

A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures

John Brazier,^{1*} Janice Connell,¹ Diana Papaioannou,¹ Clara Mukuria,¹ Brendan Mulhern,¹ Tessa Peasgood,¹ Myfawnwy Lloyd Jones,¹ Suzy Paisley,¹ Alicia O’Cathain,¹ Michael Barkham,² Martin Knapp,³ Sarah Byford,³ Simon Gilbody⁴ and Glenys Parry¹

¹Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

²Centre for Psychological Services Research, University of Sheffield, Sheffield, UK

³Centre for the Economics of Mental and Physical Health, King’s College London, London, UK

⁴Department of Health Sciences, University of York, York, UK

*Corresponding author

Declared competing interests of authors: John Brazier was a developer of the SF-6D and Michael Barkham was a developer of the CORE-OM.

Disclaimer: This report contains transcripts of interviews conducted in the course of the research and contains language that may offend some readers.

Published May 2014

DOI: 10.3310/hta18340

Scientific summary

Measures of health in mental health populations

Health Technology Assessment 2014; Vol. 18: No. 34

DOI: 10.3310/hta18340

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

There has been an increasing use of the EQ-5D and SF-36® (and its derivative the SF-6D®) in the economic evaluation of health-care interventions. These generic measures can produce health-state utility values on a scale from 0 to 1 in order to calculate quality-adjusted life-years (QALYs). These QALY calculations are used by policy-makers [e.g. the National Institute for Health and Care Excellence (NICE)] to assess the cost-effectiveness of interventions in terms of their cost per QALY gained. Generic measures of health are also being used to assess the outcome of care in routine practice (e.g. the NHS Patient Reported Outcome Measures programme). There are concerns that generic measures have been primarily designed for physical health problems and miss important aspects of the impact of mental health problems on the quality of people's lives. Another concern is that even where generic measures are found to be appropriate, they are often not used in clinical studies.

Aims and objectives

This project examines in detail the appropriateness of the EQ-5D and the SF-36 and its derivatives in populations with mental health problems in terms of their validity (construct and content) and responsiveness. This research considers groups of people with mental health problems, including those with common mental health problems (e.g. mild to moderate depression, anxiety, obsessive compulsive disorder and panic disorders), severe and complex non-psychotic disorders (e.g. personality disorder), and schizophrenia and other psychotic disorders. For those conditions where the EQ-5D and/or SF-36 is found to be valid, a further aim is to estimate functions to predict their scores from mental health-specific measures commonly used in clinical trials and other studies (also known as mapping or cross-walking).

Methods and results

The validity of the EQ-5D and SF-36 (including its derivatives, the SF-12® and SF-6D) was examined using the psychometric criteria of validity and responsiveness. Validity is the extent to which an instrument measures the concept of interest, and in this case it is the extent to which EQ-5D and SF-36 (and its derivatives) reflect the impact of mental health problems on people's health-related quality of life, whether or not they use services. Construct and content validity are assessed in this research. Construct validity is assessed in terms of the ability of these measures to reflect known-group differences (or discriminant validity) and converge with other indicators of the concept (convergent validity). Responsiveness is the ability to reflect change in the population over time. Evidence on construct validity was obtained through the reviewing of published evidence and original analysis of seven existing data sets. Content validity assesses the degree to which the items of the measures represent the domains of interest, in terms of the impact of mental health problems on the quality of people's lives. Qualitative evidence on content validity was obtained from a systematic review of the literature and analyses of interviews of people with mental health problems.

In all, four studies were undertaken to examine the appropriateness of the measures, and a fifth study to estimate mapping functions between EQ-5D and SF-6D and commonly used mental health-specific measures.

Study 1: a systematic review of the validity and responsiveness of the EQ-5D, SF-36, SF-12 and SF-6D

The aim of this study was to assess research evidence on the validity and responsiveness of the SF-36, SF-12, SF-6D and EQ-5D in mental health conditions.

Methods

Systematic reviews were undertaken in five mental health conditions. Ten databases were searched from inception to August 2009. Studies were appraised and data extracted. A narrative synthesis was performed on known-group validity, convergent validity (strength of association with related measures such as symptoms or functioning) and responsiveness (e.g. changes in scores in responders/non-responders to treatment and correlation with changes in related measures).

Results

For schizophrenia, the majority of evidence related to the SF-36 (25 studies) and EQ-5D (nine studies). Both measures demonstrated known-group differences but this was mostly limited to differences between individuals with schizophrenia and the general population. Contradictory results were found in studies measuring convergent validity and responsiveness using clinical measures of symptom severity. For bipolar disorder, 23 studies were identified, almost exclusively on the SF-36, which was able to detect known differences in symptom severity and correlated strongly with clinical measures of depression, though weakly with measures of mania. For personality disorders, the majority of studies (six out of nine) related to the EQ-5D, which reflected known-group differences and responsiveness. For depression and anxiety, 23 EQ-5D and eight SF-6D studies were identified. Both measures demonstrated good convergent validity and responsiveness for depression; however, the results on known-group differences may be driven by the presence of comorbid depression in patients with anxiety disorders.

Study 2: assessing the validity and responsiveness of the EQ-5D and SF-6D using existing data sets

The purpose of Study 2 was to generate further evidence on the validity and responsiveness of the EQ-5D and SF-6D in anxiety, depression and schizophrenia samples by undertaking secondary analysis of available data sets.

Methods

Psychometric performance of the generic preference-based measures was assessed in comparison with mental health-specific measures using samples taken from seven patient data sets (total sample size = 5748 patients). For anxiety and depression, the EQ-5D was assessed in comparison with the Hospital Anxiety and Depression Scale (HADS), and the SF-6D was assessed in comparison with the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM). For the schizophrenia samples, the EQ-5D and SF-6D were both assessed using the Brief Psychiatric Rating Scale – Expanded (BPRS-E) and Positive and Negative Syndrome Scale (PANSS). Tests of acceptability (in terms of missing data and response rates), construct validity and responsiveness were carried out.

Results

Results support the findings from the review that both the EQ-5D and SF-6D are valid in common mental health disorders, particularly in depression but also in anxiety to a lesser extent. There was evidence that both measures were acceptable to respondents, are moderately correlated with the condition-specific instruments used, and can significantly discriminate between severity groups. Both measures also displayed a moderate level of responsiveness to change. The results in schizophrenia were less clear. The EQ-5D score appears able to reflect some symptoms of schizophrenia but evidence for the convergent validity of the SF-6D was limited. Both measures were able to discriminate between severity groups, but neither measure appeared responsive to change.

Study 3: estimating mapping functions between mental health-specific measures and the EQ-5D and SF-6D

The aim of this study was to provide mapping functions to predict EQ-5D or SF-6D values from condition-specific measures used in common mental health conditions, where the generic measures were found to be valid in Studies 1 and 2.

Methods

Using data from four studies ($n = 286$, $n = 475$, $n = 394$ and $n = 213$ patients respectively) available to the authors of this report, mapping functions were estimated by regression between the EQ-5D and the HADS, and between the SF-6D and the HADS, General Health Questionnaire (GHQ-12), Patient Health Questionnaire (PHQ-9), Generalised Anxiety Disorder Assessment (GAD-7) and CORE-OM. EQ-5D and SF-6D scores and dimension-level response were used as independent variables. For each mental health-specific measure, total scores, dimension scores and item response were entered as dependent variables. For models with dimension scores and item-level responses, additive models were fitted to the data, with interactions and squared terms explored for models with dimension scores. For the EQ-5D and SF-6D scores, ordinary least squares regression was used to estimate the models, and Tobit models were explored for the EQ-5D (owing to the presence of a large proportion of responders scoring 1). Multinomial models were also fitted to the models with dimension-level response as independent variables. The performance of the models was examined using R^2 (where appropriate), Bayesian information criterion, root-mean-square error and the ability to predict scores across severity groups.

Results

The results indicated that commonly used condition-specific measures for depression and anxiety are not good predictors of EQ-5D and SF-6D scores. The HADS was found to predict EQ-5D scores poorly in two data sets; it substantially overpredicted at the worst severity and underpredicted at the lowest severity. The HADS and other measures, notably the CORE-OM, were better able to predict the SF-6D, though the mapping functions continued to suffer from some degree of over- and underprediction towards the ends of the ranges.

Study 4: a synthesis of qualitative research on the quality of life of people with mental health problems

The aim of study 4 was to find out from people with mental health problems how their problems impact on their quality of life, using a review of the qualitative literature. This review provided the basis for the topic guide used in semistructured in-depth interviews and an initial framework for their analyses.

Methods

Our research firstly involved a systematic review of qualitative studies undertaken with people with mental health problems. Studies were only included where they reported the views of people with mental health problems and supported these with participant quotes. Participants had to be either diagnosed or attending an establishment for people with mental health problems. Extensive searching was undertaken using an iterative approach up to April 2012. Framework analysis was used to allow the identification of common and variable patterns of themes within and across different studies. The searching and reviewing was undertaken by two researchers. A multidisciplinary team met regularly, in addition to meetings with clinicians and a user representative to discuss and challenge the inclusion and exclusion criteria, thematic framework, and conceptual interpretations and conclusions.

Results

A framework analysis of 13 studies revealed six major themes: well-being and ill-being; control, autonomy and choice; self-perception; belonging; activity; and hope. In particular, it underlined the complexity of the measurement of quality of life in people with mental health problems and the difficulty in separating

the symptoms of mental health from other aspects of life quality. One of the strongest themes was the importance of a sense of belonging, derived from good quality relationships. However, relationships and activity can also have a negative impact. A limitation of the review was that available studies focused on quality of life of people with severe and enduring mental health problems, particularly schizophrenia rather than affective disorder. This is addressed in the interview study.

Study 5: results from qualitative semistructured interviews of people with mental health problems

The purpose of the fifth study was to examine further the domains of quality of life that are important to people with mental health problems through interviews, and to confirm or otherwise add to the themes found in the review. To address one of the limitations of the literature review, care was taken to recruit people with a full range of problems including those with mild to moderate affective problems.

Methods

Qualitative semistructured interviews were conducted with 19 people with a broad range of mental health problems at varying levels of severity. Participants were recruited through the local Improving Access to Psychological Therapies (IAPT) service for milder depression and anxiety, and through two community mental health teams who identified those with more severe problems, including complex non-psychotic disorders and psychotic disorders. The interviews were analysed thematically using framework analysis. The themes from the review made up the initial themes of the framework.

Results

Despite widening the types and severity of mental health problems studied, our interview data fitted well with the themes from the review. Any differences tended to be within the themes and related to the degree of impact of the themes on different levels of severity, chronicity and diagnosis. With some exceptions, those with severe chronic difficulties were more likely to talk about losses and things that took quality away from life, whereas those with moderate or relatively short-lived problems spoke of the things that added quality to life. It was also found that depression had a greater impact on quality of life than other problems such as anxiety, though coping mechanisms could have a detrimental effect on other areas of life (e.g. avoidance). The only change made to the themes was that physical health was found to be more important among the interviewees than suggested by the review, so this was added as a seventh theme, as opposed to a minor sub-theme as in the review. Another important finding was a conflict between the short-term stress and effort required for a fuller life (e.g. work or social activity) and the longer-term benefits.

The content of the EQ-5D and SF-6D has been reviewed against the seven themes identified in the qualitative research and the need to cover both positive and negative aspects. In summary, the EQ-5D covers little of the content of these seven themes because of its focus on physical health. Only physical health is covered, and, to some extent, activity, which is included in a rather crude way through usual activities. The EQ-5D covers ill-being in terms of depression and anxiety but not well-being. The SF-36 covers more, through having a multi-item dimension on mental health and a vitality dimension that includes more aspects of well-being and ill-being, and some aspects of social functioning. On the other hand, like the EQ-5D, it fails to include the psychological responses of people with mental health problems to occupational and social activities such as stress, self-perception and control, autonomy and choice, and hope and hopelessness. These generic measures do not contain a sufficient proportion of the domains identified by people with mental health problems in the qualitative research.

Conclusions

The EQ-5D and SF-36 achieve an adequate level of performance in terms of the classical psychometric tests for validity and responsiveness in depression and, to some extent, in anxiety and personality disorder. The results of the mapping work, however, raise some concerns as to whether or not the EQ-5D in particular is able to fully reflect the specific impact of common mental health problems. Results from the psychometric analyses were more mixed in schizophrenia and bipolar disorder, with a suggestion that the EQ-5D and SF-36 may be reflecting depression rather than other consequences of these conditions. Although important issues were identified within the quantitative evidence, it was unclear why the evidence appears so mixed in relation to the EQ-5D and SF-36. This was the rationale for undertaking the qualitative research as this provides evidence on the content validity of these measures in people with mental health problems. This found that the generic measures did not cover many of the theories identified in the qualitative research.

Although this project has been comprehensive in the way it has adopted a multimethod approach and sought to use the best available evidence, it does have a number of limitations. Many of the quantitative studies reviewed were limited in terms of size and coverage for conditions. The studies are normally designed for other purposes, and the measures used to assess construct validity and responsiveness were often clinical measures of symptoms that may not provide a good indicator of the concepts of interest. Although the qualitative work was intended to provide a more in-depth picture of the problem, the population of people with mental health problems in the qualitative research was not comprehensive despite an extensive review of the literature and an attempt to recruit across the spectrum of conditions for the interviews. Furthermore, the samples were influenced by the clinicians used to help recruit participants.

Future research recommendations

- Further testing of construct validity and responsiveness of these generic measures should be undertaken using a larger number of data sets.
- The analysis of content validity should be extended to existing condition-specific outcome measures used in mental health research. This is important for guiding the choice of mental health measures for use in research.
- There is a case for the development of a preference-based mental health measure. This could be an enhanced version of an existing generic measure, created by adding extra dimensions (i.e. 'bolt-ons'), or a new specific mental health measure for calculating QALYs, which might include one or two physical domains but would focus on the themes identified in this research.
- The qualitative research needs to be extended to those conditions not well represented (such as obsessive compulsive disorder) and should involve participants recruited through different channels to avoid the risk of selection bias that can result from a reliance on professionals.

Funding

Funding for this study was provided by the Medical Research Council.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme or, commissioned/managed through the Methodology research programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

This issue of the Health Technology Assessment journal series contains a project commissioned/managed by the Methodology research programme (MRP). The Medical Research Council (MRC) is working with NIHR to deliver the single joint health strategy and the MRP was launched in 2008 as part of the delivery model. MRC is lead funding partner for MRP and part of this programme is the joint MRC–NIHR funding panel 'The Methodology Research Programme Panel'

To strengthen the evidence base for health research, the MRP oversees and implements the evolving strategy for high quality methodological research. In addition to the MRC and NIHR funding partners, the MRP takes into account the needs of other stakeholders including the devolved administrations, industry R&D, and regulatory/advisory agencies and other public bodies. The MRP funds investigator-led and needs-led research proposals from across the UK. In addition to the standard MRC and RCUK terms and conditions, projects commissioned/managed by the MRP are expected to provide a detailed report on the research findings and may publish the findings in the HTA journal, if supported by NIHR funds.

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Brazier *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

Professor Jane Norman Professor of Maternal and Fetal Health, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professorial Research Associate, University College London, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk