This is a repository copy of *A review of generic preference-based measures of health-related quality of life in visual disorders*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/74367/

Article:

Tosh, J., Brazier, J., Evans, P. et al. (1 more author) (Completed: 2011) A review of generic preference-based measures of health-related quality of life in visual disorders. HEDS Discussion Paper, 11/03.

HEDS Discussion Paper 11/03

A review of generic preference-based measures of health-related quality of life in visual disorders

Jonathan Tosh[1]

John Brazier[1]

Philippa Evans[1]

Louise Longworth[2]

1 – Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, UK

2 – Health Economics Research Group, Brunel University, UK

Correspondence:
Jonathan Tosh
j.tosh@sheffield.ac.uk
T: 0114 222 0830
F: 0114 272 40495

Word Count: 4,465

Abstract

Purpose: This review examines generic preference based measures and their ability to reflect health related quality of life in patients with visual disorders.

Methods: A systematic search was undertaken to identify clinical studies of patients with visual disorders where health state utility values (HSUVs) were measured and reported. Data were extracted to assess the validity and responsiveness of the measures. A narrative synthesis of the data was undertaken due to the heterogeneity between different studies.

Results: There was considerable heterogeneity in the 31 studies identified in terms of patient characteristics, visual disorders and outcomes reported. Vision loss was associated with a reduction in scores across the preference-based measure, but the evidence on validity and responsiveness was mixed. The EQ-5D's performance differed according to condition, with poor performance in age-related macular degeneration and diabetic retinopathy. The more limited evidence on the HUI-3 found it performed best in differentiating between severity groups of patients with glaucoma, AMD, cataracts and diabetic retinopathy. One study reported data on the SF-6D and showed it was able to differentiate between patients with AMD.

Conclusion: The performance of the EQ-5D in visual disorders was mixed. The HUI-3 seemed to perform better in some conditions, but the evidence on this and SF-6D is limited. More head to head comparisons of these three measures are required. The new 5-level version of EQ-5D may do better at the milder end of visual function and there is research being undertaken into adding a vision relevant dimension.

Keywords

Abbreviations

AMD – Age-Related Macular Degeneration

CS – Contrast Sensitivity

HADS – Hospital Anxiety and Depression Scale

HRQoL – Health-related Quality of Life

HSUV – Health State Utility Value

IDEEI – Impact of Dry Eyes on Everyday Life questionnaire

HUI-3 – Health Utilities Index Mark 3

NICE – National Institute for Health and Clinical Excellence

NV-AMD – Neovascular Age-Related Macular Degeneration

PL – Perception of Light

QALY – Quality Adjusted Life Year

RQKQ – Rhinconjunctivitis Quality of Life Questionnaire

SF-6D – Short Form Six Dimension

SG – Standard Gamble

TTO – Time Trade Off

VA – Visual Acuity

VAS – Visual Analogue Scale

VF-4D – Visual Function instrument (4 dimension)

VF-14 – Visual Function instrument (14 item version)

VFA – Visual Function Assessment

VFQ-25 – Visual Function Questionnaire (25 Item version)

Introduction

Health state utility values (HSUVs) (Torrance, 1986) are key parameters for economic evaluations that use a cost-utility analysis framework. Cost-utility analyses return a cost effectiveness estimate for a new intervention in terms of the incremental cost per quality adjusted life year (QALY) gained. Reimbursement agencies such as the National Institute for Health and Clinical Excellence (NICE) in the UK, as well as other similar organisations worldwide, routinely undertake cost-utility analyses for new health care interventions.

HSUVs are most frequently derived from a description of a health state by a patient using a standardised generic instrument such as the EQ-5D, the SF-6D or the HUI-3. The patient's description is valued using a set of values obtained for each possible health state taken from a general population sample. These values are elicited using standard health state valuation techniques such as visual analogue scaling (VAS), standard gamble (SG) or time trade off (TTO).

It has been shown that the different instruments for obtaining HSUVs produce different values (Longworth & Bryan, 2003; Brazier, Roberts, Tsuchiya, & Busschbach, 2004; O'brien et al., 2003). It is important to assess the validity of any health outcome instrument, including those used to obtain HSUVs, in the particular condition of interest. Three widely used measures are EQ-5D, SF-6D and HUI-3. EQ-5D is a five dimension instrument, and measures the impact of a health state on mobility, self-care, usual activities, pain/discomfort, and depression/anxiety. Each dimension is given one of three levels of severity, which results in 243 unique health states (though a 5 level version of the descriptive system has recently been developed). The SF-6D (a classification system for the SF-36 and SF-12 health questionnaires) has six dimensions and either four or six severity levels. This results in 18,000 possible unique health states. The HUI-3 has eight dimensions and either five or six severity levels, resulting in 972,000 possible unique health states. The additional dimensions and levels may increase the instruments responsiveness to particular changes in health status; however this comes at the cost of increased patient burden and completion time, as well as requiring more health state valuations and/or increased uncertainty around these values. These instruments differ in the dimensions described, the number of levels and the severity range covered which may have implications for the appropriateness of an instrument to describe particular conditions. The fewer dimensions and levels in the EQ-5D compared to other instruments has been seen by some as a limitation. However the EQ-5D has been validated in many clinical areas and has shown that it has construct validity and is responsive to change (Brazier et al., 2004).

There has been concern over the validity of the EQ-5D instrument (Brazier et al., 2004; Browne et al., 2007) and SF-6D (Browne et al., 2007) in some visual disorders. Visual disorders are a broad set of conditions that can affect a patient in a range of ways. Certain conditions are painful, can affect central or peripheral vision, can affect one or both eyes

and can therefore impact on a patient's HRQoL. The aim of this review is to examine the appropriateness of the EQ-5D, SF-6D and HUI-3 in patients within visual disorders due to the different ways particular conditions impact on HRQoL.

Methods

Search strategy and data identification

The objective of the literature review was to identify relevant journal papers reporting evidence of the performance of EQ-5D, SF-6D and HUI-3 in patients with visual disorders.

A broad search was conducted to identify studies reporting preference-based utility instruments that were used to examine the HRQoL of patients with a visual disorder. BIOSIS, CINAHL, EMBASE, MEDLINE, PsychINFO, Web of Science electronic databases and the Euroqol website were searched. A search strategy was developed following consultation with experts in information resources and health economics. The search included both free text and controlled terms. Free text words included 'visual disorder', 'euroqol', 'hui3', 'sf6d' (all with alternative spellings). Specific visual disorders were also searched, including 'cataracts', 'retinopathy' and 'macular degeneration'. The criteria for inclusion was that patients had a visual disorder, the study reported at least one from the EQ-5D, SF-6D or HUI-3 and reported another measure of quality of life (generic or condition-specific) or a measure of clinical severity. Papers that used vignettes or own health state valuations were excluded. There was no restriction relating to the type of study. Due to resource limitations only English language studies were reviewed.

Analytic strategy

Data Extraction

Data were extracted from the studies using a standardised set of forms developed for this study after reviewing forms used for similar studies in other disease areas (Papaioannou, Brazier, & Parry, 2010; Pickard, Wilke, Lin, & Lloyd, 2007). Data extracted included:

- Study Characteristics – Country, type of visual disorder, disease or treatment stage, any treatment given, study type
- Participant Characteristics – Number of participants, age, gender, ethnicity, proportion of missing data
- Instrument used – EQ-5D/SF-6D/HUI-3, other health-related utility measures, other generic measures of HRQoL, condition specific HRQoL measures and clinical measures of disease severity,

- Health state utility values – index mean, scoring algorithm, VAS mean
- Construct and convergent validity
- Responsiveness

Assessment of quality

The assessment of quality of included studies requires a different methodology from the conventional approach required for reviewing clinical evidence. Of most importance was the relevance of the study in terms of the patient population. For studies that included a mixed population of patients (i.e. with various chronic conditions), then studies were only included if they reported HSUV's for sub-groups of patients with a specific visual disorder.

Also important are response or completion rates, which may have some implications for generalisability and provide evidence on the acceptability of the questionnaire to patients.

Assessment of validity

Validity is defined as how well an instrument measures what it was intended to measure. More specifically, for instruments constructed to measure HRQoL, whether the dimensions adequately cover the key determinants of HRQoL. Criterion validity would be determined by comparing an instrument to an established gold-standard. However a gold-standard for measuring health-related utility has not been established. Therefore researchers seek evidence that supports or otherwise the validity of a measure using the idea of construct validity, which attempts to see if patterns in scores confirm prior constructs.

The most common test to identify construct validity is the 'known group' method. This compares the results of a preference based measure between groups of patients that are expected to differ in the construct. If a study presents a population stratified on the basis of a clinical indicator, then one can investigate the ability of a preference based measure to distinguish between these groups. It should be noted that the usefulness of these comparisons can be limited by sample size (especially as studies are usually not powered on a preference based measure), the appropriateness of the clinical groups defined, and exogenous factors that may influence quality of life. For instance, groups defined solely by the presence of a biomarker will not have a clinical difference that impacts on their HRQoL. Also, if patients have a number of co-morbidities then these may have a greater impact on HRQoL than the condition of interest. Known groups can be defined using a case-control analysis. However, a more stringent test is to define known groups based on different levels of condition severity.

A sub-set of construct validity is known as convergent validity. This is defined as the extent to which one measure correlates with another measure of the same concept. In this review, this would be the extent to which EQ-5D, SF-6D or the HUI-3 correlate with each other and with measures of vision problems or quality of life.

Assessment of reliability

The reliability of a measure is defined as its ability to reproduce results when measurements are repeated on an unchanged population (Brazier & Deverill, 1999). Reliability can be measured by re-testing and reporting either the correlation or difference between estimates.

Assessment of responsiveness

Responsiveness is the ability to measure a change in health status. A pre/post intervention study which reports EQ-5D, SF-6D or the HUI-3 and another valid measure of health change would allow the responsiveness of a measure due to change in health status to be identified. As with the tests of validity, it is important to consider whether the measures of health change that are being used for comparison are themselves valid. In addition, it is important to consider whether other health changes not directly related to the condition could have impacted upon health-related utility (for example, side effects of treatment).

Presentation and analysis

Tables presenting summaries of the study characteristics, analyses on the impact of visual acuity on HSUVs are presented. The analysis will be broken down into particular visual disorders to allow for conclusions to be formed both on specific disorders as well as visual disorders as a whole. Heterogeneity in the studies reviewed, in terms of study design and patient populations, means that a formal meta-analysis would be inappropriate.

Results

Search results

Bibliographic searching was completed in August 2010. A total of 1,025 potentially relevant papers were identified. Abstracts and titles for all papers were screened to identify papers meeting the inclusion criteria. 969 records were excluded, and full papers were ordered for

the remaining 56 records. After reviewing the full papers, 25 were excluded and a total of 31 papers were included in the review. A flow chart of the study selection process is shown in Figure 1.

((INSERT FIGURE 1 HERE))

The 31 studies identified from the bibliographic search are presented in Table 1 and 2. Thirty of the 31 studies were observational studies, with the remaining study being a randomised controlled trial. The selected studies were conducted in different countries; three are multi-country studies, three are in the US, five in Canada and nine are from the UK.

((INSERT TABLE 1 HERE))

((INSERT TABLE 2 HERE))

Quality of Studies

A judgement regarding the risk of bias of each study was determined by reviewing the methods of patient recruitment, and noting any missing data reported (either study drop-outs or incomplete questionnaires). A range of recruitment procedures was seen in the review, and this was usually determined by the study design. Some involved a retrospective analysis of datasets (Mittmann, Chan, Trakas, & Risebrough, 2001; Asakawa et al., 2008) with a pre-determined inclusion criteria, a number of studies were either case-controlled analyses (Polack et al., 2010; Polack et al., 2008; Polack et al., 2007), but the majority were cross-sectional observational studies. The RCT reviewed (Datta et al., 2008) had well defined inclusion criteria. As is typical, in the longitudinal studies some patients dropped out before the end of the study. The general implication of patients dropping out was that no measurements were taken. Response rates for questionnaires range from 33% to 96%, with completion rates of longitudinal studies above 85% in all but one study (range 52% to 98.1%).

Patient Characteristics

The studies identified were in a wide range of visual disorders. Nine studies were in patients with cataract (Jayamanne et al., 1999; Polack et al., 2007; Polack et al., 2008; Black et al., 2009; Asakawa et al., 2008; Datta et al., 2008; Polack et al., 2010; Conner-Spady et al., 2005). Seven studies were in patients with age related macular degeneration (AMD) (Espallargues et al., 2005; Lotery et al., 2007; Soubrane et al., 2007; Ruiz-Moreno et al., 2008; Cruess et al., 2007; Payakachat et al., 2009; Kim et al., 2010), 5 studies were in patients with glaucoma (Aspinall et al., 2008; Kobelt et al., 2006; Mittmann et al., 2001;

Thygesen et al., 2008; Montemayor et al., 2001). Three studies were in patients with conjunctivitis (Smith et al., 2005; Pitt et al., 2004; Rajagopalan et al., 2005), 2 studies were in patients with diabetic retinopathy (Smith et al., 2008; Lloyd et al., 2008) and the remaining 6 studies were in populations with various other visual conditions.

The inclusion criteria varied across the studies reviewed within each of the specific conditions. This was due to the study design, the study methodology and also in some studies the inclusion criteria was unclear. Some studies reported that patients were identified through case notes, but no more details are provided. It was noted whether AMD was bilateral or unilateral and wet or dry, whether cataracts were first or second eye, and whether glaucoma was primary or multiple.

Reliability

No tests of reliability were performed on the generic preference based measures.

Construct validity

Twenty-one of the 31 studies allow a known group analysis to be performed, 17 for the EQ-5D, 4 for the HUI-3 but no studies for the SF-6D. In 6 of the studies, groups were defined by visual acuity (VA), or by contrast sensitivity (CS). These were clearly defined groups with mean estimates of utility provided for each group (Langelaan et al., 2007; Lloyd et al., 2008; Smith et al., 2005; Soubrane et al., 2007; Thygesen et al., 2008; Espallargues et al., 2005). The remaining 25 studies had either a case-control design, had differing conditions, or did not define levels of severity.

The differences in clinical definition of groups, conditions, characteristics of patients, and study designs do not allow for a direct comparison of the utility values, or a meta-analysis.

((INSERT TABLE 3 HERE))

Convergent validity

Nine of the 31 studies reviewed provide evidence on correlation or regression between generic measure of HRQoL with either each other or with visual measures. Eight studies report evidence of convergent validity in the EQ-5D compared to a visual measure, with Espallargues et al. also reporting correlations across EQ-5D, SF-6D and HUI-3.

Glaucoma

Construct Validity

Three studies of the EQ-5D allowed an analysis of groups defined by severity of vision problems in glaucoma patients (see Table 3). Aspinall et al. present EQ-5D utility values stratified by mild, moderate and severe visual field loss. The values decrease appropriately but are not statistically significant. Kobelt et al. saw the EQ-5D decrease with increasing glaucomatous damage, but the difference between groups was not statistically significant when controlling for co-morbidity, except for the most severe group. The study by Thygesen et al. defined three groups on the basis of the Snellen score, and saw consistent ordering of mean utility values.

No such data were available for HUI-3 or SF-6D. However, one paper reported the use of HUI-3 in a case-control study, which showed a significant and appropriate difference in HUI-3 between cases and controls (Mittmann et al., 2001).

Convergent Validity

Three studies report correlation statistics for EQ-5D with VA in patients with glaucoma (Aspinall et al., 2008; Montemayor et al., 2001; Thygesen et al., 2008). Aspinall et al. found moderate and statistically significant correlations for the mobility, self-care and anxiety dimensions, along with the summed index score. However Montemayor et al. did not find significant correlations for EQ-5D with VA. The study by Thygesen et al. showed a significant correlation between VA and EQ-5D.

Responsiveness

No studies reported responsiveness in patients with glaucoma.

AMD

Construct Validity

Of the seven papers, six allowed an assessment of construct validity of the EQ-5D in people with AMD. Of these, three differentiated between groups based on severity of vision disorder and four included assessments of cases against controls (one of which also grouped by severity).

Of the case-control studies, three found that EQ-5D showed an appropriate and statistically significant reduction in HRQoL for people with AMD compared to general population controls (Lotery et al., 2007; Ruiz-Moreno et al., 2008; Soubrane et al., 2007). One reported a difference that was not statistically significant difference, but the difference was in the appropriate direction (Cruess et al., 2007).

Three studies differentiated in terms of severity: one in terms of levels of visual acuity and the other based on whether they had unilateral or bilateral AMD. Soubrane et al. showed inconsistency with the mean estimates, with normal VA (>20/40) having a worse mean utility (0.69) when compared to mild, moderate, severe and near blind utility values. This inconsistency was not seen in the VFQ-25, however the HADS anxiety dimension was also inconsistent between the normal and mild VA groups. The study did however show a significant difference between those with NV-AMD and the control group. The study was relatively large (N=401 NV-AMD patients) however proportions in each group are not provided. Kim et al. found a statistically significant difference in EQ-5D values between those with unilateral and bilateral AMD. Espallargues et al. found a consistent relationship between VA and CS with HUI-3, SF-6D, TTO and VAS but not the EQ-5D.

Convergent Validity

Only Espallergues et al. provided correlation statistics between generic and visual measures in patients with AMD. They found that the VAS, TTO, HUI-3 and SF-6D were all significantly correlated with both VA and CS. However they did not find significant correlations for EQ-5D with VA or CS.

Responsiveness

Kim et al. reported a statistically significant improvement in both the VF-4D and the EQ-5D after photodynamic therapy in patients with AMD.

Cataracts

Construct Validity

Five of the 7 studies in people with cataracts allowed an assessment of the construct validity of the EQ-5D. Conner-Spady et al. identified an appropriate but non-significant change in EQ-5D between first and second eye surgery groups. Three case-control studies conducted in different countries by Polack et al (2007,2008 and 2010) found that there were significant differences in EQ-5D between cases and controls, and found that cases were likely to report

a significant difference across all dimensions (except pain dimension in Polack et al. 2008). However, there was no strong evidence to support a significant and consistent association between the degree of VA and EQ-5D. Polack et al. 2010 reports an inconsistent association between EQ-5D and VA level.

One study reports HUI-3 values for cases and control (Asakawa et al., 2008), and identifies statistically significant and appropriate difference between the two groups.

Convergent Validity

Three studies provide evidence of the convergent validity of the EQ-5D with VA. Polack et al. 2007 and 2008 tested associations between EQ-5D and VA, with one finding that poorer VA was associated with higher odds of reporting any problem with all EQ-5D dimensions apart from anxiety (Polack et al., 2007). The other study found no significant associations between VA and EQ-5D dimensions, apart from a borderline association with self-care (p=0.05) (Polack et al., 2008). Datta et al. did not find significant correlations for EQ-5D with VA.

Responsiveness

Black et al. reported a statistically significant improvement in both the VF-14 and the EQ-5D post cataract surgery, though the later was relatively small. Conner-Spady et al. reported a statistically significant improvement in the VFA and VA post cataract surgery, but the subsequent mean improvements in EQ-VAS and EQ-5D were small and not statistically significant. This may suggest that the EQ-5D is not responsive in this population, however it should be recognised that the study was not initially powered to identify statistically significant changes, and a mean improvement was identified. Also the VAS did not change pre-post treatment, so it could be that the treatment did not significantly impact on HRQoL.

Diabetic Retinopathy

Construct Validity

Two studies reported the EQ-5D identifying a statistically significant difference between the two extreme groups, however the differences between neighbouring groups were not significant, and frequently inconsistent (Smith et al., 2008; Lloyd et al., 2008). In the study by Lloyd et al, the inconsistencies were also shown in VAS ratings of patients' own health and the HUI-3. This may be due to small sample sizes or the authors speculate whether it may be due to a loss of independence of the participants when they reach that level of severity.

Convergent Validity

Smith et al. fitted a linear regression and found visual angle to be a predictor of EQ-5D utility values. They also fitted a non-parametric ordinal logistic regression and this estimated that any degree of visual impairment would see an increased likelihood of reporting non-perfect utility values.

Responsiveness

No papers reported the responsiveness of the measures in patients with diabetic retinopathy.

Conjunctivitis

Construct Validity

All three studies allowed an assessment of construct validity of the EQ-5D in people with conjunctivitis. All three were case-control studies and showed a statistically significant difference between cases and controls (Rajagopalan et al., 2005; Pitt et al., 2004; Smith et al., 2005). Within the dimensions of the EQ-5D, The study by Pitt et al. found the pain dimension to be the only dimension to show a statistical difference. The Smith et al. study saw a significant difference across all dimensions except mobility. No studies provide evidence on the construct validity of the HUI-3 or SF-6D.

Convergent Validity and responsiveness

No papers reported on convergent validity or the responsiveness of the measures in patients with conjunctivitis.

Other Visual Conditions

Construct Validity

The remaining five studies were in unique visual conditions. Three of these studies allow an assessment of the construct validity of the EQ-5D, and two of the HUI-3.

Clark et al. and Kempen et al. reported an appropriate but non-significant difference in the EQ-5D between the control group and those with endophthalmitis and cytomegalovirus, respectively. Langelaan et al. undertook a study in visually impaired patients, and identified an appropriate but non-significant difference in the EQ-5D between low and high visual field groups, but an inconsistent and non-significant difference in the EQ-5D between low and high visual acuity groups.

Boulton et al. and Quinn et al. found the HUI-3 identified statistically significant and appropriate differences between groups of patients with unspecified blindness/visual impairment.

Convergent Validity

van Nispen et al. undertook a multivariate regression analysis in older patients with a visual impairment. They found that worsening VA was a significant risk factor for a lower EQ-5D value.

Responsiveness

None of the papers reported data on the responsiveness of the measures.

Discussion

The 31 studies found in this review show a worsening of utility values as visual impairment increases in many though not all studies. The magnitude and statistical significance of the association varied between different generic preference-based measures of health related quality of life.

The largest amount of evidence was found for the EQ-5D compared to the other generic measures and the results were mixed. Nearly all studies showed significant differences between patients with the condition and a control group without it. However, this is a very crude test of construct validity and furthermore, many were not well controlled for age and other conditions then their conclusions may be limited. Studies comparing EQ-5D scores across severity groups were more mixed, with most showing little or no difference between groups defined by clinical measures of visual impairment. There was limited evidence on responsiveness, only in the form of before and after an intervention. The few studies identified changes consistent with an effective intervention, but differences were statistically significant in only two of three studies. The assessment of convergent validity was more concerning, with several studies not demonstrating a statistically significant

correlation with clinical measures. Whilst there was less evidence for the HUI-3, all but one study demonstrated good validity; no studies assessed responsiveness. There was very limited evidence on the SF-6D in patients with visual impairment.

((INSERT TABLE 4 HERE))

The results can also be grouped by visual disorder so to examine the performance of each generic measure. A summary of the overall performance by visual disorder is provided in Table 4. The EQ-5D performs well in patients with conjunctivitis, however the evidence is limited to case-control studies and no comparison is made to either generic HRQoL or clinical measures. In patients with diabetic retinopathy, both the EQ-5D and the HUI-3 distinguished between patients with and without the condition, however some evidence showed that both instruments were unable to distinguish between severity levels The EQ-5D was found to correlate with clinical measures in patients with diabetic retinopathy. In patients with AMD, the EQ-5D distinguished between patients with and without the condition, however it was unable to distinguish between severity levels and did not correlate well with other measures. The HUI-3 and the SF-6D did however distinguish according to severity and correlated well with other measures. Case-control evidence supports the EQ-5D and HUI-3 in patients with cataracts. One study of the EQ-5D in people with cataracts showed a non-significant trend reflecting severity, however the association of EQ-5D dimensions with other measures of severity was mixed.

In patients with glaucoma the EQ-5D distinguished between different levels of severity although this was not always statistically significant. Two of three studies in this condition showed that it correlated well with other measures but the third study failed to demonstrate a relationship. The HUI-3 distinguished between cases and controls in patients with glaucoma. The EQ-5D distinguished between people with and without conjunctivitis. In the "other" category, the EQ-5D evidence is mixed, but the HUI-3 is supportive.

While there are 31 studies providing evidence on the validity and responsiveness of EQ-5D, HUI3 and SF-6D, overall the evidence base is weak. Much of the evidence is limited to comparisons with general population scores and this is rather crude. Many of the studies had small numbers, so some of the inconsistent findings and lack of statistical significance could be due to small numbers. Furthermore, this often did not control for age or co-morbidities that may correlate with visual impairment. Finally, there were very few head to head studies that really enable a true comparison of performance.

A new version of the EQ-5D has been developed with the number of levels increased to five rather than three. It is possible that this could improve the EQ-5D's ability to demonstrate differences in utility between people with milder severities of visual impairment. Further research should be conducted on the validity of the five-level version in people with visual impairment. There is also interest in adding dimensions to EQ-5D to make it more relevant

for certain conditions and there is currently a study being undertaken to look at the impact of a dimension to pick up the impact of visual impairment on HRQL. Another solution would be to develop a preference-based condition specific index using a widely used Vision specific instrument such as the VFQ-25 (Revicki et al., 2008) though there are concerns that condition specific preference-based measures may not be comparable across different medical conditions (Brazier & Tsuchiya, 2010).

Conclusion

The review has provided a narrative analysis of preference based measures in visual disorders. The broad range of values, and the differing levels of performance in terms of construct validity, convergent validity and responsiveness reflects the systematic variance attributable to different types of visual disorder, different patient populations and to study design. The number of studies investigating the EQ-5D is much larger than for HUI-3 or SF-6D. The results of this review show generally consistent results on the validity of the HUI-3 in people with visual impairment with the exception of diabetic retinopathy, the results for EQ-5D were mixed and there was little evidence on the use of the SF-6D. Responsiveness was only assessed in the EQ-5D and this was found to be consistent, but not always statistically significant. More head to head comparisons are required of these measures across visual conditions.

References

Reference List

Asakawa, K., Rolfson, D., Senthilselvan, A., Feeny, D., Johnson, J. A., Asakawa, K. et al. (2008). Health Utilities Index Mark 3 showed valid in Alzheimer disease, arthritis, and cataracts. *Journal of Clinical Epidemiology, 61,* 733-739.

Aspinall, P. A., Johnson, Z. K., Azuara-Blanco, A., Montarzino, A., Brice, R., Vickers, A. et al. (2008). Evaluation of quality of life and priorities of patients with glaucoma. *Investigative Ophthalmology & Visual Science, 49,* 1907-1915.

Black, N., Browne, J., van der Meulen, J., Jamieson, L., Copley, L., Lewsey, J. et al. (2009). Is there overutilisation of cataract surgery in England? *British Journal of Ophthalmology, 93,* 13-17.

Boulton, M., Haines, L., Smyth, D., Fielder, A., Boulton, M., Haines, L. et al. (2006). Health-related quality of life of children with vision impairment or blindness. *Developmental Medicine & Child Neurology, 48,* 656-661.

Brazier, J. & Deverill, M. (1999). A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Economics, 8,* 41-51.

Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics, 13,* 873-884.

Brazier, J. & Tsuchiya, A. (2010). Preference-based condition specific measures of health: what happens to cross programme comparability. *Health Economics, 19,* 125-129.

Browne, J., Jamieson, L., Lewsey, J., van der Meulen, J., Black, N., Cairns, J. et al. (2007). Patient Reported Outcome Measures (PROMs) in elective surgery: report to the Department of

Health. *Health Services Research Unit, London School of Hygiene & Tropical Medicine & Clinical Effectiveness Unit, Royal College of Surgeons of England*.

Clark, A., Ng, J. Q., Morlet, N., Tropiano, E., Mahendran, P., Spilsbury, K. et al. (2008). Quality of life after postoperative endophthalmitis. *Clinical & Experimental Ophthalmology, 36,* 526-531.

Conner-Spady, B. L., Sanmugasunderam, S., Courtright, P., Mildon, D., McGurran, J. J., Noseworthy, T. W. et al. (2005). The prioritization of patients on waiting lists for cataract surgery: validation of the Western Canada waiting list project cataract priority criteria tool. *Ophthalmic Epidemiology, 12,* 81-90.

Cruess, A., Zlateva, G., Xu, X., Rochon, S., Cruess, A., Zlateva, G. et al. (2007). Burden of illness of neovascular age-related macular degeneration in Canada. *Canadian Journal of Ophthalmology, 42,* 836-843.

Datta, S., Foss, A. J., Grainge, M. J., Gregson, R. M., Zaman, A., Masud, T. et al. (2008). The importance of acuity, stereopsis, and contrast sensitivity for health-related quality of life in elderly women with cataracts. *Investigative Ophthalmology & Visual Science, 49,* 1-6.

Espallargues, M., Czoski-Murray, C. J., Bansback, N. J., Carlton, J., Lewis, G. M., Hughes, L. A. et al. (2005). The impact of age-related macular degeneration on health status utility values. *Investigative Ophthalmology & Visual Science, 46,* 4016-4023.

Jayamanne, D. G., Allen, E. D., Wood, C. M., Currie, S., Jayamanne, D. G., Allen, E. D. et al. (1999). Correlation between early, measurable improvement in quality of life and speed of visual rehabilitation after phacoemulsification. *Journal of Cataract & Refractive Surgery, 25,* 1135-1139.

Kempen, J. H., Martin, B. K., Wu, A. W., Barron, B., Thorne, J. E., Jabs, D. A. et al. (2003). The effect of cytomegalovirus retinitis on the quality of life of patients with AIDS in the era of highly active antiretroviral therapy. *Ophthalmology, 110,* 987-995.

Kim, J., Kwak, H. W., Lee, W. K., Kim, H. K., Kim, J., Kwak, H. W. et al. (2010). Impact of photodynamic therapy on quality of life of patients with age-related macular degeneration in Korea. *Japanese Journal of Ophthalmology, 54,* 325-330.

Kobelt, G., Jonsson, B., Bergstrom, A., Chen, E., Linden, C., Alm, A. et al. (2006). Cost-effectiveness analysis in glaucoma: what drives utility? Results from a pilot study in Sweden. *Acta Ophthalmologica Scandinavica, 84,* 363-371.

Langelaan, M., de Boer, M. R., van Nispen, R. M., Wouters, B., Moll, A. C., van Rens, G. H. et al. (2007). Impact of visual impairment on quality of life: a comparison with quality of life in the general population and with other chronic conditions. *Ophthalmic Epidemiology, 14,* 119-126.

Lloyd, A., Nafees, B., Gavriel, S., Rousculp, M. D., Boye, K. S., Ahmad, A. et al. (2008). Health utility values associated with diabetic retinopathy. [Review] [25 refs]. *Diabetic Medicine, 25,* 618-624.

Longworth, L. & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics, 12,* 1061-1067.

Lotery, A., Xu, X., Zlatava, G., Loftus, J., Lotery, A., Xu, X. et al. (2007). Burden of illness, visual impairment and health resource utilisation of patients with neovascular age-related macular degeneration: results from the UK cohort of a five-country cross-sectional study. *British Journal of Ophthalmology, 91,* 1303-1307.

Mittmann, N., Chan, D., Trakas, K., & Risebrough, N. (2001). Health utility attributes for chronic conditions. *Disease Management & Health Outcomes, 9,* 11-21.

Montemayor, F., Sibley, L. M., Courtright, P., Mikelberg, F. S., Montemayor, F., Sibley, L. M. et al. (2001). Contribution of multiple glaucoma medications to visual function and quality of life in patients with glaucoma. *Canadian Journal of Ophthalmology, 36,* 385-390.

O'Brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P., & Brazier, J. (2003). A view from the bridge: agreement between the SF 6D utility algorithm and the Health Utilities Index. *Health Economics, 12,* 975-981.

Papaioannou, D., Brazier, J., & Parry, G. (2010). The validity and responsiveness of the SF-36, SF-12, EQ-5D and SF-6D in schizophrenia: a systematic review. *Under Review*.

Payakachat, N., Summers, K. H., Pleil, A. M., Murawski, M. M., Thomas, J., III, Jennings, K. et al. (2009). Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Quality of Life Research, 18,* 801-813.

Pickard, A. S., Wilke, C. T., Lin, H. W., & Lloyd, A. (2007). Health utilities using the EQ-5D in studies of cancer. *Pharmacoeconomics, 25,* 365-384.

Pitt, A. D., Smith, A. F., Lindsell, L., Voon, L. W., Rose, P. W., Bron, A. J. et al. (2004). Economic and quality-of-life impact of seasonal allergic conjunctivitis in Oxfordshire. *Ophthalmic Epidemiology, 11,* 17-33.

Polack, S., Eusebio, C., Fletcher, A., Foster, A., Kuper, H., Polack, S. et al. (2010). Visual impairment from cataract and health related quality of life: results from a case-control study in the Philippines. *Ophthalmic Epidemiology, 17,* 152-159.

Polack, S., Kuper, H., Mathenge, W., Fletcher, A., Foster, A., Polack, S. et al. (2007). Cataract visual impairment and quality of life in a Kenyan population. *British Journal of Ophthalmology, 91,* 927-932.

Polack, S., Kuper, H., Wadud, Z., Fletcher, A., Foster, A., Polack, S. et al. (2008). Quality of life and visual impairment from cataract in Satkhira district, Bangladesh. *British Journal of Ophthalmology, 92,* 1026-1030.

Quinn, G. E., Dobson, V., Saigal, S., Phelps, D. L., Hardy, R. J., Tung, B. et al. (2004). Health-related quality of life at age 10 years in very low-birth-weight children with and without threshold retinopathy of prematurity. *Archives of Ophthalmology, 122,* 1659-1666.

Rajagopalan, K., Abetz, L., Mertzanis, P., Espindle, D., Begley, C., Chalmers, R. et al. (2005). Comparing the discriminative validity of two generic and one disease-specific health-related quality of life measures in a sample of patients with dry eye. *Value in Health, 8,* 168-174.

Revicki, D. A., Nafees, B., Meguro, M., Kowalski, J., Walt, J., Hansen, J. et al. (2008). Item reduction of National Eye Institute Visual Function Questionnaire - 25 Using Rasch Analysis. In.

Ruiz-Moreno, J. M., Coco, R. M., Garcia-Arumi, J., Xu, X., Zlateva, G., Ruiz-Moreno, J. M. et al. (2008). Burden of illness of bilateral neovascular age-related macular degeneration in Spain. *Current Medical Research & Opinion, 24,* 2103-2111.

Smith, A. F., Pitt, A. D., Rodruiguez, A. E., Alio, J. L., Marti, N., Teus, M. et al. (2005). The economic and quality of life impact of seasonal allergic conjunctivitis in a Spanish setting. *Ophthalmic Epidemiology, 12,* 233-242.

Smith, D. H., Johnson, E. S., Russell, A., Hazlehurst, B., Muraki, C., Nichols, G. A. et al. (2008). Lower visual acuity predicts worse utility values among patients with type 2 diabetes. *Quality of Life Research, 17,* 1277-1284.

Soubrane, G., Cruess, A., Lotery, A., Pauleikhoff, D., Mones, J., Xu, X. et al. (2007). Burden and health care resource utilization in neovascular age-related macular degeneration: findings of a multicountry study. *Archives of Ophthalmology, 125,* 1249-1254.

Thygesen, J., Aagren, M., Arnavielle, S., Bron, A., Frohlich, S. J., Baggesen, K. et al. (2008). Late-stage, primary open-angle glaucoma in Europe: social and health care maintenance costs and quality of life of patients from 4 countries. *Current Medical Research & Opinion, 24,* 1763-1770.

Torrance, G. W. (1986). Measurement of health state utilities for economic appraisal:: A review. *Journal of health economics, 5,* 1-30.

van Nispen, R. M., de Boer, M. R., Hoeijmakers, J. G., Ringens, P. J., van Rens, G. H., van Nispen, R. M. A. et al. (2009). Co-morbidity and visual acuity are risk factors for health-related quality of life decline: five-month follow-up EQ-5D data of visually impaired older patients. *Health & Quality of Life Outcomes, 7,* 18.

Table 1: Characteristics of included studies

Study reference (Author, Year)	Country	Disease/treatment stage	Study type (e.g. cross sectional, RCT, cohort)
GLAUCOMA			
(Aspinall et al., 2008)	UK	Glaucoma and no other ocular co morbidity	Cross-sectional study
(Kobelt et al., 2006)	Sweden	Ocular hypertension or open-angle glaucoma	Cross-sectional study
(Mittmann et al., 2001)	Canada	Glaucoma – a subset from a study on a range of chronic conditions	Cross-sectional study
(Montemayor et al., 2001)	Canada	COAG, normal-pressure glaucoma or suspected glaucoma with treatment	Cross-sectional study
(Thygesen et al., 2008)	Multi-country	Late-stage primary open-angle glaucoma	Case review
AMD			
(Cruess et al., 2007)	Canada	NV-AMD	Cross-sectional observational study
(Espallargues et al., 2005)	UK	Wet or dry AMD	Cross-sectional study
(Kim et al., 2010)	Korea	-	Cohort study
(Lotery et al., 2007)	UK	Bilateral subfoveal NV-AMD	Cross-sectional study
(Payakachat et al., 2009)	Multi-country	Wet AMD	Cross-sectional study
(Ruiz-Moreno et al., 2008)	Spain	Bilateral NV-AMD	Prospective case-control study
(Soubrane et al., 2007)	Multi-country	NV-AMD	Cross-sectional study
CATARACTS			
(Asakawa et al., 2008)	Canada	+/- other co-morbidities	Cross-sectional study
(Black et al., 2009)	UK	First or second eye	Prospective cohort study
(Conner-Spady et al., 2005)	Canada	-	Cohort study
(Datta et al., 2008)	UK	Bilateral cataracts in over 70s	Secondary analysis of RCT
(Jayamanne et al., 1999)	UK	First Eye	Prospective study
(Polack et al., 2007)	Kenya	-	Case-control study
(Polack et al., 2008)	Bangladesh	-	Case-control study
(Polack et al., 2010)	The Philippines	Over 50s	Case control study
DIABETIC RETINOPATHY			
(Lloyd et al., 2008)	UK	Diabetic Retinopathy due to diabetes	Cross-sectional study
(Smith et al., 2008)	US	Type-2 diabetes	Cross-sectional study
CONJUNTIVITIS			
(Pitt et al., 2004)	UK	-	Cohort study
(Rajagopalan et al., 2005)	Multi-country	Non-Sjogren's Keratoconjunctivitis or Sjogren's Syndrome	Cross-sectional study
(Smith et al., 2005)	Spain	-	Cohort study
OTHER VISUAL DISORDERS			
(Boulton et al., 2006)	UK	Vision Impairment or blindness in children	Cross-sectional study
(Clark et al., 2008)	Australia	Post cataract surgery endophthalmitis	Cohort study
(Kempen et al., 2003)	US	Cytomegalovirus retinitis in patients with AIDS	Prospective cohort study
(Langelaan et al., 2007)	Netherlands	Low-vision patients	Cross-sectional study
(Quinn et al., 2004)	US	Retinopathy of Prematurity	Cohort study
(van Nispen et al., 2009)	Netherlands	Vision Impairment in older people	Observational study

Table 2: Instruments used

Study reference (Author, Year)	Generic Preference Based			Direct valuation	Rating scale	Condition specific HRQoL instruments and measures of clinical severity				
	EQ-5D	SF-6D	HUI-3	TTO	VAS	VFQ (20/25)	VF (14/4D)	RQLQ	VFA	IDEEL
GLAUCOMA										
(Aspinall et al., 2008)	✓			✓						
(Kobelt et al., 2006)	✓				✓					
(Mittmann et al., 2001)			✓							
(Montemayor et al., 2001)	✓								✓	
(Thygesen et al., 2008)	✓									
AMD										
(Cruess et al., 2007)	✓					✓				
(Espallargues et al., 2005)	✓	✓	✓	✓	✓		✓			
(Kim et al., 2010)	✓						✓			
(Lotery et al., 2007)	✓					✓				
(Payakachat et al., 2009)	✓					✓				
(Ruiz-Moreno et al., 2008)	✓					✓				
(Soubrane et al., 2007)	✓					✓				
CATARACTS										
(Asakawa et al., 2008)			✓							
(Black et al., 2009)	✓						✓			
(Conner-Spady et al., 2005)	✓				✓				✓	
(Datta et al., 2008)	✓						✓			
(Jayamanne et al., 1999)	✓				✓					
(Polack et al., 2007)	✓					✓				
(Polack et al., 2008)	✓					✓				
(Polack et al., 2010)	✓					✓				
DIABETIC RETINOPATHY										
(Lloyd et al., 2008)	✓		✓			✓				
(Smith et al., 2008)	✓									
CONJUNCTIVITIS										
(Pitt et al., 2004)	✓				✓	✓		✓		
(Rajagopalan et al., 2005)	✓				✓					✓
(Smith et al., 2005)	✓				✓	✓		✓		
OTHER VISUAL DISORDERS										
(Boulton et al., 2006)			✓							
(Clark et al., 2008)	✓			✓		✓				
(Kempen et al., 2003)	✓				✓					
(Langelaan et al., 2007)	✓									
(Quinn et al., 2004)			✓							
(van Nispen et al., 2009)	✓									

VFQ20/25 = Visual Function Questionnaire 20/25. VF-14/4D = Visual Functional Questionnaire 14/4 dimension. RQLQ = Rhinoconjunctivitis Quality of Life Questionnaire. VFA = Visual Function Assessment. IDEEI = Impact of Dry Eyes on Everyday Life questionnaire.

Table 3: Utility, Visual Acuity and VAS values

Study reference (Author, Year)	Instrument	Index (mean (SD))	VA (logMAR mean (SD) unless specified)	VAS (mean (SD))
GLAUCOMA				
(Aspinall et al., 2008)	EQ-5D	0.76 (0.19)	Median group 6/12 or better in both eyes	
(Kobelt et al., 2006)	EQ-5D	0.80 (0.23)	(right/left) 0.76/0.74 (0.30/0.29)	74.7 (18.2)
(Mittmann et al., 2001)	HUI-3	0.924 (0.086)		
(Montemayor et al., 2001)	EQ-5D	0.89 (range -0.08 to 1.00)	-0.10 (0.17)	
(Thygesen et al., 2008)	EQ-5D	0.65 (0.28)	Best/worst eye 0.28(0.26) / 0.14(0.18)	
AMD				
(Cruess et al., 2007)	EQ-5D	0.64 (0.52,0.76)	0.66 (0.64)	
(Espallargues et al., 2005)	EQ-5D	0.72 (0.22)	Better seeing eye 1.01 (0.67)	65.0
	SF-6D	0.66 (0.14)	Worse seeing eye 1.68 (0.75)	
	HUI-3	0.34 (0.28)		
(Kim et al., 2010)	EQ-5D	Pre-treatment 0.729 (0.236) Post-treatment 0.793 (0.222)		
(Lotery et al., 2007)	EQ-5D	0.67	0.26 (0.19)	
(Payakachat et al., 2009)	EQ-5D	0.7711 (0.21)	Median group for Better and worse eye groups: 20/80 to 20/160	
(Ruiz-Moreno et al., 2008)	EQ-5D	0.68 (0.62,0.74) 95% CI		
(Soubrane et al., 2007)	EQ-5D	0.65	0.6 (0.7)	
CATARACTS				
(Asakawa et al., 2008)	HUI-3	-		
(Black et al., 2009)	EQ-5D	0.81 (0.23)		
(Conner-Spady et al., 2005)	EQ-5D	First eye surgery 0.80 (0.20) Second eye surgery 0.78 (0.20)	First eye, second eye VA op 0.58 (0.30), 0.52 (0.27) VA non-op 0.43 (0.28), 0.29 (0.21)	First eye surgery 77.5 (17.9) Second eye surgery 77.1 (16.6)
(Datta et al., 2008)	EQ-5D	Median 0.73 (0.26)	0.28 [0.16;0.42]	
(Jayamanne et al., 1999)	EQ-5D	-	90% had VA 6/18-6/60	
(Polack et al., 2007)	EQ-5D	-	Median group <6/60 >3/60	
(Polack et al., 2008)	EQ-5D	-	Median group <3/60 > PL	
(Polack et al., 2010)	EQ-5D	-	Median group <3/60 > PL	
DIABETIC RETINOPATHY				

(Lloyd et al., 2008)	EQ-5D, HUI-3	By level of VA: No DR - 0.83 (0.20), 0.81 (0.20) 6/6-6/9 - 0.83 (0.20), 0.78 (0.22) 6/12-6/18 - 0.50 (0.30), 0.30 (0.38) 6/24-6/36 - 0.68 (0.29) , 0.61 (0.35) 6/60-6/120 - 0.53 (0.47), 0.52 (0.50) Counting fingers - 0.34 (0.36), 0.37 (0.00)	By level of VA: No DR - 33% 6/6-6/9 - 45% 6/12-6/18 - 9% 6/24-6/36 - 7% 6/60-6/120 - 5% Counting fingers - 2%	By level of VA: No DR 75.0 (20.6) 6/6-6/9 - 73.8 (18.1) 6/12-6/18 - 57.7 (22.7) 6/24-6/36 - 65.9 (21.1) 6/60-6/120 - 52.3 (31.3) Counting fingers -50.1 (9.0)
(Smith et al., 2008)	EQ-5D	0.8 (0.18)	Median group >20/20	
CONJUNCTIVITIS				
(Pitt et al., 2004)	EQ-5D	-		SAC 81.69 (14.89) Control 84.92 (12.54)
(Rajagopalan et al., 2005)	EQ-5D	Control 0.87 (0.03) Non-SS KCS 0.82 (0.02) SS 0.74 (0.03)		Control 88.93 (2.06) Non-SS KCS 82.45 (1.19) SS 66.94 (2.43)
(Smith et al., 2005)	EQ-5D	-		SAC 80.09 (15.24) Control 83.34 (11.86)
OTHER VISUAL DISORDERS				
(Boulton et al., 2006)	HUI-3	0.34 (0.43)		
(Clark et al., 2008)	EQ-5D	Cases 0.66 (0.32) Controls 0.81 (0.25)		
(Kempen et al., 2003)	EQ-5D	No CMV = 0.71 Long standing CMV = 0.75 Newly diagnosed CMV = 0.75	No CMV = 91 (median) Long standing CMV = 88 Newly diagnosed CMV = 88	No CMV = 72.5 Long standing CMV = 72.2 Newly diagnosed CMV = 63.9
(Langelaan et al., 2007)	EQ-5D	0.73 (0.22)	Functional Acuity Score 38.61 (26.5)	
(Quinn et al., 2004)	HUI-3	All subjects 0.59 (0.39) Blind or low vision in better eye 0.25 (0.37) Sighted in better eye 0.78 (0.25) No-ROP subjects 0.90 (0.16)	All subjects - 20/63 Blind or low vision in better eye - Blind Sighted in better eye - 20/40 No-ROP subjects - 20/20	
(van Nispen et al., 2009)	EQ-5D	Respondents 0.69 (0.24) Non-respondents 0.57 (0.29)	Respondents 0.55 [0.42;0.77] Non-respondents 0.52 [0.41;0.80]	

PL = Perception of light.

Table 4: Overall performance by visual disorder

		EQ-5D	HUI-3	SF-6D
Glaucoma	Severity	✓ ✓ ✓	.	.
	Case-control	.	✓	.
	Correlation	✓ x ✓	.	.
	Responsiveness	.	.	.
AMD	Severity	✓ x x	✓	✓
	Case-control	✓ ✓ ✓ ✓	.	.
	Correlation	x	✓	✓
	Responsiveness	✓	.	.
Cataracts	Severity	✓ ✓ x x	.	.
	Case-control	✓ ✓ ✓	✓	.
	Correlation	x ✓ x	.	.
	Responsiveness	✓ ✓	.	.
Diabetic Retinopathy	Severity	x x	x	.
	Case-control	✓ ✓	✓	.
	Correlation	✓	.	.
	Responsiveness	.	.	.
Conjunctivitus	Severity	.	.	.
	Case-control	✓ ✓ ✓	.	.
	Correlation	.	.	.
	Responsiveness	.	.	.
Other	Severity	?	✓ ✓	.
	Case-control	✓ ✓	.	.
	Correlation	✓	.	.
	Responsiveness	.	.	.

KEY

✓ statistically significant

✓ trend meeting prior expectation but difference not statistically significant

x Inconsistent or non-significant correlation

? mixed

. No evidence

Figure 1: Flow diagram showing selection of studies

```
┌─────────────────────────────────────┐
│ Number of potentially relevant       │
│ records                              │
│                                      │
│ 1025                                 │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐        ┌─────────────────────────────────────┐
│ Number of citations screened         │───────▶│ Number of citations excluded         │
│                                      │        │                                      │
│ 1025                                 │        │ 969                                  │
└─────────────────────────────────────┘        └─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐        ┌─────────────────────────────────────┐
│ Number of full text articles assessed│───────▶│ Number of full text articles excluded│
│                                      │        │                                      │
│ 56                                   │        │ 25                                   │
└─────────────────────────────────────┘        └─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│ Number of studies included in review │
│                                      │
│ 31                                   │
└─────────────────────────────────────┘
```