Article:
Wickramasekera, N. and Tubeuf, S. (2023) Measuring quality of life for people with common mental health problems. Journal of Mental Health, 32 (1). pp. 3-10. ISSN: 0963-8237

https://doi.org/10.1080/09638237.2020.1818190

Measuring quality of life for people with common mental health problems

Nyantara Wickramasekera[*1], Sandy Tubeuf[2]

[1 *] N.Wickramasekera@sheffield.ac.uk – Health Economics and Decision Science, University of Sheffield, 30 Regent Street, Sheffield, S1 4DA, United Kingdom. Tel: +44(0)1142224349

[2] Institut de Recherche Santé et Société, Université Catholique de Louvain, Clos Chapelle-aux-Champs, 30, Bte B1. 30. 15 B-1200 Brussels

Acknowledgements: We are grateful to the ECO study team lead by Dr Helen Elsey. We thank the information specialist team – Judy Wright, Thomas Veale and Rocio Rodriguez Lopez – for the literature search. Also, we would like to thank Professor Allan House for reviewing a preliminary version of the paper.

Conflict of Interest: The authors declare that no conflict exists.

Author contributions: All co-authors designed and carried out the research, conducted the analyses and wrote the paper.

Funding: The study was funded by the NIHR public health programme (project number 11/3050/08)

Data Availability: Participant consent was taken only for the purpose of this study therefore the data set is not publically available.

Ethical approval: Ethical approval was granted by the University of Leeds School of Medicine Research Ethics Committee. The study also received approvals from the Ministry of Justice and the probation services involved (SoMREC/13/014 in December 2013 and NOMS approval in February 2014 (NOMS 2013-257)). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. Informed consent was obtained from all individual participants included in the study.

Abstract

Background: It is important to identify valid and acceptable outcome measures so that interventions evaluating common mental health problems can be assessed appropriately. Some advocate the use of generic preference-based measures claimed to be applicable for all health interventions, but others argue that they are insensitive for common mental health problems. The aim of this paper is to evaluate the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM), to be used in cost-effectiveness studies in people with common mental health problems.

Method: The CORE-OM measure was tested for completeness, acceptability and responsiveness in a pilot study. Analyses for missing data, distribution of scores, and standardised response means (SRMs) were calculated.

Results: Missing data did not exceed 5% for any of the CORE-6D items both at baseline and follow-up. The overall comprehension rate was high, and only 19 participants (14%) requested clarifications to complete the questionnaire. As expected in a feasibility study, there was a small and non-significant SRM.

Conclusion: CORE-OM is a valid and acceptable instrument to evaluate quality of life for people with common mental health problems. More research is needed with larger sample sizes to compare CORE-6D with other condition specific quality of life instruments.

Key words

Mental Health, Preference based measure, Utility, Feasibility, Acceptability

Measuring quality of life

Introduction

Generic quality of life measures, widely used to assess effectiveness in cost-effectiveness analyses, are problematic to use among people with mental health problems[1] (Brazier, 2010). There are several widely available quality of life measures used in mental health research (Burckhardt and Anderson, 2003; Gilbody, Richards, & Barkham, 2007; Gray and Mellor-Clark, 2007; Keetharuth et al., 2018; Skevington, Lotfy, & O'Connell, 2004; Thorgrimsen et al., 2003; van Nieuwenhuizen, Janssen-de Ruijter, & Nugter, 2017), however only a small sample of these can be used to generate utilities, which are key to measure QALYs (quality-adjusted life years), the preferred outcome for cost-effectiveness analyses in the UK (NICE, 2013). Generic measures such as EQ-5D (Herdman et al., 2011) and SF-36 (Brazier, Usherwood, Harper, & Thomas, 1998) can be used to generate QALYs but they have been found not robust enough to measure quality of life for people with common mental health problems because they overly focus on physical health and do not contain sufficient dimensions to capture aspects important to people with common mental health problems (Brazier, 2010; Finch, Brazier, & Mukuria, 2018; Keetharuth, et al., 2018; Papaioannou, Brazier, & Parry, 2011; Saarni et al., 2010). Evidence shows that they are not sensitive to capture symptom reduction in people with psychotic disorders and personality disorders (Barton et al., 2009; Papaioannou, et al., 2011; Saarni, et al., 2010). Despite this, EQ-5D has been adopted as the standard outcome measure for regulatory bodies such as the National Institute for Health and Care Excellence (NICE) (NICE, 2013). These concerns are raised because insensitive measures may underestimate the effectiveness of interventions (Saarni, et al., 2010), and thereby undermine health care decision making. Consequently, health

[1] Mental health conditions is not the only special case and generic health-related quality of life quality measures have been found not sensitive in other diseases areas (Longworth et al., 2014).

economists and decision makers require robust quality of life instruments that can be used in cost-effectiveness studies of interventions for people with common mental health problems. Given the limitations of generic measures, there is a growing interest in using condition specific mental-health instruments to capture changes in quality of life for people with common mental health problems. We identified two quality of life instruments specific to mental health, the Clinical Outcome in Routine Evaluation 6 Dimensions (CORE-6D) and the Recovering Quality of Life (ReQoL), which both enable the calculation of utilities and ultimately QALYs for cost-effectiveness analyses. The CORE-6D is based on six questions of the Clinical Outcome in Routine Evaluation-Outcome Measure, which is a 34-item questionnaire developed in the UK for routine use in psychological services (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005; Barkham et al., 2001; Barkham, Mellor-Clark, Connell, & Cahill, 2006; Muehlenkamp and Gutierrez, 2007) and it has been valued to generate utilities and QALYs. The ReQoL was recently developed especially for people experiencing mental health difficulties (Keetharuth, et al., 2018) and a valuation survey is currently underway to produce utility scores for ReQoL (Keetharuth et al., 2017). For the purpose of our study, we chose to survey CORE-OM instead of ReQoL because it is a well-established instrument, which has been used in other trials (Barkham, et al., 2005; Barkham, et al., 2001; Barkham, et al., 2006; Muehlenkamp and Gutierrez, 2007) and weights were already available to transform CORE-6D into utilities.

In the current study, we explored the appropriateness of using the CORE-OM instrument to evaluate a social farming intervention for people with common mental health problems. Social farming (or care farming) is a complex intervention including outdoor activities such as horticulture, livestock farming, gardening, conservation and woodwork. There is growing evidence that outdoor and nature-based interventions promote mental health and physical wellbeing (Bragg and Atkins, 2016; Clatworthy, Hinds, & Camic, 2013; Murray

et al., 2019). These benefits are strongly valued by social care commissioners and local authorities are often in charge of financing and developing outdoor interventions. Research in this area is needed to capture the benefits to the service users in this context and to conduct economic evaluations demonstrating the value for money. The aim of this study is to assess the feasibility, acceptability and the responsiveness of CORE-OM, and the related preference-based measure CORE-6D, as a mental health-related quality of life measure that has been used to evaluate a complex outdoor intervention for people with common mental health problems.

Methods

Study design

We used data from a feasibility study, the ECO study, which aimed to assess the feasibility of conducting a cost-effectiveness study using care farming as an intervention for probation service users experiencing common mental health problems and serving community orders (Elsey et al., 2018). The study was carried out in three sites in England and participants were included in the study if they were adult probationers (18 years or above) who were serving a community order as mandated by a judge. We selected the CORE-OM outcome measure to be tested in the ECO study because CORE-OM had previously been used to evaluate quality of life of adult probationers (Horton et al., 2014; Tapp, Fellowes, Wallis, Blud, & Moore, 2009) and the included participants were likely to have common mental health problems. Also, probationers are identified as a group of individuals with a higher proportion of mental health problems (Seymour, 2010). Ethical approval to carry out the pilot study was granted by the University of Leeds Medicine and Health ethics committee and written informed consent was obtained from all participants. The protocol of the main study is available elsewhere (Elsey et al., 2014).

Measuring quality of life

The participants in the intervention arm attended a care farm. It is a complex social intervention that provides farming activities for therapeutic purposes (Bragg, Egginton-Metters, Elsey, & Wood, 2014). The defining characteristics of a care farm involve both a "farming" and "care" component. The participants in the care farming arm looked after farm animals and did horticultural activities. The participants for the comparator were chosen from the same area as the selected care farms and participated in a wide range of activities such as working in a charity warehouse sorting second-hand clothes, and programs to address alcohol misuse, domestic violence, and anger-management or drink driving.

Data collection was carried out at two time points. The primary outcome was quality of life measured by the CORE-OM and the secondary outcome measure was mental well-being measured using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) (Tennant et al., 2007). Baseline data were collected at the start of the service of a community order and at follow-up six months later. A research assistant was present during baseline data collection to provide assistance to the participants. Participants asked for oral clarification if they found the questions difficult to understand. This information was recorded using a form, which the research assistant completed for each participant at baseline.

CORE-OM

The CORE-OM is a 34-item questionnaire developed in the UK, for routine use in psychological services (Barkham, Evans, Margison, & McGrath, 1998; Barkham, et al., 2005; Barkham, et al., 2001; Barkham, et al., 2006; Evans et al., 2000; Muehlenkamp and Gutierrez, 2007). It is comprised of four domains: well-being, problems (anxiety, depression, physical problems and trauma), functioning (general functioning, close relationships and social relationships), and risk to self and others. Using CORE-OM, Mavranezouli et al. (2011) developed the CORE-6D preference based measure which can be used to generate

utilities and QALYs (Mavranezouli, Brazier, Young, & Barkham, 2011). The CORE-6D is a 6-item instrument, assessing quality of life across two main dimensions - emotional and physical - each with three severity response levels, therefore, describing 729 unique health states. To produce utility scores for these health states, a valuation survey of a subset of these health states was undertaken in a representative sample of the UK population using the time trade-off technique and an algorithm was developed to link all possible health states to utility values (Mavranezouli, Brazier, Rowen, & Barkham, 2012).

Analysis

Completeness: It was assessed by analyzing the rates of missing values for each of the 34 items of the CORE-OM questionnaire at baseline and at follow-up. We also analyzed the completeness by calculating the response rate at follow-up.

Acceptability: We assessed the acceptability of the questionnaire to participants by systematically screening and coding the form filled by the research assistant, to identify if a particular question of the CORE-OM questionnaire was repeatedly reported as challenging to understand.

Distribution of scores: We explored the utility scores of participants using their individual responses to the CORE-OM questionnaire and converting them to CORE-6D using the algorithm developed by Mavranezouli et al. (Mavranezouli, et al., 2012). The algorithm was applied using STATA version 13. Utility scores were calculated for each participant at baseline and at follow-up with available answers to each item used for CORE-6D. Furthermore, we examined the presence of floor and ceiling effects because this impairs the ability of instruments to detect decreases or increases in quality of life. The ceiling and floor effects were calculated by the percentage of responses achieving the highest and lowest scores above the commonly used 15% threshold (Terwee et al., 2007). According to Terwee et al. (2007) if more than 15% of the responses achieved the lowest or highest possible score then the authors argue that it is likely that extreme values will be missing and consequently the sample presents limited content validity and the reliability of the collected data is reduced.

Measuring quality of life

Responsiveness: It is the ability of the CORE-6D instrument to detect change overtime. To assess responsiveness, we compared the baseline and follow-up mean scores of CORE-6D and WEMEBS which is a widely used mental wellbeing measure (Tennant, et al., 2007). We also calculated standardised response means (SRMs) which enables us to compare the magnitude of change measured by the two questionnaires. SRMs were calculated by taking the average difference divided by the standard deviation of the differences between the paired measurements (Husted, Cook, Farewell, & Gladman, 2000). Cut-off values of 0.20, 0.50 and 0.80 represent small, moderate, large changes (Husted, et al., 2000).

Results

Sample characteristics

One hundred and thirty-four participants were recruited to the ECO study, out of which 50 attended care farm intervention and 84 attended other community orders. The majority of study participants were male (70.9%) (see Table 1). The age range was between 18 and 65 years old with a mean age of 34 (SD 11). The mean WEMWBS score at baseline was 50.2 (SD 10.1) out of a total highest score of 70. Self-reported data shows that 29.9% of the participants reported using substances in the past four weeks. The most commonly used substance was cannabis (20.1%). The main results for the trial are reported elsewhere (Elsey, et al., 2018).

[Insert Table 1 here]

Completeness

Missing data were above 5% for only one item of the CORE-OM at baseline, and did not exceed 5% for any of the items at follow-up (Table 2). Missing data ranged between 0.75%

to 8.21% at baseline and 1.43% to 4.29% at follow-up. Question 30 ("*I have thought I am to blame for my problems and difficulties*") had the highest number of missing data points at baseline and follow-up. Of those who completed the follow-up questionnaire, 12 items had no missing data and overall, compared to baseline, missing data for follow-up questionnaires were low.

Missing data did not exceed 5% for any of the CORE-6D items neither at baseline nor at follow-up. For the CORE-6D items, missing data ranged between 0.75% to 4.48% at baseline and 1.43% to 2.86% at follow-up. At baseline, the highest number of missing data was found for question 21 ('*I have been able to do most things I needed to'*) with 6 missing data points. At follow-up question 21 and questions 8 ('*I have been troubled by aches, pains or other physical problems'*) had the highest number of missing data (2 each). Thus, all of the CORE-6D items experienced a low level of missing data.

[Insert Table 2 here]

Data completeness was hindered by two key challenges. First, the rate of missing data at baseline was distorted because of human error. When the questionnaire was administered to participants, 2 pages of the CORE-OM questionnaire were not included. This meant that 19 participants only received questions 1 to 10, hence increasing the number of missing data for questions 11 to 34. Second, the response rate at 6-months follow-up was only 52%. Sixty-four participants were lost to follow-up for reasons related to the study design. Follow-up questionnaires were intended to be completed face-to-face at the end of the participant's community order, however, participants who arranged to work their unpaid hours on planned days could change, as could the actual project site and a result the research assistant was unable to conduct the follow-up questionnaires face-to-face. Moreover, throughout this study probation services were undergoing a major reorganization dissolving probation trusts in

2014 and subsequently privatizing these services, this and the closure of one of the care farm sites had a significant impact on the response rates.

Acceptability

A research assistant was present when the participants first completed the questionnaire to provide assistance if needed. Of the 134 participants, 19 (14%) requested clarification from the research assistant to complete the CORE-OM questionnaire. Two or more participants needed clarification on questions 3, 17, 19 and 31. Two participants requested clarification about the question "I have someone to turn to for support" (Q3). Some asked for explanations of the words they did not understand namely, the meaning of overwhelmed (Q17), warmth and affection (Q19), and optimistic (Q31). Furthermore, two participants also highlighted that it was unclear whether to fill the CORE-OM questionnaire based on their experience of the past week or based on their overall experience in life; this happened despite the fact that the questionnaire clearly stated that participants should respond based on their experience over the last week.

Distribution of scores

At baseline and follow-up, a full range of responses was observed for the majority of items in the CORE-6D. Table 3 shows the proportion of participants within each level and dimension of the CORE-6D instrument. The highest proportion of responses was reported for question 16, which asked the participants to rate whether they had made plans to end their life, in the past week. This is however, expected as our sample reported mild to moderate symptoms, but Q16 may be more useful for a sample with severe mental health problems.

Measuring quality of life

We transformed the CORE-6D item scores into utilities following Mavranezouli et al. (Mavranezouli, et al., 2012). The overall utility score distributions were skewed to the right, indicating that participant's quality of life was high at baseline and follow-up (see Figures 1&2). The proportion of responses at the worst utility score at baseline and follow-up was less than 2%, which shows absence of floor effects. There was however, some ceiling effects as the proportion of scores at the highest level ranged from 22% at baseline and 17% at follow-up, which is slightly above the 15% threshold. So, this may suggest that changes in scores will be difficult to measure for those probationers at the highest possible score level.

[Insert Table 3 here; insert Figures 1&2 here]

Responsiveness

In both the CORE-6D and WEMWBS mean scores improved from baseline to follow-up. In a complete case sample of 50 participants, the means and standard deviations for CORE-6D at baseline and follow-up were 0.83 (0.16) and 0.85 (0.11), respectively (Table 4). In the same sample, the means and standard deviations for WEMWEBS at baseline and follow-up were 51.66 (9.42) and 53.92 (9.56), respectively. As expected in a feasibility study, no significant difference was found in the SRM statistics for both CORE-6D and WEMWBS. The SRMs for CORE-6D was 0.16 (small change) and the SRM for WEMWEBS was 0.27 (small change) with a 0.11 point difference in SRM statistic between the two instruments.

Discussion

CORE-OM is a valid and acceptable instrument to evaluate quality of life for people with common mental health problems. Completion rates were high for all the items in the CORE-

6D instrument. There was a small, non-significant SRM for both CORE-6D and WEMWEBS, this might be due to the fact that only 50 complete cases were included in the analysis and this pilot study was not powered to measure change over time.

Low levels of nonresponse show that the participants did not find the questionnaire difficult to understand or unduly challenging. Moreover, this is confirmed by the fact that only 14% of the participants requested help or clarification from the research assistant while completing the questionnaire. Nevertheless, question 30 – 'I have thought I am to blame for my problems and difficulties' - had a missing rate above the 5% threshold commonly considered acceptable for questionnaires, at baseline. It is unclear why nonresponse was high for this question, especially because the participants did not ask the research assistant to clarify the question. However, the literature shows that probationers struggle to come to terms with the attribution of blame and accountability (Batson, Gudjonsson, & Gray, 2010), thus this might have contributed to the high nonresponse rate. While this issue is observed with probationers, it is also known that probationers are likely to be a group with poor literacy and so this might not be observed for other people with common mental health problems who generally have a better literacy (Seymour, 2010).

Despite the skewed distribution of CORE-6D utility scores, all response categories were used for the CORE-6D items at baseline. This shows that the included participants experienced high quality of life, and also shows that the sample was heterogeneous. Nevertheless, the instrument allowed capturing this heterogeneity, which is evidenced by the large standard deviation and wide range of utility scores especially at baseline. Moreover, question 16 – '*I made plans to end my life*' – had over 90% of participants selecting "*not at all*". This may limit the items ability to detect moderate changes but given the negative nature of the question, and the overall high quality of life of the sample, it is not surprising that the vast majority selected "*not at all*". While there was no floor effects, there was some

indication that ceiling effects were present, and this may reduce the ability of the instruments to discriminate between participants at the healthy end of the scale. However, as mentioned above this might be because the included sample contained a large proportion of individuals reporting that they are healthy both according to CORE-OM and WEMWEBS. In addition, previous studies have reported ceiling effects of up to 25% (Ferreira, Ferreira, Ribeiro, & Pereira, 2016) in the widely used generic preference-based measure EQ-5D-3L; our results are similar but lower.

CORE-OM is a valuable measure to use for empirical studies among this population because of its ability to derive utilities using the CORE-6D instrument. CORE-6D has not been widely adopted in health economics empirical studies possibly because EQ-5D has long been the NICE preferred measure to generate utilities and ultimately quality-adjusted life years (QALYs). However, given the limitations of EQ-5D to measure quality of life of people with common mental health problems, this study suggests that CORE-6D could be used to generate utilities in this population. Moreover, CORE-OM presents the advantage of being suitable as a primary outcome in trials evaluating participants with common mental health problems and as a secondary outcome via CORE-6D for the health economics component. The use of the same measure would allow a shorter data collection package and consequently, reduce response fatigue and burden to the participants (Rolstad, Adler, & Rydén, 2011). If CORE-OM is the primary outcome of a study, the trial research manager will maximise collection of CORE-OM data from the participant and thus, health economists will benefit by being able to use CORE-6D with potentially lower levels of missing data.

Limitations

There are several limitations to this study. Due to recruitment difficulties the number of participants included in the study at follow-up was low; the response rate at 6-months follow-up was only 52%. Follow-up data assessments were missing because service users

often did not attend their last day at the project. As a result, changes in utility scores from baseline to follow-up were calculated for a smaller subset of the sample. If used in a full cost-effectiveness analysis, these data would require a strategy for imputing data for the missing responses items (Faria, Gomes, Epstein, & White, 2014).

We also found that the respondents followed up were older and used fewer substances. This should be considered when reviewing the conclusions of the study. Moreover, due to human error baseline CORE-OM data collection was not maximised. This is because two pages of the CORE-OM questionnaire were not included in the questionnaire pack. This error was corrected only after 19 participants filled in the incomplete questionnaire.

Further research

At the time this study was conducted between 2014 and 2015, CORE-OM was selected as the best instrument to assess the quality of life for people with common mental health problems. Since then however, ReQoL instrument was developed especially for people experiencing mental health difficulties (Keetharuth, et al., 2018). Moreover, to produce utility scores for ReQoL a valuation survey is currently underway (Keetharuth, et al., 2017). Consequently, potential work could be done to compare CORE-OM and ReQoL in the future and provide guidance on which instrument is recommended for use for subgroups of people such as those people with severe psychotic illnesses as opposed to people with common mental health problems. It is a well-known issue that we need a more appropriate instrument to measure the health-related quality of life of individuals with common mental health problems. This study shows that CORE-OM and the related CORE-6D can be recommended in future studies to assess quality of life for people with common mental health problems and evaluate the effectiveness and cost-effectiveness of mental health interventions.

References

Barkham, M., Evans, C., Margison, F., & McGrath, G. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health, 7*(1), p 35.

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *The British Journal of Psychiatry, 186*(3), pp. 239-246.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69*(2), p 184.

Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE System. *Counselling and Psychotherapy Research, 6*(1), pp. 3-15.

Barton, G. R., Hodgekins, J., Mugford, M., Jones, P. B., Croudace, T., & Fowler, D. (2009). Measuring the benefits of treatment for psychosis: validity and responsiveness of the EQ–5D. *The British Journal of Psychiatry, 195*(2), pp. 170-177.

Batson, A., Gudjonsson, G., & Gray, J. (2010). Attribution of blame for criminal acts and its relationship with psychopathy as measured by the Hare Psychopathic Checklist (PCL-SV). *Journal of Forensic Psychiatry & Psychology, 21*(1), pp. 91-101.

Bragg, R., & Atkins, G. (2016). A review of nature-based interventions for mental health care. *London: Natural England.*Retrieved from http://publications.naturalengland.org.uk/publication/4513819616346112

Bragg, R., Egginton-Metters, I., Elsey, H., & Wood, C. (2014). *Care farming: Defining the 'offer' in England.* .

Brazier, J. (2010). Is the EQ–5D fit for purpose in mental health? *The British Journal of Psychiatry, 197*(5), pp. 348-349.

Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of clinical epidemiology, 51*(11), pp. 1115-1128.

Burckhardt, C. S., & Anderson, K. L. (2003). The Quality of Life Scale (QOLS): reliability, validity, and utilization. *Health and quality of life outcomes, 1*(1), p 60.

Clatworthy, J., Hinds, J., & Camic, P. (2013). Gardening as a mental health intervention: a review. *Mental Health Review Journal, 18*(4), pp. 214-225.

Elsey, H., Bragg, R., Elings, M., Cade, J., Brennan, C., Farragher, T., Wickramasekera, N. (2014). Understanding the impacts of care farms on health and well-being of disadvantaged populations: a protocol of the Evaluating Community Orders (ECO) pilot study. *BMJ open, 4*(10), p e006536.

Elsey, H., Farragher, T., Tubeuf, S., Bragg, R., Elings, M., Brennan, C., Richardson, Z. (2018). Assessing the impact of care farms on quality of life and offending: a pilot study among probation service users in England. *BMJ open, 8*(3), p e019296.

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health*

Faria, R., Gomes, M., Epstein, D., & White, I. R. (2014). A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics, 32*(12), pp. 1157-1170.

Ferreira, L. N., Ferreira, P. L., Ribeiro, F. P., & Pereira, L. N. (2016). Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. *Health and quality of life outcomes, 14*(1), p 89.

Finch, A. P., Brazier, J. E., & Mukuria, C. J. T. E. J. o. H. E. (2018). What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *The European Journal of Health Economics, 19*(4), pp. 557-570.

Gilbody, S., Richards, D., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ–9 and CORE–OM. *Br J Gen Pract, 57*(541), pp. 650-652.

Gray, P., & Mellor-Clark, J. (2007). *CORE: a decade of development* Rugby: Page Bros.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research, 20*(10), pp. 1727-1736.

Horton, M., Wright, N., Dyer, W., Wright-Hughes, A., Farrin, A., Mohammed, Z., Tennant, A. (2014). Assessing the risk of self-harm in an adult offender population: an incidence cohort study.

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: a critical review and recommendations. *Journal of clinical epidemiology, 53*(5), pp. 459-468.

Keetharuth, A., Brazier, J., Connell, J., Bjorner, J. B., Carlton, J., Buck, E. T., Croudace, T. (2018). Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *The British Journal of Psychiatry, 212*(1), pp. 42-49.

Keetharuth, A., Brazier, J., Connell, J., Carlton, J., Taylor Buck, E., & Ricketts, T. (2017). *Development and Validation of the Recovering Quality of Life (ReQoL) Outcome Measures*. http://www.eepru.org.uk/article/development-and-validation-of-the-recovering-quality-of-life-reqol-outcome-measure/

Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernandez Alava, M., Mukuria, C., Evans, P. (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technology Assessment*

Mavranezouli, I., Brazier, J., Rowen, D., & Barkham, M. (2012). Estimating a Preference-Based Index from the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM) Valuation of CORE-6D. *Medical Decision Making*, p 0272989X12464431.

Mavranezouli, I., Brazier, J., Young, T., & Barkham, M. (2011). Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). *Quality of life research, 20*(3), pp. 321-333.

Muehlenkamp, J. J., & Gutierrez, P. M. (2007). Risk for suicide attempts among adolescents who engage in non-suicidal self-injury. *Archives of suicide research, 11*(1), pp. 69-82.

Murray, J., Wickramasekera, N., Elings, M., Bragg, R., Brennan, C., Richardson, Z., Shickle, D. (2019). The impact of care farms on quality of life, depression and anxiety among different population groups: A systematic review. *Campbell Systematic Reviews, 15*(4), p e1061.

NICE. (2013). Guide to the methods of technology appraisal 2013. Retrieved from https://www.nice.org.uk/article/pmg9/chapter/the-reference-case.

Papaioannou, D., Brazier, J., & Parry, G. (2011). How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review. *Value in Health, 14*(6), pp. 907-920.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in Health, 14*(8), pp. 1101-1108.

Saarni, S. I., Viertiö, S., Perälä, J., Koskinen, S., Lönnqvist, J., & Suvisaari, J. (2010). Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. *The British Journal of Psychiatry, 197*(5), pp. 386-394.

Seymour, L. (2010). *Promoting and protecting offenders' mental health and wellbeing*. London

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of life research, 13*(2), pp. 299-310.

Tapp, J., Fellowes, E., Wallis, N., Blud, L., & Moore, E. (2009). An evaluation of the Enhanced Thinking Skills (ETS) programme with mentally disordered offenders in a high security hospital. *Legal and Criminological Psychology, 14*(2), pp. 201-212.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and quality of life outcomes, 5*(1), p 63.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology, 60*(1), pp. 34-42.

Thorgrimsen, L., Selwood, A., Spector, A., Royan, L., de Madariaga Lopez, M., Woods, R., & Orrell, M. (2003). Whose quality of life is it anyway?: The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. *Alzheimer Disease & Associated Disorders, 17*(4), pp. 201-208.

van Nieuwenhuizen, C., Janssen-de Ruijter, E., & Nugter, M. (2017). *Handleiding Manchester Short Assessment of Quality of Life (MANSA)* Eindhoven: Stichting QoLM NL.

Tables

Table 1: Demographic characteristics of participants

Patient characteristics	Baseline (n=134)
Age (years), mean (SD), [range]	34.2 (11.4) [18 – 65]
Female, n (%)	39 (29.1)
Male, n (%)	95 (70.9)
WEMWEBS, mean (SD)[n]	50.2(10.1)[124]
Substance misuse, n (%)	40 (29.9)

Table 2: Missing data of the CORE-OM questionnaire

CORE-OM Item	Total missing at baseline (n=134)	Percent missing at baseline (n=134)	Total missing at follow-up (n=70)	Percent missing at follow-up (n=70)
1	1	0.75	0	0
2	4	2.99	1	1.43
3	3	2.24	1	1.43
4	2	1.49	1	1.43
5	3	2.24	2	2.86
6	2	1.49	1	1.43
7	3	2.24	0	0
8	3	2.24	2	2.86
9	1	0.75	0	0
10	2	1.49	1	1.43
11	7*	5.22	2	2.86
12	7*	5.22	1	1.43
13	6*	4.48	2	2.86
14	5*	3.73	0	0
15	5*	3.73	1	1.43
16	5*	3.73	0	0
17	6*	4.48	2	2.86
18	5*	3.73	0	0
19	6*	4.48	2	2.86
20	6*	4.48	2	2.86
21	6*	4.48	2	2.86
22	6*	4.48	1	1.43
23	5*	3.73	0	0
24	6*	4.48	0	0
25	5*	3.73	2	2.86
26	6*	4.48	0	0
27	6*	4.48	0	0
28	6*	4.48	2	2.86
29	4*	2.99	2	2.86
30	11*	8.21	3	4.29
31	7*	5.22	2	2.86
32	6*	4.48	1	1.43
33	4*	2.99	0	0
34	4*	2.99	0	0
*** The sample has lost an additional 19 answers because of 2 pages of the questionnaire were not given to the participants to complete due to human error.**				

Table 3: Percentage of responses by dimension and level for CORE-6D

	Item	(%) of responses for each level of severity				
		Not at all	Only occasionally	Sometimes	Often	Most of the time
Baseline	Q1	57.1	10.5	19.5	8.3	4.5
	Q8	41.2	10.7	25.2	11.5	11.5
	Q15	73.6	6.4	10.9	7.3	1.8
	Q16	91.8	1.8	4.5	0.9	0.9
	Q21	44	26.6	22.9	2.8	3.7
	Q33	67.6	12.6	12.6	5.4	1.8
Follow-up	Q1	64.3	11.4	18.6	2.9	2.9
	Q8	32.4	16.2	25	19.1	7.4
	Q15	79.7	8.7	8.7	2.9	0
	Q16	91.4	2.9	1.4	2.9	1.4
	Q21	47.1	20.6	27.9	1.5	2.9
	Q33	77.1	8.6	10	2.9	1.4

Q1: I have felt terribly alone and isolated

Q8: I have been troubled by aches, pains or other physical problems

Q15: I have felt panic or terror

Q16: I made plans to end my life

Q21: I have been able to do most things I needed to

Q33: I have felt humiliated or shamed by other people

Table 4: Baseline and follow-up outcomes scores and standard response means

Survey	Baseline Mean (SD) (N=50)	Follow-up Mean (SD) (N=50)	SRM (95% CIs)
CORE-6D	0.83 (0.16)	0.85 (0.11)	0.16 (-0.15 to 0.41)
WEMWBS	51.66 (9.42)	53.92 (9.56)	0.27 (-0.09 to 0.60)

Figures

Figure 1: Utility Score Distribution at Baseline

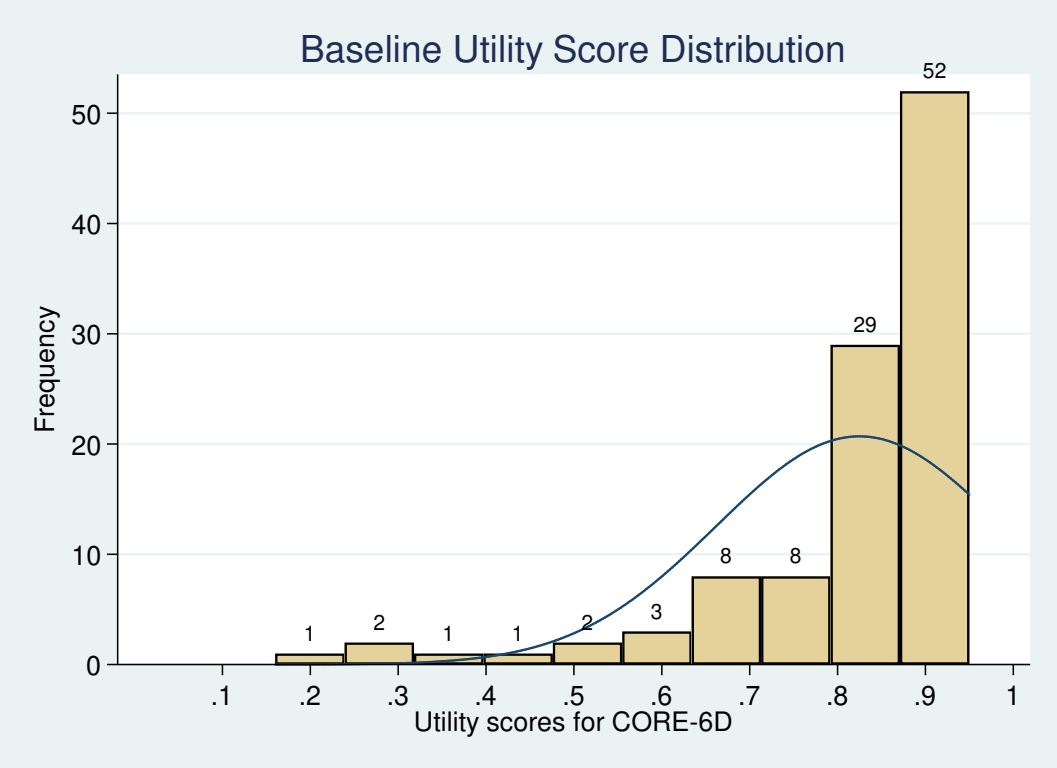

Figure 2: Utility Score Distribution at Follow-up