
Article:
Rowen, D., Carlton, J., McDool, E. orcid.org/0000-0002-3530-7921 et al. (83 more authors) (2025) Psychometric performance of a new condition-specific preference-weighted measure, VILL-UI, and EQ-5D-5L in patients with age-related macular degeneration: A MACUSTAR Study Report. Value in Health. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2025.04.2155

Psychometric performance of a new condition-specific preference-weighted measure, VILL-UI, and EQ-5D-5L in patients with age-related macular degeneration: A MACUSTAR Study Report

Authors: Donna Rowen, PhD[1*], Jill Carlton, PhD[1], Emily McDool[1], PhD, Frank G. Holz[2], MD, PhD, Nadia Zakaria, PhD[3], Jan H. Terheyden J H, MD[2], Robert P. Finger, MD, PhD,[2], on behalf of the MACUSTAR Consortium

[1] Sheffield Centre for Health and Related Research, University of Sheffield, UK

[2] Department of Ophthalmology, University of Bonn, Germany

[3] Novartis AG, Basel, Switzerland

*Corresponding author information:

Donna Rowen, PhD

Sheffield Centre for Health and Related Research, University of Sheffield

UK, South Yorkshire, Sheffield

Email: d.rowen@sheffield.ac.uk

Phone: +44114 222 0728

MACUSTAR consortium members: H. Agostini, I. D. Aires, L. Altay, R. Atia, F. Bandello, P. G. Basile, J. Batuca, C. Behning, M. Belmouhand, M. Berger, A. Binns, C. J. F. Boon, M. Böttger, J. E. Brazier, C. Carapezzi, J. Carlton, A. Carneiro, A. Charil, R. Coimbra, D. Cosette, M. Cozzi, D. P. Crabb, J. Cunha-Vaz, C. Dahlke, H. Dunbar, R. P. Finger, E. Fletcher, M. Gutfleisch, F. Hartgers, B. Higgins, J. Hildebrandt, E. Höck, R. Hogg, F. G. Holz, C. B. Hoyng, A. Kilani, J. Krätzschmar, L. Kühlewein, M. Larsen, S. Leal, Y. T. E. Lechanteur, D. Lu, U. F. O. Luhmann, A. Lüning, N. Manivannan, I. Marques, C. Martinho, A. Miliu, K. P. Moll, Z. Mulyukov, M. Paques, B. Parodi, M. Parravano, S. Penas, T. Peters, T. Peto, S. Priglinger, R. Ramamirtham, R. Ribeiro, D. Rowen, G. S. Rubin, J. Sahel, C. Sánchez, O. Sander, M. Saßmannshausen, M. Schmid, S. Schmitz-Valckenberg, J. Siedlecki, R. Silva, E. Souied, G. Staurenghi, J. Tavares, D. J. Taylor, J. H. Terheyden, A. Tufail, P. Valmaggia, M. Varano, A. Wolf, N. Zakaria

Precis: Comparison of psychometric performance of a new condition-specific preference-weighted measure and generic preference-weighted EQ-5D-5L in patients with age-related macular degeneration

Author contributions:

All authors certify that they meet the ICMJE criteria for authorship.

Conflict of Interest Disclosures: Rowen, Carlton report receiving grants from the EuroQol Research Foundation for other studies. Terheyden reports receiving funding from Heidelberg Engineering, Optos, Carl Zeiss Meditec, CenterVue (devices), Novartis, Okko (recipient). Finger reports receiving funding from Bayer, Ellex, Novartis, Novartis, Opthea, Alimera, Santhera, Roche/Genentech, CentreVue, Zeiss. Rowen is a member of the EuroQol Group.

Funding/Support: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116076. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The communication reflects the author's view and neither IMI nor the European Union, EFPIA, or any Associated Partners are responsible for any use that may be made of the information contained therein.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgements: None.

Ethics Approval and Consent to Participate: For the MACUSTAR study, local ethics committees at each of the study sites reviewed and approved the protocols. All patients provided written informed consent to participate in those studies and for the use of the study data for research.

Abstract

Aims: The Vision Impairment in Low Luminance-Utility Index (VILL-UI) is a novel preference-weighted measure for use in patients with age-related macular degeneration (AMD). No evidence exists on its psychometric performance, nor its performance in comparison to generic preference-weighted measure, EQ-5D-5L commonly used in economic evaluation. This study compares the psychometric performance of VILL-UI and EQ-5D-5L in patients with AMD.

Methods: Assessments of feasibility, convergent/divergent validity and known-group validity of VILL-UI and EQ-5D-5L are undertaken using MACUSTAR data at baseline, 12, 24 and 36 months. Analyses are undertaken separately using UK and German preference weights for both measures.

Results: The sample with complete responses (n=586) had mean age 71.9 years (standard deviation 6.9), 65.2% women, with predominantly intermediate AMD (87.2%). VILL-UI and EQ-5D-5L are feasible for completion, though VILL-UI has fewer usable responses due to its response options (baseline 89% vs 100%). EQ-5D-5L has high ceiling effects, with around one third of participants reporting the best health state compared to under 8% for VILL-UI. Convergent validity between EQ-5D-5L and VILL-UI utilities and dimensions where a relationship is expected is low, with divergent validity demonstrated where expected. VILL-UI detected statistically significant differences in known-groups for visual acuity, visual function and AMD stage across most timepoints, with little evidence of known-group validity for EQ-5D-5L.

Conclusions: VILL-UI is appropriate for use in future AMD studies to inform economic evaluation. VILL-UI has superior performance to EQ-5D-5L for known-group validity and has fewer ceiling effects, but has fewer usable responses.

Key words: age-related macular degeneration; EQ-5D-5L, VILL-UI, psychometrics

Highlights:

What is already known on this topic: The VILL-UI is a newly developed condition-specific preference-weighted measure for use in patients with age-related macular degeneration (AMD), but its psychometric performance is unknown.

What this study adds: This study is the first to assess the psychometric performance of VILL-UI, also in comparison to generic EQ-5D-5L, using a large multicountry sample of patients with AMD.

How this study might affect research, practice or policy: This study generates a new knowledge about the performance of VILL-UI and EQ-5D-5L in patients with AMD to inform the use and interpretation of these measures in economic evaluation in AMD.

Introduction

Many international reimbursement agencies and decision-making bodies recommend use of a generic preference-weighted measure (PWM) to assess benefits of healthcare and treatments using economic evaluation [1]. A PWM consists of a classification system used to assign a participant to a health state, and corresponding preference weights that generate utility values for all health states [2]. Utilities reflect how good or bad health states are on a scale where 1 equals full health, 0 reflects a health state deemed equivalent to being dead, and values below zero indicate the health state is considered worse than being dead. A generic PWM measures generic health-related quality of life (HRQoL) and can be used across all patient groups and conditions, enabling comparability and consistency in evaluations and decisions across different conditionsNICE, the National Institute for Health and Care Excellence in England and Walesrecommend the use of generic EQ-5D [3], a widely used generic PWM that has both a three-level version (EQ-5D-3L) [4] and a newer five-level version (EQ-5D-5L) [5]. . EQ-5D-5L is often specifically named for use by international reimbursement agencies [1].

Whilst generic PWMs are recommended for use in all patient groups, they are not necessarily appropriate for use in people with all health conditions, as their generic content may not reflect all that is important to people with a certain condition [6]. For instance, the five dimensions covered by EQ-5D-5L do not cover the impact of sensory perception other than pain. Generic PWMs may also not capture improvements or deteriorations in health that have been demonstrated clinically (see for example a recent review assessing meaningful change [7]). This can mean that they do not detect change where it has occurred, or detect a lower change than suggested clinically. This impacts on the results of cost-effectiveness analyses, as the smaller change in benefit can make the treatments appear less cost-effective, and potentially more expensive than the cut-off required for funding new treatments.

Age-related macular degeneration (AMD) is a condition of central, specialized area of the retina responsible for central vision. It is the most common cause of severe visual loss in industrialized countries and affects almost 25% of the population 60+ in Europe [8]. Early and intermediate AMD are not associated with legal blindness but impair vision in low luminance and low contrast situations, whilst late stage AMD is almost always associated with a severe loss of vision and ultimately blindness. Total annual costs due to AMD are estimated to exceed 2 billion Euros in the European Union alone [9]. Visual impairment is one area where EQ-5D has been found to have poor psychometric performance [10,11]. The performance of EQ-5D-3L in AMD in particular has been assessed previously. Studies have found weak evidence of construct validity [12,13] and the only study assessing responsiveness found EQ-5D-3L was responsive to treatment [14]. Whilst three studies found that EQ-5D-3L could distinguish between people with AMD and a control group [15-17], and three studies found some evidence of known-group validity [12,14,18], six studies found no evidence of known-group validity meaning EQ-5D-3L could not distinguish between different severity groups where we would expect it to be able to clinically [13,15-17,19-20]. To our knowledge, there have been no published studies assessing the psychometric performance of EQ-5D-5L in people with AMD.

The VILL-UI is a new AMD-specific PWM designed for use in people with AMD that can generate health state utilities reflecting vision-related quality of life (VRQoL) for use in economic evaluation [21]. The VILL-UI was derived from an existing patient-reported outcome measure (PROM), the VILL-33 [22,23], and has both United Kingdom (UK) and

German preference weights [21]. However, the psychometric performance of VILL-UI has not yet been examined. Since NICE recommend the use of EQ-5D where it is appropriate, comparing the psychometric performance of VILL-UI and EQ-5D-5L in people with AMD is of interest for two reasons. Firstly, to determine differences found by the measures, and secondly, to assess whether VILL-UI can detect change or differences in groups of different severity of AMD.

This study assesses and compares the psychometric performance of VILL-UI and EQ-5D-5L in patients with AMD using a large multi-country AMD patient dataset, and comparisons are made using both UK and German preference weights for each measure.

Methods

VILL-UI and EQ-5D-5L

The Vision Impairment in Low Luminance (VILL-33) measure is a self-report PROM, that has been recently developed for use in patients with AMD [22,23]. The VILL-33 consists of 33 items which focus on visual impairment and VRQoL under challenging luminance and contrast conditions. The VILL-33 has been shown to be content valid, construct valid, criterion valid, test-retest reliable, internally consistent and responsive to changes over time [23-25]. The VILL-UI is an AMD-specific PWM derived from the VILL-33 that reflects the three domains of the VILL-33: reading and accessing information; mobility and safety; and emotional well-being [21]. Reading and accessing information is assessed using "recognizing small objects in dim lighting (e.g. coins)" ("information") and "reading print against a colourful background (e.g. a brochure)" ("reading"), and each have four severity levels (1 = none, 2 = a little, 3 = a lot, 4 = can't do). Mobility and safety is assessed jointly using the combination of items "seeing steps or curbs in the dark" and "feeling unsafe as a pedestrian or cyclist at dawn or at night" with eight severity levels. Emotional well-being is assessed using "feel worried that your eyesight might get worse" with four severity levels (1 = never, 2 = sometimes, 3 = often, 4 = always). VILL-UI is generated from VILL-33 data, but all VILL-33 items have a response option "Don't do this for other reasons" and VILL-UI utilities cannot be generated where participants select this response for one or more items used in the VILL-UI classification system. UK and German preference weights were generated using modelled data from an online discrete choice experiment (DCE) survey with a duration attribute undertaken with a representative sample of members of the public [21]. The UK preference weights have a utility range of 1 to -0.084 and the German preference weights range from 1 to -0.182.

The EQ-5D-5L is a generic PWM with 5 dimensions of mobility (i.e. walking about), self-care (i.e. washing or dressing yourself), usual activities, pain/discomfort, and anxiety/depression. Each dimension has five levels of severity (1 = none, 2 = slight, 3 = moderate, 4 = severe, 5 = extreme/unable). UK utilities have been generated using the current method recommended by NICE (mapping [26] onto UK EQ-5D-3L utilities [27]), and German utilities have been generated using the German preference weights [28] with a range of 1 to -0.661.

Sample

Analyses are conducted using the MACUSTAR study dataset [29,30]. MACUSTAR is a longitudinal, prospective cohort study on AMD, registered on clinicaltrials.gov under NCT03349801. The study is conducted at 20 clinical sites across Europe (Denmark, France,

Germany, Italy, Netherlands, Portugal and UK) and included patients with early, intermediate and late AMD. VILL-33 and EQ-5D-5L questionnaires were self-completed unless patients requested interviewer administration, administered in different languages across different countries with translation following a standardized protocol described previously [23]. In several scientific advice procedures conducted by the MACUSTAR consortium with the European Medicine Agency (EMA), EMA concluded that the MACUSTAR sample is representative of the general European AMD population and thus results can be generalized (see respective EMA letters of support available at www.ema.eu). MACUSTAR data until four years of follow-up was collected April 2018 to April 2023 and accessed 8th November 2023. Four timepoints are used where both VILL-UI and EQ-5D-5L data are available: baseline; 12 months; 24 months; 36 months. Feasibility assessments are conducted on all data (there are missing data for EQ-5D-5L and VILL-UI) since they assess missing data (baseline n=661; 12 months n=474; 24 months; n=428; 36 months n=422). All remaining analyses use complete data for EQ-5D-5L and VILL-UI, that is, only including cases where both EQ-5D-5L and VILL-UI were completed at that timepoint with no missing data (baseline n=586; 12 months n=416; 24 months n=378; 36 months n=367). This approach is taken to ensure that any differences in results across EQ-5D-5L and VILL-UI across the assessments are not due to differences in samples.

Analyses

The analyses report the sample characteristics and assess the feasibility, convergent validity and known-group validity of VILL-UI and EQ-5D-5L using both UK and German preference weights to generate utilities for each measure. There is no gold standard PWM for vision or AMD, and hence analyses report comparative performance of these two measures. The statistical analysis plan was informed by recent psychometric analyses of preference weighted measures [31,32].

Feasibility assessments examine the practicality of the measures for completion by people with AMD. This is assessed using missing data for each EQ-5D-5L and VILL-UI dimension, since high levels of missing data indicate a lack of evidence for acceptability and feasibility.

Floor and ceiling effects are examined for each dimension and the utilities of each measure. Where there are large proportions of responses observed at the least severe responses there is an inability to capture an improvement in health (referred to here as ceiling effects). Where there are large proportions of responses observed at the most severe level there is an inability to capture a deterioration in health (referred to here as floor effects). For a clinical population, large floor and ceiling effects indicate that the measure is unlikely to fully capture clinical changes in the condition. At the measure level, ceiling and floor effects are flagged where these are >15% of participants (for all dimensions combined) [33,34].. For patients reporting full health in EQ-5D-5L, their responses to VILL-UI are examined to assess the ability of VILL-UI to capture the impact of visual impairment where no impact was indicated using EQ-5D-5L. It is hypothesised that EQ-5D-5L dimensions have larger ceiling effects than VILL-UI dimensions.

Convergent validity assessments examine the strength of association between EQ-5D-5L and VILL-UI. Evidence of convergent validity is determined by whether moderate (0.41-0.60) or good (0.61-0.80) (see [35]) agreement is observed where these are motivated theoretically (e.g. emotional wellbeing dimension in VILL-UI and anxiety/depression dimension in EQ-5D-5L). Divergent validity is demonstrated where the measures or

dimensions are not correlated where this is also theoretically motivated (e.g. information and reading dimensions in VILL-UI and EQ-5D-5L mobility and pain/discomfort), which is expected to be the case in this instance for several dimensions given the different content (VRQoL vs generic HRQoL) of the measures. Correlations between VILL-UI and EQ-5D-5L utilities is assessed using Pearson correlation coefficients, and correlations between dimensions is assessed using Spearman rank correlation coefficients. Moderate correlations are hypothesised between: VILL-UI mobility and safety and EQ-5D-5L mobility; VILL-UI emotional well-being and EQ-5D-5L anxiety/depression; VILL-UI information, reading, mobility and safety and EQ-5D-5L usual activities. It is hypothesised that the other dimensions are not correlated, meaning that there is divergent validity, since they capture different aspects.

Known-group validity assessments examine the ability to differentiate between groups of different severity. This is assessed using VILL-UI and EQ-5D-5L utilities and dimensions using the distribution of responses, including mean across subgroups and across different time points, and using effect sizes calculated using Cohen's D. Evidence of known-group validity using effect sizes considers 0.2-0.49 as small, 0.5-0.79 as moderate, and ≥0.8 to be large effect sizes [36]. For differences in means across groups, known-group validity is determined by whether there is a statistically significance difference at the 5% level across known groups (using the p-value of a t-test) and whether the direction of the difference is in accordance with clinical expectation (according to the clinical authors). The known groups that are examined are AMD severity (early/intermediate vs late), visual acuity using BCVA (best-corrected visual acuity, no impairment vs any impairment) [37], visual function (no dysfunction vs dysfunction [37]) and presence of late AMD in the fellow eye (no presence vs presence). Clinical expectation was that utilities would be higher (i.e. indicating lower severity) for the lower severity groups (early/intermediate AMD, no impairment, no dysfunction) in comparison to the higher severity group (late AMD, any impairment, dysfunction). Visual dysfunction was defined as a visual acuity below the 5th percentile of healthy control participants in the MACUSTAR study [37]. Analyses are conducted separately for each of the four timepoints for AMD stage, but for other severity groups analysis is only undertaken at baseline as the severity groupings were not available in later timepoints. It is hypothesised that VILL-UI has moderate known-group validity and that EQ-5D-5L does not reflect all known-group differences.

Change over time was explored using a subsample of participants with complete VILL-UI and EQ-5D-5L data at all timepoints, using utility values at each timepoint for all participants and by AMD severity (early/intermediate vs late).

Analyses were undertaken using Stata version 18.

Results

Table 1 reports sample characteristics. The sample at baseline (n=586) has mean age 71.9 years (standard deviation 6.9) and 65.2% are female. The sample has predominantly intermediate AMD (87.2% at baseline to 83.3% at 36 months).

Figures 1 and A1 in supplementary materials show the distribution of EQ-5D-5L and VILL-UI utilities by timepoint. EQ-5D-5L has a narrower range than VILL-UI, particularly when using German weights. Both measures have a right-skewed distribution, though EQ-5D-5L has larger ceiling effects and VILL-UI is less skewed.

Both VILL-UI and EQ-5D-5L are feasible. In terms of missing data, at baseline all 661 participants fully completed EQ-5D-5L whereas 586 completed VILL-UI (see Table A1 in supplementary materials). The percentage of participants with missing data (responses are unable to be used to generate utilities) for VILL-UI varies from 10.3% to 13.0% by timepoint and from 0% to 2.1% for EQ-5D-5L.

Table 2 shows the distribution of dimension responses and mean (standard deviation) utility for each measure at each timepoint. EQ-5D-5L has high ceiling effects (e.g. at baseline between 50.9% and 95.4% of observations in contrast to 21.5% to 62.1% for VILL-UI). Due to differences in country preference weights, mean VILL-UI utility per timepoint is lower using German weights, whereas mean EQ-5D-5L utility is higher using German weights.

Table 3 shows the distribution of VILL-UI dimension responses at each timepoint when the participant is in EQ-5D-5L full health. The proportion of participants at full health using EQ-5D-5L, (i.e. ceiling response to all dimensions), varies from 32.3% at 36 months to 37.0% at baseline, with all timepoints exceeding the 15% cut-off. In contrast, the proportion of participants in the best state using VILL-UI varies from 6.6% at 24 months to 7.3% at baseline (see Table A2 in supplementary materials for EQ-5D-5L responses for participants reporting best state using VILL-UI). Table 3 shows that all VILL-UI dimensions capture VRQoL problems not reflected in EQ-5D-5L when the participant is in good health, with notably only 29.2% to 33.3% of participants at the least severe level for VILL-UI emotional wellbeing, despite having no problems in the anxiety/depression EQ-5D-5L dimension.

Convergent and divergent validity between EQ-5D-5L dimensions and VILL-UI dimensions is shown in Table 4 at baseline and Tables A3 to A5 in the supplementary materials for the other timepoints. In contrast to our hypotheses, low correlations between dimensions are observed where moderate correlations were expected. Divergent validity is observed across the remaining EQ-5D-5L and VILL-UI dimensions as expected, indicating that these dimensions capture different aspects.

Correlations between EQ-5D-5L and VILL-UI utilities are low, where for the 12 month timepoint only there is moderate correlation.

Known-group validity assessments are shown in Table 5. Known-group validity assessed using AMD stage was significantly demonstrated at baseline and 36 months for VILL-UI (both country weights) using t-tests and with moderate or large effect sizes, and across all timepoints the ordering of mean values across severity groups was in accordance with clinical expectation even when not significant. For EQ-5D-5L this was only significant using German weights at baseline, though the effect size was small, and at the 12 month timepoint the mean values for the different AMD stage groups were not in accordance with clinical expectations.

VILL-UI detected a statistically significant difference using t-tests in known-groups for visual function and visual acuity (see Table 5). EQ-5D-5L detected a statistically significant difference for visual function but not for visual acuity. All of the effect sizes were small for visual function and visual acuity.

The ordering of VILL-UI utilities across severity groups in the fellow eye (AMD stages according to the Beckman classification) was in accordance with clinical expectations but was not statistically significant. EQ-5D-5L was not in accordance with clinical expectations. All effect sizes were small for presence of late AMD in the fellow eye.

For all participants with utilities at all timepoints, VILL-UI utilities demonstrate change over time (see Supplementary Materials Table A6) that would be expected clinically, where utilities decrease over time and are lower for participants with more severe AMD. EQ-5D-5L utilities do not decrease as expected at 24 months, and using the German weights do not reflect severity at 24 and 36 months.

Discussion

This study is the first to assess the psychometric performance of the newly developed VILL-UI PWM derived from the VILL-33 and EQ-5D-5L in patients with AMD, and the first to compare the performance of the two measures. The study is of international importance and significance due to the use of a multi-country, multi-centre dataset [29], and the application of different country value sets for each measure. The results provide valuable information to those conducting quality of life studies in AMD and undertaking cost-effectiveness analyses of interventions in AMD.

The VILL-UI has superior performance for known-group validity to EQ-5D-5L, with fewer ceiling effects, but higher missing data. EQ-5D-5L has little evidence of known-group validity, and large ceiling effects with over one third of patients reporting full health. There is low correlation between EQ-5D-5L and VILL-UI utilities and dimensions common to each measure where moderate correlation may be expected. Divergent validity is demonstrated where expected between dimensions that would not be expected to be related. Change over time is demonstrated as expected for VILL-UI but is not clearly observed for EQ-5D-5L.

Differences in country preference weights for UK and Germany leads to differences in mean utilities. There are greater similarities between UK utilities than between German utilities, meaning smaller differences between VILL-UI and EQ-5D-5L at the mean level using UK weights than using German weights. This demonstrates the need to understand the implications of specific country preference weights as well as the performance of the dimensions that generate the utilities.

One limitation of VILL-UI is the large proportion of data regarded as missing/where a utility score cannot be generated (10.3% to 13.0% across different timepoints). The missingness mainly occurs (see supplementary materials Table A1) due to the response option “Don’t do this for other reasons” and for participants selecting this response for VILL-UI dimensions, no utility value can be generated. For any dataset this means that some participant data will not be able to be used to generate utilities, and it is more likely those participants will have comorbidities (as this may mean they are unable to do things for other reasons) meaning that their responses are not missing at random. This potential level of missing data should be considered in sample size calculations when collecting VILL/VILL-UI data, and in subsequent data analyses. It can be considered advantageous that the “Don’t do for other reasons” response option acts as a filter where utilities will only be generated for participants where the recorded impacts are due to their eyesight. Therefore even though it is not commonly observed for PROMs to include a response option “Do not do for other reasons”, it may be considered advantageous for ensuring impairment is due to the health condition being assessed, and this is commonly implemented in PROMs for eye conditions [38,39].

A single vision “bolt-on” dimension to EQ-5D has been explored [40-43] to improve the psychometric performance of EQ-5D-5L in vision. However, one additional general dimension on vision is unlikely to reflect all aspects of importance not included in EQ-5D (e.g. vision in dark surroundings). EQ-5D with a bolt-on will have different properties and

likely different preference weights, reducing comparability of results to other conditions where EQ-5D or EQ-5D with an alternative bolt-on is used, though this comparability is often the key argument for using a generic PWM.

Study limitations comprise that the MACUSTAR sample used here included the sample used to derive the VILL-UI classification system. However only the baseline sample was used to derive the VILL-UI classification, whereas the analyses undertaken here use three additional timepoints. For the known-group validity assessments of AMD stage the VILL-UI has much higher effect sizes for baseline than for the other timepoints, and one possibility is that this is because this was the data used to select the VILL-UI classification system based on psychometric performance. For the other known-group analyses, severity groups could only be defined at baseline. Therefore, repeating known-group analyses in another dataset would be beneficial. The study was unable to assess responsiveness to treatment as the data was from a non-interventional study, and assessment of responsiveness of VILL-UI and EQ-5D-5L to treatment in people with AMD is recommended.

The MACUSTAR dataset is a multi-country dataset, and in the analyses reported here the UK and German preference weights are applied to the entire dataset though this includes participant data collected outside of UK and Germany. This increases sample size, enables like-for-like comparisons of the application of the UK and German weights, and relies on the assumption that different language versions of the measures do not affect participant responses. This is often done in multi country datasets when used to inform economic evaluation, where preference weights for one country are applied to the entire sample. .

One factor that may have affected results is the different recall period of the measures. EQ-5D-5L has a recall period of ‘today’, whereas VILL-33 asks about your health during the ‘past month’.

The divergent validity between many EQ-5D-5L and VILL-UI dimensions, and low convergent validity in dimensions where a relationship would be expected, indicate that each measure captures different aspects. VILL-UI focuses on aspects of VRQoL that are impacted by AMD, to provide utilities that directly and better capture the impact of visual impairment, but cannot be used to capture generic HRQoL. In contrast, EQ-5D-5L covers general aspects of HRQoL and does not have a dimension related to vision, meaning HRQoL impact from visual impairment is captured via its impact on other dimensions. EQ-5D-5L has the benefit of comparability when used across different patient groups and treatments, and is able to capture the impact of comorbidities and potentially wider side effects from treatments. However EQ-5D-5L may underestimate treatment effects in people with AMD as it is not able to reflect known clinical effects.

The evidence provided here demonstrates that EQ-5D-5L is not fully appropriate for use in people with intermediate AMD, and details the impact on utilities (and hence QALYs) if VILL-UI was used instead of EQ-5D-5L. NICE [3], for example, allow the use of condition-specific PWMs when evidence shows that EQ-5D is not appropriate. The NICE methods guide [3] suggests that inappropriateness is demonstrated using evidence of lack of content validity, construct validity and responsiveness, using a synthesis of peer-reviewed literature. Qualitative work used to develop the VILL-33 identified aspects of importance to patients that are not included in EQ-5D (in particular, reading and accessing information) [22]. The evidence presented here shows that EQ-5D-5L performs poorly for convergent validity and known-group validity in AMD, indicating a lack of construct validity in AMD, and our study is the first to our knowledge to assess the psychometric performance of EQ-5D-5L in AMD.

The NICE methods guide [3] further suggests detailing the methods used to generate utilities from a condition-specific PWM, their validity and how the methods impact on utilities. The details of validity of VILL-UI and how VILL-UI utilities differ to EQ-5D-5L utilities, using both UK and German preference weights, have been provided here. We have demonstrated that the VILL-UI is appropriate for use in future AMD studies to inform economic evaluation. The results indicate VILL-UI is more appropriate for use for patients with intermediate AMD than EQ-5D-5L on the grounds of psychometric performance.

References

[1] Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *The European Journal of Health Economics* 21 2020:21(8);1245-1257.

[2] Brazier J, Ratcliffe J, Saloman J & Tsuchiya A (2016) Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press.

[3] National Institute for Health and Care Excellence. NICE, National Institute of Health and Care Excellence. NICE health technology evaluations: the manual. 2022, http://www.nice.org.uk/process/pmg36 NICE: London. Accessed 30th September 2024.

[4] Brooks R. EuroQol: the current state of play. *Health Policy.* 1996;37(1):53-72.

[5] Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 2011;20:1727–36.

[6] Rowen, D., Brazier, J., Ara, R. *et al.* The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. *PharmacoEconomics* **35** (Suppl 1), 33–41 (2017).

[7] Payakachat N, Ali MM, Tilford JM. Can The EQ-5D Detect Meaningful Change? A Systematic Review. Pharmacoeconomics. 2015 Nov;33(11):1137-54.

[8] Ling Wong W, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2014;2(2):e106-e116.

[9] Cruess AF, Zlateva G, Xu X, Soubrane G, Pauleikhoff D, Lotery A, et al. Economic burden of bilateral neovascular age-related macular degeneration: multi-country observational study. *PharmacoEconomics*;26.

10] Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. The European Journal of Health Economics 2018;19(4): 557-570.

[11] Tosh J, Brazier J, Evans P, Longworth L. A review of generic preference-based measures of health-related quality of life in visual disorders. Value in Health, 2012;15(1):118-27.

[12] Au Eong KG, Chan EW, Luo N, et al. Validity of EuroQOL-5D, time trade-off, and standard gamble for age-related macular degeneration in the Singapore population. *Eye*. 2012;26(3):379-88.

[13] De Sousa Peixoto R, Krstic L, Hill SCL. et al. Predicting quality of life in AMD patients—insights on the new NICE classification and on a bolt-on vision dimension for the EQ-5D. *Eye* 2021;35:3333–3341.

[14] Kim J, Kwak HW, Lee WK, Kim HK. Impact of photodynamic therapy on quality of life of patients with age-related macular degeneration in Korea. *Jpn J Ophthalmol* 2010;54:325–30.

[15] Lotery A, Xu X, Zlatava G, Loftus J. 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. *Br J Ophthalmol* 2007;91:1303–7

[16] Ruiz-Moreno JM, Coco RM, Garcia-Arumi J, et al. Burden of illness of bilateral neovascular age-related macular degeneration in Spain. *Curr Med Res Opin* 2008;24:2103–11.

[17] Soubrane G, Cruess A, Lotery A, et al. Burden and Health Care Resource Utilization in Neovascular Age-Related Macular Degeneration: Findings of a Multicountry Study. *Arch Ophthalmol*. 2007;125(9):1249–1254.

[18] Choi S, Park SM, Jee D. Utility values for age-related macular degeneration patients in Korea. PLoS One. 2018 Jul 31;13(7):e0201399.Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155-159.

[19] Cruess A, Zlateva G, Xu X, Rochon S. Burden of illness of neovascular age-related macular degeneration in Canada. *Can J Ophthalmol* 2007;42:836-843.

[20] Espallargues M, Czoski-Murray CJ, Bansback NJ, et al: The impact of age-related macular degeneration on health status utility values. *Invest Ophthalmol Vis Sci* 2005;46:4016-4023.

[21] Rowen D, Carlton J, Terheyden JH, et al. Development and Valuation of a Preference-Weighted Measure in Age-Related Macular Degeneration From the Vision Impairment in Low Luminance Questionnaire: A MACUSTAR Report. *Value in Health*, 2024:27(5);642-654.

[22] Pondorfer SG, Terheyden JH, Overhoff H, et al. Development of the Vision Impairment in Low Luminance Questionnaire. *Translational vision science & technology* 2021;10(1).

[23] Terheyden JH, Pondorfer SG, Behning C, et al; MACUSTAR consortium. Disease-specific assessment of Vision Impairment in Low Luminance in age-related macular degeneration - a MACUSTAR study report. *Br J Ophthalmol*. 2023;107(8):1144-1150.

[24] Terheyden JH, Mekschrat L, Ost RAD, et al. Interviewer Administration Corresponds to Self-Administration of the Vision Impairment in Low Luminance (VILL) Questionnaire. *Transl Vis Sci Technol*. 2022;11(4):21.

[25] Terheyden JH, Gerhards J, Ost RAD, et al. Patient-reported vision impairment in low luminance predicts multiple falls. *BMC Geriatr*. 2023;23(1):583.

[26] Hernández Alava M, Pudney S, Wailoo A. Estimating the Relationship Between EQ-5D-5L and EQ-5D-3L: Results from a UK Population Study. *PharmacoEconomics*. 2023;41(2):199-207.

[27] Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095-108.

[28] Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. *PharmacoEconomics* 2018;36:663-674.

[29] Finger RP, Schmitz-Valckenberg S, Schmid M, et al. MACUSTAR: Development and Clinical Validation of Functional, Structural, and Patient-Reported Endpoints in Intermediate Age-Related Macular Degeneration. *Ophthalmologica*. 2019;241:61–72.

[30] Terheyden JH, Holz FG, Schmitz-Valckenberg S, et al. (2020): Clinical study protocol for a low-interventional study in intermediate age-related macular degeneration developing novel clinical endpoints for interventional clinical trials with a regulatory and patient access intention—MACUSTAR. *Trials* 21.

[31] Jones R, Mulhern B, McGregor K, et al, on behalf of the Quality Of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team. Psychometric Performance of HRQoL Measures: An Australian Paediatric Multi-Instrument Comparison Study Protocol (P-MIC). *Children*. 2021;8(8):714.

[32] Mukuria C, Rowen D, Acaster S, et al. Psychometric performance of the CFQ-R-8D compared to the EQ-5D-3L and SF-6D in people with cystic fibrosis. *J Patient Rep Outcomes*. 2024;8(1):24.

[33] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293-307

[34] Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. 2007;60(1):34-42.

[35] Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 1977;*33*(1):159–174.

[36] Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–9.

[37] Dunbar HMP, Crabb D, Behning C, et al. Further evidence of heterogeneous visual function deficits in intermediate AMD. A MACUSTAR study report. *Invest. Ophthalmol. Vis. Sci*. 2023;64(8):2760.

[38] Mangione CM, Lee PP, Pitts J, et al. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Archives of ophthalmology* 1998;116(11):1496-504.

[39] Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. *Investigative ophthalmology & visual science* 2002;43(4):927-35.

[40] Gandhi M, Ang M, Teo K, et al. A vision 'bolt-on' increases the responsiveness of EQ-5D: preliminary evidence from a study of cataract surgery. The European Journal of Health Economics : HEPAC : Health Economics in Prevention and Care. 2020 Jun;21(4):501-511

[41] Luo N, Wang X, Ang M, Finkelstein EA, Aung T, Wong TY, Lamoureux E. A Vision "Bolt-On" Item Could Increase the Discriminatory Power of the EQ-5D Index Score. Value Health. 2015 Dec;18(8):1037-42

[42] Sampson, Christopher & Addo, Rebecca & Haywood, Philip & Herdman, Michael & Janssen, M.F. & Mulhern, Brendan & Page, Katie & Reardon, Oona & Rodes Sanchez, Marina & Schneider, Justine & Shah, Koonal & Thetford, Clare. (2019). Development of EQ-5D-5L bolt-ons for cognition and vision. Value in Health. 22.

[43] Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T & Longworth L (2015) An Exploratory Study to Test the Impact on Three “Bolt-On” Items to the EQ-5D. Value in Health, 18(1), 52-60.

Table 1: The MACUSTAR study sample, by timepoint

Characteristic	Level	Baseline, N=586		12 months N=416		24 months N=378		36 months N=367	
		n	%	n	%	n	%	n	%
Sex	Female	382	65.2%	263	63.2	248	65.6	246	67.0
	Male	204	34.8%	153	36.8	130	34.4	121	33.0
AMD stage	Early AMD	34	5.8	26	6.3	18	4.8	17	4.6
	Intermediate AMD	511	87.2	359	86.3	327	86.5	302	83.3
	Late AMD	41	7.0	31	7.5	33	8.7	48	13.1
BCVA (best-corrected visual acuity)	No impairment	439	74.9	346	83.2	314	83.1	309	84.2
	Impairment	105	17.9	70	16.8	64	16.9	58	15.8
	Missing	42	7.2	0	0	0	0	0	0
Visual function	No dysfunction	96	16.4	85	20.4	77	20.4	75	20.4
	Dysfunction	364	62.1	265	63.7	235	62.2	233	63.5
	Missing	126	21.5	66	15.9	66	17.5	59	16.1
Presence of late AMD in the fellow eye	No presence	505	86.2	389	93.5	351	92.9	340	92.6
	Presence	39	6.7	26	6.3	26	6.9	27	7.4
	Missing	42	7.2	1	0.2	1	0.3	0	0
Age	Mean (s.d.)	71.9 (6.92)		71.6 (6.90)		71.2 (6.97)		70.7 (6.89)	

Table 2: VILL-UI and EQ-5D-5L dimension responses and utilities, by timepoint

VILL-UI dimension	Level	Baseline, %	12 months, %	24 months, %	36 months, %
Information	1	37.5	32.2	28.0	26.7
	2	45.2	50.7	50.5	50.7
	3	14.5	16.1	20.4	20.7
	4	2.7	1.0	1.1	1.9
Reading	1	53.9	48.1	42.3	39.5
	2	34.0	41.8	41.5	45.8
	3	10.4	10.1	15.3	14.4
	4	1.7	0	0.8	0.3
Mobility and safety	1	62.1	61.8	57.4	51.5
	2	26.6	25.7	27.5	33.5
	3	2.7	4.1	4.5	3.3
	4	0.2	1.2	0.5	0.3
	5	5.5	5.5	7.4	6.0
	6	1.5	1.0	2.4	4.1
	7	0.2	0.0	0.0	0.0
	8	1.2	0.7	0.3	1.4
Emotional wellbeing	1	21.5	24.8	24.6	23.4
	2	41.6	44.5	46.6	45.5
	3	27.0	26.2	22.8	22.6
	4	9.9	4.6	6.1	8.5
EQ-5D-5L dimension					
Mobility	1	72.2	74.8	72.8	70.8
	2	16.6	14.9	16.4	17.2
	3	9.6	8.4	9.0	9.0
	4	1.7	1.9	1.9	3.0
	5	0.0	0.0	0.0	0.0
Self-care	1	95.4	96.2	97.6	96.5
	2	3.2	2.9	1.9	3.3
	3	0.9	0.7	0.3	0.3
	4	0.3	0.2	0.3	0
	5	0.2	0	0	0
Usual activities	1	82.8	86.1	85.7	82.0
	2	12.3	9.6	8.2	11.2
	3	3.8	3.6	5.0	6.0
	4	0.9	0.7	1.1	0.8
	5	0.3	0	0	0
Pain/discomfort	1	50.9	46.2	48.2	46.1
	2	28.8	35.8	32.5	31.9
	3	16.7	13.2	17.2	18.8
	4	3.1	4.8	2.1	3.3
	5	0.5	0	0	0
Anxiety/depression	1	68.6	70.7	71.2	66.2
	2	21.0	18.0	19.8	25.3
	3	8.4	10.1	7.1	6.5

VILL-UI dimension	Level	Baseline, %	12 months, %	24 months, %	36 months, %
	4	1.5	0.7	1.6	1.6
	5	0.5	0.5	0.3	0.3
VILL-UI, UK weights	Mean (SD)	0.804 (0.191)	0.810 (0.165)	0.783 (0.187)	0.770 (0.195)
VILL-UI, German weights	Mean (SD)	0.756 (0.236)	0.765 (0.211)	0.734 (0.240)	0.715 (0.249)
EQ-5D-5L, UK weights	Mean (SD)	0.836 (0.167)	0.836 (0.160)	0.840 (0.156)	0.828 (0.156)
EQ-5D-5L, German weights	Mean (SD)	0.909 (0.136)	0.910 (0.131)	0.919 (0.114)	0.908 (0.121)

Notes: For VILL-UI dimensions of Information and Reading: level 1= none; level 2 = a little; level 3 = a lot; level 4 = can't do. For VILL-UI dimension of Mobility and safety (mobility/safety items): level 1 = none/never; level 2 = a little/sometimes; level 3 = a little/often; level 4 = a little/always; level 5 = a lot/often; level 6 = a lot/always; level 7 = can't do/often; level 8 = can't do/always. For VILL-UI dimension of emotional wellbeing: level 1 = never; level 2 = sometimes; level 3 = often; level 4 = always. For EQ-5D-5L dimensions: level 1 = none; level 2 = slight; level 3 = moderate; level 4 = severe; level 5 = extreme/unable.

Table 3: VILL-UI dimension responses when patients report full health on EQ-5D-5L, by timepoint

VILL-UI dimension	Level	Baseline (n=217)		12 months (n=137)		24 months (n=135)		36 months (n=118)	
		n	%	n	%	n	%	n	%
Information	1	94	43.3	54	39.4	48	35.6	40	33.9
	2	99	45.6	72	52.6	70	51.9	56	47.5
	3	21	9.7	10	7.3	14	10.4	21	17.8
	4	3	1.4	1	0.7	3	2.2	1	0.9
Reading	1	138	63.6	78	56.9	74	54.8	57	48.3
	2	57	26.3	53	38.7	46	34.1	46	39.0
	3	18	8.3	6	4.4	14	10.4	15	12.7
	4	4	1.8	0	0	1	0.7	0	0
Mobility and safety	1	166	76.5	103	75.2	100	74.1	79	67.0
	2	40	18.4	24	17.5	28	20.7	29	24.6
	3	3	1.4	5	3.7	2	1.5	4	3.4
	4	1	0.5	2	1.5	0	0	0	0
	5	6	2.8	2	1.5	3	2.2	5	4.2
	6	0	0	1	0.7	2	1.5	0	0
	7	0	0	0	0	0	0	0	0
	8	1	0.5	0	0	0	0	1	0.9
Emotional wellbeing	1	60	27.7	40	29.2	45	33.3	38	32.2
	2	93	42.9	63	46.0	63	46.7	58	49.2
	3	51	23.5	31	22.6	22	16.3	15	12.7
	4	13	6.0	3	2.2	5	3.7	7	5.9

Notes: For VILL-UI dimensions of Information and Reading: level 1= none; level 2 = a little; level 3 = a lot; level 4 = can't do. For VILL-UI dimension of Mobility and safety (mobility/safety items): level 1 = none/never; level 2 = a little/sometimes; level 3 = a little/often; level 4 = a little/always; level 5 = a lot/often; level 6 = a lot/always; level 7 = can't do/often; level 8 = can't do/always. For VILL-UI dimension of emotional wellbeing: level 1 = never; level 2 = sometimes; level 3 = often; level 4 = always.

Table 4: Correlations between VILL-UI and EQ-5D-5L dimensions and utilities, reported at baseline

	VILL-UI Information	VILL-UI Reading	VILL-UI Mobility and safety	VILL-UI Emotional wellbeing	VILL-UI utilities, UK weights	VILL-UI utilities, German weights
EQ-5D-5L Mobility	0.12	0.11	0.26	0.09		
EQ-5D-5L Self-care	0.06	0.13	0.16	0.04		
EQ-5D-5L Usual activities	0.20	0.23	0.30	0.14		
EQ-5D-5L Pain/discomfort	0.14	0.11	0.30	0.12		
EQ-5D-5L Anxiety/ depression	0.04	0.09	0.11	0.22		
EQ-5D-5L utilities, UK weights					0.30	
EQ-5D-5L utilities, German weights						0.34

Notes: Pearson correlation is reported for dimensions and Spearman rank correlation is reported for utilities.

Table 5: Assessing known-group validity according to AMD stage, visual function, visual acuity and presence of late AMD in fellow eye, by timepoint

Timepoint	Measure	Early/Intermediate AMD (mean)	n	Late AMD (mean)	n	Accordance with clinical expectation	P-value	Effect size
Baseline	VILL-UI, UK weights	0.833	545	0.420	31	Yes	<0.001	2.600
	VILL-UI, German weights	0.790		0.304		Yes	<0.001	2.418
	EQ-5D-5L, UK weights	0.839		0.790		Yes	0.072	0.292
	EQ-5D-5L, German weights	0.913		0.865		Yes	0.029	0.354
12 months	VILL-UI, UK weights	0.812	401	0.763	15	Yes	0.259	0.297
	VILL-UI, German weights	0.767		0.715		Yes	0.349	0.247
	EQ-5D-5L, UK weights	0.835		0.855		No	0.647	-0.121
	EQ-5D-5L, German weights	0.910		0.917		No	0.826	-0.058
24 months	VILL-UI, UK weights	0.784	345	0.770	33	Yes	0.664	0.079
	VILL-UI, German weights	0.736		0.712		Yes	0.585	0.100
	EQ-5D-5L, UK weights	0.841		0.828		Yes	0.647	0.084
	EQ-5D-5L, German weights	0.919		0.912		Yes	0.746	0.059
36 months	VILL-UI, UK weights	0.785	319	0.670	48	Yes	<0.001	0.599
	VILL-UI, German weights	0.735		0.585		Yes	<0.001	0.612
	EQ-5D-5L, UK weights	0.830		0.815		Yes	0.544	0.094
	EQ-5D-5L, German	0.909		0.903		Yes	0.761	0.047

	weights							
		Visual function - No dysfunction (mean)	n	Visual function - Dysfunction (mean)	n			
Baseline	VILL-UI, UK weights	0.883	96	0.816	364	Yes	<0.001	0.439
	VILL-UI, German weights	0.854		0.766		Yes	<0.001	0.449
	EQ-5D-5L, UK weights	0.879		0.824		Yes	0.004	0.329
	EQ-5D-5L, German weights	0.938		0.902		Yes	0.023	0.262
		Visual acuity (BCVA) – No impairment (mean)	n	Visual acuity (BCVA) – Impairment (mean)	n			
Baseline	VILL-UI, UK weights	0.847	439	0.776	105	Yes	<0.001	0.484
	VILL-UI, German weights	0.808		0.715		Yes	<0.001	0.489
	EQ-5D-5L, UK weights	0.844		0.821		Yes	0.190	0.143
	EQ-5D-5L, German weights	0.916		0.898		Yes	0.198	0.140
		No presence of late AMD in fellow eye (mean)	n	Presence of late AMD in fellow eye (mean)	n			
Baseline	VILL-UI, UK weights	0.834	505	0.824	39	Yes	0.668	0.071
	VILL-UI, German weights	0.791		0.775		Yes	0.619	0.083
	EQ-5D-5L, UK weights	0.838		0.864		No	0.331	-0.162
	EQ-5D-5L, German weights	0.911		0.938		No	0.218	-0.205

Notes: Effect size is calculated using Cohen's D.

Figure 1: Distribution of EQ-5D-5L and VILL-UI utilities, at baseline