

Known group validity assesses the extent to which scores on an instrument differ across groups where they are expected to differ (for example severity groups). This was measured using ANOVA difference tests, and by calculating effect sizes between the groups, which provide a standard indicator of the size of the difference, and are calculated as Score 2 – Score 1/Standard deviation of Score 1. Effect sizes of less than 0.2 are considered small, 0.5 moderate, and 0.8 large [19]. We assessed differences in EQ-5D-3L and NEWQOL-6D utility scores at baseline across a range of groups:

a. Total number of seizures at baseline (defined as four groups: 1-3, 4-5, 6-9 and 10+ total lifetime seizures).
b. Self-reported health status (defined as three groups from a five point Likert scale: very good/excellent, good, and fair/poor)
c. Remission status (two groups: remission and no remission, where remission was defined as no attack in the past year). This was carried out at year 1 and year 2, and can be interpreted as known group validity because the utility of those in different frequency groups can be assessed at one time point.
d. HADS anxiety and depression dimension caseness cut offs at baseline (defined as no case (a score below 8) and possible case (a score of 8+)).

Responsiveness

Responsiveness assesses the ability of an instrument to detect changes in HRQL over time. However, responsiveness is difficult to prove as there is no 'true' reference measure of HRQL. Therefore clinical variables that are expected to have an effect on HRQL are used to identify groups of patients whose HRQL is expected to have changed, and the responsiveness of the utility measures is compared across the groups. We assessed responsiveness for full completers at each time point, and compared baseline to year 1, and baseline to year 2, by assessing mean change in utility overall, and across identified change groups using the Standardised Response Mean (SRM; (T2 – T1)/SD of change), which is a standard indicator of change across measures and time points. SRMs of less than 0.2 are considered small, 0.5 moderate, and 0.8 large [19]. The following analyses were carried out:

a. Assessment of floor and ceiling effects: If a large proportion are at the floor (lowest possible score) or ceiling (highest possible score) then this impairs the ability of the measure to pick up decreases or increases in HRQL respectively.
b. Mean change and SRMs based on self-reported health transition anchor (assessing whether health has improved, stayed the same or worsened over the last year).
c. Mean change and SRMs based on remission, defined as those with no seizures during the past year. This was done between Y1 and Y2 as it is difficult to interpret baseline frequency in terms of when the attacks occurred outside the SANAD study period. although the sample consists of newly diagnosed patients, the exact time of seizure occurrence is unclear.
d. Change in the actual NEWQOL-6D and EQ-5D-3L health states reported over time

Results:

Sample characteristics

The characteristics of the sample at baseline are reported in Table 1, with the sample size at years 1 and 2 also indicated. A slight majority of the sample (55%) were male and the mean age was 40. The mean EQ-VAS score was 68, with a wide range of health reported (from 5 to 100).

Descriptive analysis

The completion rate of EQ-5D-3L was higher than NEWQOL-6D. Descriptive statistics for EQ-5D-3L and NEWQOL-6D are reported in Table 2. The mean EQ-5D-3L scores are lower than the NEWQOL-6D across each time point, with a larger standard deviation. The median of the EQ-5D-3L is higher in each case, reflecting the skewed nature of the data. The EQ-5D-3L data is bimodal, with a large number of responses at 1, whereas NEWQOL-6D is unimodal. Bimodal EQ-5D-3L data is common due to the difference in utility values between the best state (11111 with a utility of 1), and the next best state (11211 with a utility of 0.883).

Concurrent validity

Correlations indicate moderate convergence between the EQ-5D-3L and NEWQOL-6D utility scores at baseline (0.617), year one (0.651) and year two (0.647). However, the correlations between the

NEWQOL-6D and EQ-5D-3L dimensions at baseline reported in Table 3 indicate a low level of overlap except for the depression dimension. only baseline correlations are reported as the correlation pattern is similar at both follow up time points. Figure 1 displays the Bland Altman plot at baseline, with the central line indicating the mean, and the lines either side indicating 2 x SD away from the mean, and shows that disagreement between the measures is larger at the more severe end of the scale, where the mean of the measures is low, and many data points are outside the limits of agreement. The relationship is similar at both follow up time points.

Known group validity

Table 4 demonstrates that both EQ-5D-3L and NEWQOL-6D significantly discriminate between groups defined by number of seizures (EQ-5D-3L: ($F_{(2,1543)}$ = 28.02, p = 0.000); NEWQOL-6D: ($F_{(2,1496)}$ = 32.50, p = 0.000)), with a small to moderate effect size. For self-reported health the differences are significant (EQ-5D-3L: ($F_{(2,1557)}$ = 333.03, p = 0.000); NEWQOL-6D: ($F_{(2,1505)}$ = 212.72, p = 0.000)) with a moderate to large effect size. A large effect size is seen for both measures between groups defined by impact of attack score (EQ-5D-3L: ($F_{(1,808)}$ = 127.33, p = 0.000); NEWQOL-6D: ($F_{(1,793)}$ = 174.07, p = 0.000)), and HADS-A (EQ-5D-3L: ($F_{(1,1526)}$ = 418.51, p = 0.000); NEWQOL-6D: ($F_{(1,1483)}$ = 670.57, p = 0.000) and HADS-D (EQ-5D-3L ($F_{(1,1538)}$ = 625.82, p = 0.000); NEWQOL-6D ($F_{(1,1493)}$ = 792.24, p = 0.000) cut off scores. A moderate to large effect size for remission status was observed for both EQ-5D-3L and NEWQOL-6D.

Responsiveness

EQ-5D-3L has a large ceiling effect at each time point, with between 31.6% reporting the 'best' health state (11111) at baseline. This may limit the ability of the descriptive system to measure increases in health, and is a common finding with EQ-5D-3L (Brazier et al., 2004). NEWQOL-6D does not have floor or ceiling effects. Table 5 reports the responsiveness of EQ-5D-3L and NEWQOL-6D. NEWQOL-6D has larger a SRM than EQ-5D-3L for those who report remission in number of seizures and also improved health over the last year, with utility values improving . EQ-5D-3L is slightly more responsive in those that reported worsening health, although the SRMs are very similar. Overall the SRMs are generally in the low to moderate range overall, with the largest SRMs reported between baseline and year one. The frequency of respondents changing response category and overall utility

value varies for each of the preference based measure. NEWQOL-6D demonstrates a higher level of change than EQ-5D-3L between baseline and year 1 at the utility (85% vs. 62%) and response category level. At the response level, the number of patients changing category ranges from 33% (stigma) to 60% (worry) change for NEWQOL-6D, and 12% (mobility) to 35% (anxiety/depression) for the EQ-5D-3L.

Discussion

There is some evidence from earlier research that EQ-5D-3L may not demonstrate a high level of validity for use in the measurement of HRQL in epilepsy [5-7]. We have carried out psychometric analysis to compare the recently developed epilepsy specific preference based measure NEWQOL-6D and the EQ-5D-3L in a sample of patients with newly diagnosed focal epilepsy. This adds to the past work investigating the psychometric validity of the EQ-5D-3L, as well as establishing evidence about the validity of the NEWQOL-6D.

The results suggest that there are similarities and differences between the measures, which impact on the psychometric characteristics, and are likely to affect subsequent cost effectiveness analyses. Firstly, the mean EQ-5D-3L scores are lower, and this is commonly found when comparing EQ-5D-3L to other generic and condition specific preference based measures due to comparative differences in the range of the utility scales and the concepts measured by the descriptive systems [21-22]. Secondly, in terms of convergence, the utility scales are moderately correlated, but there are low correlations between the descriptive systems. This has been shown elsewhere [21-23] and suggests that these instruments are measuring some overlapping concepts, but there some important differences, and the utility scales, developed using the same method, are complementary.

In terms of other differences between the descriptive systems, EQ-5D-3L has a large ceiling effect, a common finding [24] which suggests that EQ-5D-3L is not sensitive to small differences in health (i.e. the difference between none and some problems is too large to pick up small changes), but also cannot pick up improvements in health at follow up amongst those who report the best EQ-5D-3L health state at baseline. NEWQOL-6D does not display a ceiling effect, and has a larger number of respondents changing the reported health state between time points, in part due to the number of

health states described (4096 vs. 243). This may mean that NEWQOL-6D could be more sensitive in less severe epilepsy groups who are likely to be at the upper end of the utility scale. It should be noted that the five level EQ-5D (EQ-5D-5L) [25] may help resolve the ceiling effect issue to some extent, and increases the number of health states to 3125. EQ-5D-5L should now be considered for inclusion in trials where cost effectiveness analysis is required. EQ-5D-3L has a higher overall response rate, as it is administered as a five item questionnaire whereas NEWQOL-6D is derived from a much longer battery measure. Furthermore, the EQ-5D-3L appeared before the instruments included in the NEWQOL battery [16] in the SANAD study questionnaire, and therefore the later measures may be affected by questionnaire fatigue. However both have response rates over 90%. Finally, the recall period for EQ-5D relates to 'today', and this may limit its sensitivity in a chronic, episodic disease such as epilepsy.

In terms of other psychometric indicators, the EQ-5D-3L and NEWQOL-6D perform similarly overall which suggests that both measures have validity for use in the economic evaluation of treatments for newly diagnosed focal epilepsy. Both display generally good discriminant validity between clinical and general health severity groups, with the same category effect size. Therefore both measures can be used with confidence to distinguish severity groups. Agreement between the measures is better at the less severe end of the utility scale, which is a common finding given differences in the range reported by the measures, with EQ-5D-3L having a larger range than most utility instruments [26]. The evidence for responsiveness is less clear. There is an indication that NEWQOL-6D has slightly better responsiveness amongst those who report general health and seizure frequency improvement, but the measures are more similar when assessing those who report worsening health between time points. Further research using clinical rather than self-reported indicators of change would be useful.

The general determinants of HRQL in epilepsy, which is determined not just by efficacy but also by comorbidities, treatment safety and baseline severity of the condition [27] may provide context for some of the results reported. For example, in terms of the descriptive system, both measures are highly sensitive to anxiety and depression which are well established epilepsy comorbidities that are related to HRQL regardless of seizure frequency [28]. However, anxiety and depression are not correlated entirely with treatment efficacy in epilepsy so they don't effect much change in overall utility

in studies of antiepileptic treatment. It is therefore difficult to delineate the impact of HRQL in epilepsy on the sensitivity of and change in utility values measured (rather than at the dimension level) given the overall similar findings.

The analysis reported here raises wider issues about the use and acceptability of generic and condition specific preference based measures. For example, EQ-5D-3L allows for comparisons of utilities across health conditions which in some cases are limited by the psychometric validity of the instrument in that area. By definition, condition specific measures do not allow for comparisons across health areas [29,30], but may provide a set of complementary utility values can be assessed alongside, for example, EQ-5D, to allow for more holistic measurement of the HRQL impacts of the condition. This might be particularly useful in settings where a particular measure is recommended, but evidence relating to that measure in certain health conditions is mixed (for example NICE recommending the EQ-5D in England) [12]. Furthermore, condition specific measures such as NEWQOL-6D may be limited in populations where there are comorbidities (as it is not designed to pick up changes in those). Further psychometric comparisons of generic and condition specific preference based measures across a range of health areas is warranted.

The differences in the psychometric indicators may reflect in some ways into differences in cost effectiveness results, where the incremental cost-effectiveness ratio using NEWQOL-6D values may be different when compared with EQ-5D-3L. Further work could investigate this issue in more detail across different trial datasets.

There are limitations to this analysis. Data were available from only one source (the SANAD trial that was also used to develop NEWQOL-6D) and replication of the results on a different sample with different external clinical indicators and types of epilepsy would be of interest. The time period studied in SANAD, where the measures were taken once a year over a two-year observation period is quite different to the typical RCT duration used for drug registration purposes. We also did not have access to data about the type of seizure experienced by those included in SANAD, and this may have helped inform the utility values used. The baseline seizure frequency data is difficult to interpret, as a patient with 2 seizures in the 12 months prior to starting treatment may have had them in the space of a few

weeks before starting treatment, or over a longer period. The analysis could also not fully assess the responsiveness of the measures by varying levels of response to AED treatment as we could only define patients as seizure-free, or having one or more seizures. Finally, using samples based on those fully completing each of the measures also excludes a number of respondents, and potentially those in poorer health who may have not managed to complete the long NEWQOL battery, but is a valid criterion given that we are trying to compare the performance of measures across matched samples.

In conclusion, there is the indication that NEWQOL may better reflect differences across groups particularly at the least severe end but both measures perform at a similar level. . Therefore there is some evidence that both the EQ-5D-3L and NEWQOL-6D have validity for providing utility values for use in the economic evaluation of interventions for newly diagnosed focal epilepsy, and subsequent decision making. Given the widespread use of EQ-5D-3L (and its use across conditions given its generic nature), and the subsequent development of the EQ-5D-5L which may increase responsiveness, both measures could be used alongside each other to provide complementary utility values from a generic and condition specific perspective, and more holistic measurement of HRQL in epilepsy.

References

1 Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. Epilepsia. Apr 2010;51(4):676-685.

2 Thurman DJ, Beghi E, Begley CE et al. "Standards for epidemiologic studies and surveillance of epilepsy." Epilepsia 2011;52(S7):2–26.

3 Shorvon SD, Reynolds EH. Unnecessary polypharmacy for epilepsy. BMJ 1977;1:1635–37.

4 Shorvon SD, Reynolds EH. Reduction of polypharmacy for epilepsy. BMJ 1979;2:1023–25

5 Stavem K, Bjornaes H, Lossius MI. Properties of the 15D and EQ-5D-3L utility measures in a community sample of people with epilepsy. Epilepsy Res 2001;44:179-89.

6 Selai CE, Trimble MR, Price MJ, et al. Evaluation of health status in epilepsy using the EQ-5D-3L questionnaire: a prospective, observational, 6-month study of adjunctive therapy with anti-epileptic drugs. Curr Med Res Opin 2005;21(5):733-9.

7 Langfitt JT, Vickrey, BG, McDermott MP. Validity and responsiveness of generic preference-based HRQOL instruments in chronic epilepsy. Qual Life Res 2006;15:899–914.

8 Böckerman P, Johansson E, Saarni SI. Do Established Health-related Quality-of-life Measures Adequately Capture the Impact of Chronic Conditions on Subjective Well-being? Health Pol 2011;100:91-5.

9 Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Latimer N et al. The development of a QALY measure for epilepsy: NEWQOL-6D. Epilepsy and Behaviour 2012;24:36-43.

10 Marson AG, Al-Kharusi AM, Alwaidh M, et al. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazeprine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. Lancet 2007; 369: 1000-1015.

11 Brooks R. EuroQol: The current state of play. Health Policy 1996;37:53-72.

12 National Institute for Health and Care Excellence. Guides to the methods for technology appraisal. NICE 2013.

13 Pharmaceutical Benefits Advisory Committee. Guidelines to preparing submissions to the Pharmaceutical Benefits Advisory Committee. Australian Department of Health 2013.

14 Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. CADTH 2006.

15 Dolan P. Modeling valuations for EuroQol health states. Medical Care 1997; 35(11):1095-108.
16 Abetz L, Jacoby A, Baker GA, McNulty P. Patient-based assessments of quality of life in newly diagnosed epilepsy patients: validation of the NEWQOL. Epilepsia 2000; 41(9): 1119-28.
17 Mulhern B, Rowen D, Snape A, et al. Valuations of epilepsy-specific health states: a comparison of patients with epilepsy and the general population. Epilepsy and Behavior 2013;36:12-7.
18 Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983;67(6):361-70.
19 Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum 1988.
20 Bland MJ, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet . 1986;327(8476):307-10.
21 Rowen D, Young T, Brazier J, Gaugris S. Comparison of Generic, Condition-Specific, and Mapped Health State Utility Values for Multiple Myeloma Cancer. Value Health 2012;15(8):1059-68.
22 Mulhern B, Rowen D, Brazier J, et al. Development of DEMQOL-U and DEMQOL-Proxy-U: Generation of preference based indices from DEMQOL and DEMQOL-Proxy for use in economic evaluation. Health Technology Assessment 2013;17(5).
23 Mulhern B, Mukuria C, Barkham M, et al. Using preference based measures in mental health conditions: The psychometric validity of the EQ-5D-3L and SF-6D. Brit J Psychiatry. 10.1192/bjp.bp.112.122283.
24 Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D-3L and SF-6D across seven patient groups. Health Econ 2004;13(9):873-84.
25 Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D-3L (EQ-5D-3L-5L). Qual Life Res 2011;20(10):1727-36.
26 Whitehurst D, Bryan S. Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D-3L and SF-6D) are Not Interchangeable. But Why Should We Expect Them to be? Value Health 2011;14(4):531–8.
27 Taylor R, Sander J, Taylor R, Baker G. Predictors of health-related quality of life and costs in adults with epilepsy: A systematic review. Epilepsia, 2011;52(12):2168–80.
28 Kwan P, Yu E, Leung H, Leon T, Mychaskiw M. Association of subjective anxiety, depression, and sleep disturbance with quality-of-life ratings in adults with epilepsy. Epilepsia 2009;50(5):1059–66.

29 Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, Barkham M, Ibbotson R. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technology Assessment 2012;16:32.

30 Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Econ 2010;19(2):125-29.

Table 1: Background characteristics (baseline)

		N (%)
N	Baseline	1611
	Year 1	1288
	Year 2	1134
Male		885 (55.2)
Age		
	Mean (SD)	39.7 (16.5)
	Range	16 – 86
In employment		1297 (78.8)
Marital status		
	Married/partner	885 (55.1)
	Single	548 (34.1)
	Divorced/separated	134 (8.3)
	Widowed	38 (2.4)
EQ-5D-3L VAS		
	Mean (SD)	68.14 (20.7)
	Range	5 – 100

Table 2: Descriptive analysis of EQ-5D-3L and NEWQOL-6D utility scores

	Baseline	One year	Two years
EQ-5D-3L			
N fully completing (% completion)	1563 (98)	1244 (98)	1091 (98)
Mean	0.735 (0.30)	0.769 (0.29)	0.789 (0.28)
Median	0.848	0.848	0.848
Range	-0.38 to 1	-0.454 to 1	-0.239 to 1
NEWQOL-6D			
N fully completing (% completion)	1508 (94)	1156 (90)	1023 (90)
Mean	0.766 (0.13)	0.798 (0.13)	0.805 (0.13)
Median	0.786	0.832	0.844
Range	0.341 to 0.957	0.341 to 0.957	0.341 to 0.957

Table 3: EQ-5D-3L and NEWQOL-6D dimension level correlations (baseline)

	EQ-5D-3L				
	Mobility	Self-Care	Usual activities	Pain/discomfort	Anxiety/depression
NEWQOL-6D					
Worry	0.221	0.169	0.326	0.243	0.327
Depression	0.235	0.217	0.355	0.303	0.633
Memory	0.267	0.207	0.346	0.271	0.319
Concentration	0.289	0.237	0.388	0.265	0.326
Control	0.307	0.255	0.387	0.288	0.357
Stigma	0.243	0.213	0.305	0.230	0.261

Table 4: Known group validity of EQ-5D-3L and NEWQOL-6D utilities

	EQ-5D-3L				NEWQOL-6D			
	N	Mean (SD)	ES	Sig	N	Mean (SD)	ES	Sig
No of seizures last year (BL)				0.000				0.000
1-3	474	0.811 (0.26)			468	0.802 (0.11)		
4-9	447	0.737 (0.29)	0.28		432	0.765 (0.12)	0.34	
10+	625	0.676 (0.33)	0.21		599	0.738 (0.14)	0.23	
Self reported health								
Very good/excellent	545	0.893 (0.18)		0.000	515	0.830 (0.10)		0.000
Good	581	0.787 (0.22)	0.59		554	0.776 (0.11)	0.54	
Fair/poor	482	0.489 (0.35)	1.35		439	0.678 (0.13)	0.89	
Remission status (Y1)								
No remission (>1 attack)	742	0.701 (0.32)		0.000	695	0.762 (0.13)		0.000
Remission (0 attacks)	495	0.876 (0.20)	0.88		456	0.855 (0.10)	0.93	
Remission status (Y2)								
No remission (>1 attack)	513	0.700 (0.31)		0.000	479	0.756 (0.14)		0.000
Remission (0 attacks)	573	0.869 (0.22)	0.77		540	0.849 (0.09)	1.03	
HADS-A cut off								
No case	743	0.880 (0.19)		0.000	718	0.840 (0.08)		0.000
Possible case	785	0.599 (0.32)	1.48		767	0.696 (0.13)	1.80	
HADS-D cut off								
No case	1036	0.848 (0.20)		0.000	999	0.820 (0.09)		0.000
Possible case	504	0.501 (0.34)	1.02		496	0.657 (0.13)	1.81	

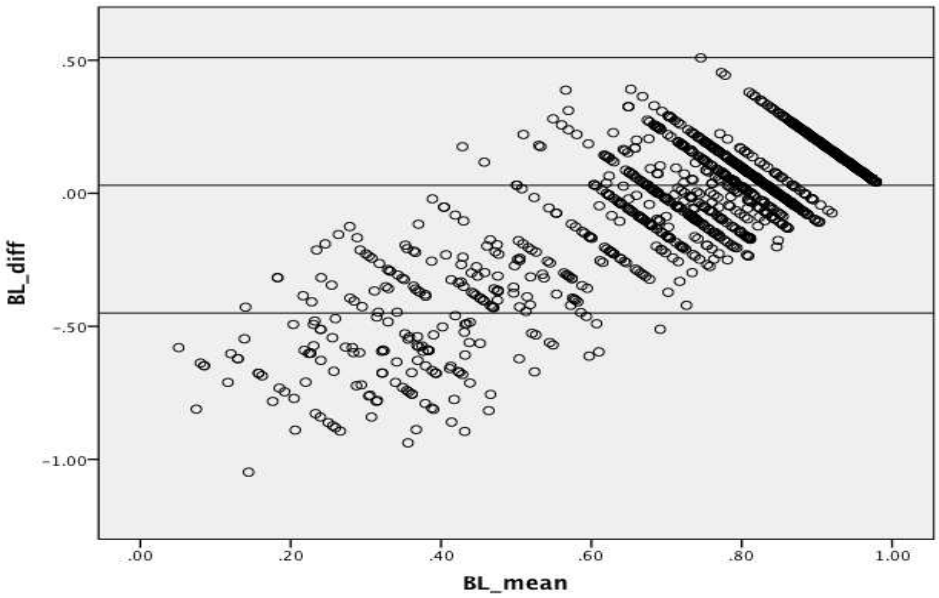
ES: Effect size

Table 5: Responsiveness of EQ-5D-3L and NEWQOL-6D

	T0-T1		T1-T2	
	Mean Utility Change	SRM	Mean Utility Change	SRM
EQ-5D-3L				
Overall	0.025	0.10	0.005	0.03
Self-report health change anchor				
Improved	0.076	0.38	0.051	0.30
Same	0.019	0.09	0.002	0.01
Declined	-0.092	0.29	0.084	0.35
Seizure frequency				
Remission	0.045	0.23	0.031	0.18
NEWQOL-6D				
Overall	0.027	0.27	-0.002	0.02
Self-report health change anchor				
Improve	0.056	0.52	0.022	0.24
Same	0.018	0.20	-0.008	0.11
Decline	-0.022	0.19	-0.036	0.31
Seizure frequency				
Remission	0.047	0.53	0.019	0.23

SRM = standardised response mean

Figure 1: Bland Altman plot at baseline of EQ-5D-3L and NEWQOL-6D



Nb: The x axis is the mean utility score of the EQ-5D-3L and NEWQOL-6D, and the he y axis is the difference in utility score between the measures