MEASURING AND VALUING GLAUCOMA HEALTH STATES

MEASURING AND VALUING GLAUCOMA HEALTH STATES

by

KEVIN M. KENNEDY, BA, MA

A Thesis submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy

McMaster University

McMaster University DOCTOR OF PHILOSOPHY (2022) Hamilton, Ontario (Health Research)

TITLE: Measuring and valuing glaucoma health states

AUTHOR: Kevin M. Kennedy

Bachelor of Arts (Carleton University)

Master of Arts (Laurentian University)

SUPERVISOR: Dr. Feng Xie

NO. PAGES: 90

LAY ABSTRACT

Cost-utility is a method used to support decision making for which treatments are available and how much they should cost. Utility is a value that reflects the severity of a health state and the preferences of a population for that health state. Patients with glaucoma lose their vision over time, resulting in worse health states and eventually blindness. It is difficult to measure the additional value of treatments in glaucoma because the usual measures of health state do not capture changes in health experienced by patients with glaucoma. Previous cost-utility studies use a variety of sources to estimate the utility of preferences. The objective of this dissertation was to describe the use and application of a preference-based glaucoma-specific measure of utility (HUG-5), validate the measure with glaucoma patients, and develop an algorithm that assigns preference weights to the measure. This work identifies common issues in cost-utility studies of glaucoma interventions, describes the properties of the HUG-5 in a sample of glaucoma patients and reports a preference-weighted algorithm for converting HUG-5 health states to utilities that describe the preferences of the US general population.

ABSTRACT

Cost-utility is a method that integrates clinical factors of disease progression with cost and consequence to patient's overall health state. Cost-utility analyses describe the trade-offs that are consequences of interventions and are commonly used to support health policy decision making. Well constructed cost-utility analyses can improve access patients have to effective interventions and move past ineffective or exceedingly costly interventions. The methods employed to assess incremental effectiveness of glaucoma interventions have advanced over time, where investigators have adapted to standards for reporting and modeling strategies. However, most cost-utility analyses use a restrictive set of utility values from foundational, but outdated work. Glaucoma cost-utility analyses combine utility values from patient preferences, on multiple scales of measurement. These methodological flaws in the use and reporting of utility values limit the internal validity of the study and external validity in generalizing the results to broader populations. Common generic preference-based measures of health utility, including the EuroQol 5 Dimension 3/5 Level (EQ-5D-3/5L), lack the sensitivity to detect health state changes in patients with glaucoma.

The three main investigations of this dissertation aimed to describe the current methodological issues in glaucoma cost-utility investigations, validate a condition-specific preference-based measure (HUG-5) in a sample of patients with various stages of glaucoma, and generate a preference-based weighting algorithm for glaucoma health states for US general population. A systematic review of cost-utility analyses in the literature revealed that most all cost-utility analyses failed to document rationale for including health utility estimates or conduct a review of appropriate utilities for relevant health states. Most authors of cost-utility analyses did not recognize or describe the limitations of combining utilities from multiple scales of measurement and from different preference sources. In a psychometric validation of the HUG-5, a sample of patients with glaucoma completed the NEI-VFQ-25, the HUG-5, and the EQ-5D-5L to describe their quality of life and health state. The HUG-5 dimensions described similar dimensions measured by National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) scales and differentiated from unrelated dimensions with adequate test-retest reliability on 3 month follow up. To model utility values to HUG-5 health states, a multi-attribute disutility function (MADUF) was estimated from the preferences of the US general population. A preference-weighting algorithm assigned utility values to the 3125 health states measured by the HUG-5. In validating the MADUF with directly elicited marker health states, the MADUF performed similar in mean absolute error relative to other related studies for the Health Utilities Index Mark 3 (HUI-3) and Patient-Oriented Prostate Utility Scale (PORPUS).

This work has significant implications for investigators evaluating glaucoma interventions with cost-utility comparisons, providing an alternative measure of utility on the policy scale with a condition-specific preference-based measure.

ACKNOWLEDGEMENTS

I will be forever grateful for the patients, clinicians, colleagues, and mentors for their commitment, insights, and experiences that allowed for this work to be completed.

To Feng, your guidance has fostered a well-rounded understanding of the methods and considerations to effectively elicit preferences for health states. Your willingness to explore, attempt new methods, and always give critical thought towards the implications of decisions has had a lasting impression on me. Principles of science and philosophy were rarely lost on you in our discussions. I will make a point of carrying these principles with me for the rest of my career. Beyond mentorship, I have valued you in my life as a friend. Some highlights for me include you attending the hamilton concert band with your family, inviting our small group for dinner with your family, and participating in my swim workouts with Blast Triathlon on weekends. Thank you for helping me to develop my skills and my character to be who I am today. To Jean-Eric, your friendly presence is infectious and always put a smile on my face. I will miss the drop-in visits, the interesting problems you would share that we could solve, and being the temporary in-house IT support/artwork hanger at the CRL. Thank you for helping me to learn how I can apply myself more to participate in the communities around me. The opportunities you have directly facilitated for me continue to grow. To Simon, since meeting you at the Euroqol annual meeting in my first year, I can see why Feng has such high praise for your character, your work, and good nature. I will always remember the shared experiences in Portugal and New Orleans. Thank you for stepping up and joining my committee late in my PhD process. Your comments have always help me identify glaring gaps in what I had thought I knew and how I can improve my reporting of the importance and meaning of the findings.

To the patients at PRISM Eye Institute who have participated in my studies, I have learned much about how patients face and adapt to changes in vision. I have developed a strong sense of humility as a consequence of your deep reflections. Without your willingness to share, I would not have found such purpose in my work. From your stories, I have developed a rich understanding of the impact that access to interventions has on your daily living and the quality of your lives. To Dr. Ahmed and Dr. Podbielski, thank you for your commitment and ongoing engagement with my work and making collecting data at PRISM a very positive experience.

To my friends in the HRM program, thank you for all of the memories, particularly the HRMSA events, workshops, and meetings. It has been an honour getting to know you and I place a high value on the friendships we have developed over the years. I look forward to witnessing the heights I am confident you will achieve. Special recognition to Fabrice, Dena, Lyubov, and Emily. I would not been able to pursue my dream of research without the continued support of my family, my fiance Shelby, and my friends. Thank you for believing in me.

For the memory of my father, Frank E. Kennedy

Table of Contents

PREFACE

This doctoral dissertation is a complete summary of the research conducted and prepared for the fulfillment of the Doctor of Philosophy degree at McMaster University. Three original stand-alone articles were prepared for publication in clinical and health economics journals, therefore this is some overlap in their content. This dissertation is presented as a 'sandwich thesis', where 3 research articles are preceded with an overview introduction chapter and proceeded by a conclusive chapter, which highlights the main conclusions, lessons learned and discusses future topics of investigation. The three articles are organized to first, present an overview of cost-utility studies in glaucoma, with particular attention to health state utility values. Second, the article assesses the psychometric validity of the HUG-5 [1] measure in a sample of patients with glaucoma, describing the validity and reliability in contrast with other condition-specific and generic measures of health status. Third, the article defines a value function for the HUG-5 on the scale of perfect/full health to death from a representative sample of the US population.

The chapters contained herein reflect the work of a team of researchers with diverse training backgrounds in clinical ophthalmology, health research methods, and health economics. Dr. Feng Xie is the senior author on all 3 papers. His contributions included developing the research plan, revisions to protocols and manuscripts as well as critical guidance on research methods, data collection and data analysis techniques. Dr. Jean-Eric Tarride and Dr. Simon Pickard served as committee members throughout the course of this research. Both provided substantial feedback on Chapters 3 and 4. Chapter 2 was published prior to establishing the thesis committee. Dr. Ike Ahmed and Dr. Dominik Podbielski are ophthalmologists with specialty training in glaucoma. In addition to facilitating access to patients, both contributed to the design and review of the methodological review and psychometric validation of the HUG-5 (Chapter 2 and 3). Dr. Keean Nanji and Dr. Gurkaran Sarohia are currently ophthalmology residents. At the time of data collection, both clinicians were attending medical school and contributed to the data collection and provided critical review of manuscripts. Dr. Nanji made substantial contributions to Chapter 3 and Dr. Sarohia assisted with Chapter 2. Dr. Sergei Muratov outlined much of what is contained in this dissertation in the mixed-methods development of the HUG-5, and made substantial contributions in the design and critical review of Chapter 3. As primary author, I designed and wrote each publication independently before circulating for revisions. I reviewed and consolidated the literature from Chapter 2. I directly collected patient data over several months for Chapter 3, I developed and tested the software used to collect study data for Chapter 4 and managed the general population participants. I was responsible for all study ethics applications, data analysis, creation of tables and figures describing study results. Chapters 3 and 4 were

[1] The HUG-5 stands for Health Utility for Glaucoma - 5 Dimensions and is a preference-based disease-specific measure of health utility. The measure dimensions include: 1) visual discomfort (physiological and psychological), 2) mobility concerns (limitations to spatial awareness), 3) daily life activities (close up work, household chores), 4) emotional health (negative thoughts, worries and fears), and 5) social activities (getting out with friends, large social gatherings, incidents of social embarrassment).

possible due to funding from the Glaucoma Research Society of Canada. The three research papers included in this dissertation include:

Chapter 2: Review Kennedy, K. & Sarohia, G. & Podbielski, D. & Pickard, A.S. & Tarride, J.E. & Ahmed, IIK. & Xie, F.(2022) Systematic methodological review of health state values in glaucoma cost-utility analyses. Submitted to: *JAMA Ophthalmology*

Chapter 3: Validation Kennedy, K. & Podbielski, D. & Ahmed, IIK. & Xie, F. (2019). Disease-specific preference-based measure of glaucoma health states: HUG-5 psychometric validation. Published in: *Journal of Glaucoma*.

Chapter 4: Valuation Kennedy, K & Pickard, S & Tarride, J.E. & Xie, F. (2022). Resurrecting muti-attribute utility function: developing a value set for Health Utility for Glaucoma-5 (HUG-5) for the US. Submitted to: *Value in Health*

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The three investigations contained within this dissertation aim to describe the current use of health utility values in cost-utility literature of glaucoma interventions, validate an alternative for estimating health utilities for glaucoma health states with the condition-specific, preference-based measure, the Health Utility for Glaucoma (HUG-5), and assign a preference algorithm for HUG-5 health states from the US general population.

1.1 Background

Policy decisions guiding health care resource allocation operate under the constraint of scarcity. Treatment alternatives are considered and selected to maximize the value of limited resources.[1] Health technology assessment (HTA) is a method of integrating clinical, ethical, social, legal, and economic information to inform these policy decisions.[2,3] Health policy decisions often are made in a context that requires comparison between treatments with multiple outcomes, or for prioritizing therapies for disease states that are not directly comparable.[4] Cost-utility analysis is an economic assessment of the costs and benefits of implementing interventions over time. Health utility is used in the estimation of quality-adjusted life years (QALY). In cost-utility studies, QALY estimates are the measure of effect and applied as a difference value between intervention and control groups in the denominator of the incremental cost-utility ratio (ICUR). ICUR (cost of intervention B – cost of intervention A / QALY of intervention B – QALY of intervention A) are the primary outcomes of cost-utility studies. ICUR values are affected by the costs, model structure, and utility values chosen. The accepted policy scale to support decisions that consider treatments across diseases and interventions measures utility values anchored at 1 for perfect or full health and 0 for dead. Utility is weighted by preferences of a population. Regulatory bodies that use cost-utility analyses to inform their decisions recommend utility values derived from the preferences of the general population of which they are responsible for, or at least a comparable population.

Glaucoma is a progressive condition, where patients experience peripheral to central vision loss over time. Glaucoma can go undetected for years until the patient experiences changes in their vision. The most prevalent type of glaucoma is primary open-angle glaucoma (POAG). POAG is a progressive optic neuropathy in which the drainage angle for fluid in the eye remains open. Elevated intraocular pressure (IOP) is a risk factor for progression in POAG. Less common types include closed-angle glaucoma (ACG) and normal tension glaucoma (NTG). ACG can present suddenly or gradually, where as NTG only presents gradually and is not associated with elevated IOP. When glaucoma is detected, treatment is preventative and aims to slow disease progression. Interventions aim to delay and prevent severe damage to patient vision. Current treatments focus on reducing IOP. Over the course of a patient's life after diagnosed with glaucoma, they often receive multiple combinations of topical eye drops, laser surgical interventions, and in cases of uncontrolled IOP, incisional surgeries.[5] Ophthalmologists use objective measurements such as cup-to-disc ratio, corneal thickness, age, previous progression, and clinical judgment to determine how frequently individual patients should be monitored.[6] Each eye with glaucoma is often classified by Hodapp-Parrish-Anderson (HPA) stages of glaucoma, and grouped into the mutually exclusive categories of OHT (ocular hypertension, pre-glaucoma), mild glaucoma (0db to -6db), moderate glaucoma (-6db to -12db) advanced glaucoma (-12db to -20db), blindness in one eye, and blindness in both eyes.[7]

As glaucoma is a progressive condition, much of the potential damage can be avoided through early detection and intervention. Governments and health care authorities apply targeted screening programs to identify people with increased risk of developing glaucoma.[8] Direct costs to patients with glaucoma are expensive, including hospital visits, glaucoma related procedures, and IOP lowering medications. The recurring costs and loss of productivity due to limited vision have a pronounced impact on the lives of patients and consequently on societies. After a patient is diagnosed with glaucoma, they are often followed for the rest of their life at 3, 6, or 12-month intervals depending on their risk of progression. In the evaluation of emerging glaucoma interventions, CADTH, a Canadian HTA authority has cited a high degree of uncertainty of the economic evidence supporting reimbursement of micro-invasive glaucoma surgeries.[9] Schmier et al. described that glaucoma economic evaluations are not comparable.[10]

Health utilities are measured alongside clinical trials or observational studies using indirect and direct measures. Indirect health state utilities are measured from generic (e.g., EQ-5D-3L/5L, HUI-3) or condition-specific (e.g. for glaucoma, GUI, HUG-5) preference-based measures of health state. Direct health utilities are measured from patients, whom respond to stated preference exercises (e.g., Standard Gamble [SG], Time Trade-Off (TTO), Discrete Choice Experiments [DCE], and Visual Analog Scale [VAS]). Direct patient preferences have been associated with greater utility estimates for health states, relative to the utility assigned by the general population.[11,12] Generic measures are often favoured by HTA agencies due to the use of a general health sate which facilitate comparisons across conditions to support broad HTA decision processes. However, generic measures have been demonstrated to lack sensitivity in measuring health states affected by changes to sensory systems (such as hearing or vision). Condition-specific preference-based measures (CS-PBMs) have been proposed as an alternative to generic measures that have lacked sufficient sensitivity or responsiveness for patients with disease states.[13,14]

Both indirect and direct methods of estimating utilities for health state are used to generate utilities or values that are applied in cost-utility analyses of glaucoma interventions. The NEI-VFQ-25 is a measure of quality of life related to visual function. The NEI-VFQ-25 has a utility mapped function described as the VFQ-UI. However, the VFQ-UI does not finely discriminate between levels of visual impairment among glaucoma patients. Ongoing development of the EQ-5D has seen the emergency of bolt-on attributes that allow for an additional dimension of health state to be measured. However, in reducing glaucoma to a single dimension of vision does not effectively capture the attributes of living with glaucoma important to patients. Glaucoma is unique, as it may not affect central vision until end-stage and requires ongoing management to prevent further vision loss. In measuring glaucoma health states, directly elicited utilities are described as elicited on multiple different scales of measurement.[15] Further, utilities and values from multiple sources are combined as best-estimates for health states. The HUG-5 was developed to reflect the experiences of patients living with glaucoma and to be used as a condition-specific preference-based measure of health state utility to facilitate cost-utility analyses with utilities for glaucoma health states on the policy scale of perfect/full health to death. The HUG-5 is a preference-based descriptive system that describes glaucoma health states. The measure was developed with a mixed-methods approach. Initially, items were identified by a systematic literature then compiled into an interview guide. Patients with glaucoma were sampled with a maximum variation strategy to capture information as to what living with glaucoma means to patients with different durations of illness, number of eye drops, surgical interventions, and severity. The feedback from patients was elicited through face-to-face semi-structured interviews. After consolidated the information into themes, identifying attributes, and descriptions, 3 versions of the draft questionnaire were pilot tested with a separate, small group of patients.[16] The next steps of measure development include assessing the psychometric properties of the HUG-5 and defining a preference-weighted utility algorithm on the scale of Perfect Health to Death.

1.2 Chapter Objectives and Contents

Chapter 2 is a manuscript titled 'SYSTEMATIC METHODOLOGICAL REVIEW OF HEALTH STATE VALUES IN GLAUCOMA COST-UTILITY ANALYSES' prepared for submission to JAMA: Ophthalmology. In brief, utilities and values used in investigations of cost-utility analyses of glaucoma interventions have significant impact on the conclusions of incremental value, ultimately affecting policy decisions which can impact uptake and access to emerging technologies. This chapter systematically reviews the published literature to describe how clinicians and health economists have previously evaluated the incremental value of glaucoma interventions and their approach in determining which utilities are used for modeling health outcomes over time.

The goals of this chapter are to:

1) Summarize the cost-utility literature assessing the incremental value of glaucoma interventions
2) Describe the range of utility and values used for glaucoma health states across different elicitation scales and preference sources
3) Review the reporting quality of the cost-utility literature in their use of health state utilities and values

Chapter 3 is a manuscript previously published in a 2019 issue of the Journal of Glaucoma titled 'DISEASE-SPECIFIC PREFERENCE-BASED MEASURE OF GLAUCOMA HEALTH STATES: HUG-5 PSYCHOMETRIC VALIDATION'. Building on the previous work of Dr. Muratov, the HUG-5 is contrasted with a condition-specific measure of health-related quality of life and a generic preference-based measure to describe the psychometric properties for patients with glaucoma. The patients were sampled from a southern Ontario eye-clinic and supported by ophthalmologist researchers of PRISM Eye Institute.

The goal of this chapter is to describe the validity, reliability, and sensitivity as evidence for use of the HUG-5 as an instrument of measuring the health states of patients with glaucoma.

Chapter 4 is a manuscript submitted to Value in Health titled 'RESURRECTING MULTI-ATTRIBUTE UTILITY FUNCTION; DEVELOPING A VALUE SET FOR HEALTH UTILITY FOR GLAUCOMA (HUG-5)'. For the HUG-5 to meet the intended use of a 'preference-based' measure, a scoring algorithm that assigns utilities to HUG-5 health states is required. The scoring algorithm was generated from a multi-attribute disutility method and describes at 3125 HUG-5 states. The tasks within this study required respondents to trade off between risky outcomes and certain outcomes, similar to how people make real health care decisions. This study defines the first glaucoma-specific preference-based measure weighted by the preferences of the US population on the policy scale of perfect/full health and death.

The goal of this chapter is to generate a preference-weighted scoring algorithm for HUG-5 health states with the multi-attribute disutility function method.

2 SYSTEMATIC METHODOLOGICAL REVIEW OF HEALTH STATE VALUES IN GLAUCOMA COST-UTILITY ANALYSES

Prepared for submission as: Kennedy, K., & Sarohia G.S., & Pickard, A.S., & Tarride, J.E., & Xie, F. (2022). *JAMA: Ophthalmology.* (To be Submitted July 2022)

2.1 Foreward

This first paper was conceptualized to understand how utility values are identified, described, and included in cost-utility studies of glaucoma interventions. The purpose of this is to understand the potential impact of the HUG-5 as a condition-specific preference-based measure to set a standard of health utility values for glaucoma health states.

In this systematic review, we identify and describe 43 cost-utility studies of glaucoma interventions published from 2005 to 2022 that used Markov state transitions, micro simulations, discrete event simulations and decision tree models to determine long term costs and utilities of pharmacological interventions, screening, and monitoring programs. We found that few cost-utility studies of glaucoma interventions describe key characteristics of the source of utilities and their relevance to the decision context the cost-utility study informs. These studies often combined estimates from multiple sources to inform economic models with utilities that are elicited on different scales or applied simplified linear transformations of clinical features to utility values. This review describes that few cost-utility studies describe important rationale for using health state utilities. Including additional details on the search, appraisal, selection, and inclusion process of health utilities improves transparency, generalizability and supports the assessment of the validity of study conclusions. Future investigations should aim to use health utilities on the same scale of measurement across health states and consider the source and relevance to the decision context/purpose of conducting that cost-utility study.

Overall, we found that few glaucoma cost-utility studies adhere to reporting standards for use and selection of health state utilities. Authors of future CUA are directed to ISPOR resources for reporting standards and approaches. Within CUA, reviews of available utilities for health states are important to justify their use. This review has identified a growing need for reliable estimates of health state utility on the policy scale of perfect health and death to support economic comparisons of glaucoma interventions.

2.2 Introduction

Health care resources are finite and should be allocated efficiently to maximize the value to the population. Cost-utility analysis (CUA) is a framework to measure this incremental value between alternative healthcare strategies. CUA may apply decision analytic models to synthesize available evidence on costs and effects. A recommended outcome for the effect measure in CUA is quality adjusted life-years (QALY). QALY combines quantity and quality of life with the quality weight often derived from preference-weighted health state utilities.

While several guidelines have been published to improve the methods and reporting of economic evaluations, ophthalmologists have reported a desire for specific guidelines regarding the economic assessments specifically for glaucoma.[17] Glaucoma is a collection of optic neuropathies with a global prevalence of 2.4%, affecting 68 million people over 40 in 2021.[18,19] There is a growing need for effective, long-term therapies that reduce burden on patients and health care systems. Due to the complex nature of iterative treatment to minimize vision loss, there is a current need for high quality evidence to provide patients with reliable, effective, and affordable care.[20] Previous studies have aggregated evidence to describe the available literature measuring the cost-utility of glaucoma screening programs[21,22], primary open-angle glaucoma progression[23], ophthalmic medications[24–26], and micro-invasive glaucoma surgeries.[27] While each of these reviews address important policy decisions of resource allocation, none have explored the use and reporting of health utilities and preference elicitation methods. This is important as a recent study has identified concerns with reporting quality of health utilities in published CUA (e.g. including health utilities from different samples, estimation methods and preference weights) which may lead to unreliable evidence.[28] Previous reviews of glaucoma economic evaluations have described problematic use of vision-scale health utilities (perfect vision to blindness) and incorporating both direct patient health utilities with indirect health utilities from generic preference-based measures. One original study evaluated health utilities across 5 common ophthalmic disease, describing differences between scales of measurement, direct and indirect measures.[15] Lee et al identified that early measures of health utilities in glaucoma were measured with on a scale of perfect vision and blindness, describing systematic differences and highlighted that health utilities of vision-related health states should be measured on the full/perfect health to death scale. A previous review of economic studies evaluating ophthalmic drugs with a focus on health utilities found that choosing different techniques to derive health utilities will affect the results and conclusions of a CUA, inhibiting comparability between CUA.[29]

This methodological review describes the state of practice among published glaucoma CUA studies. The primary objective was to describe the characteristics of glaucoma CUA by focusing on the sources and range of health utilities and the quality of reporting.

2.3 Methods

A systematic search strategy was developed to identify CUA of glaucoma screening, treatment, or monitoring interventions (SUPPLEMENT 1). Search terms were generated from published glaucoma systematic reviews with economic evaluation terms. The search strategy was reviewed by an ophthalmologist with specialized training in glaucoma and a staff librarian with training in developing systematic review strategies. Medline OVID, Cumulated Index to Nursing and Allied Health Literature (CINHAL), Embase OVID, Web of Science, Scopus, Biosis previews, the Health Economic Evaluations Database, and the NHS Economic Evaluation Database (NHS EED) were searched for articles meeting inclusion criteria from 1989 to May 24, 2022.

Review was restricted to include articles written in English. Titles, abstracts and subsequently full texts were screened by GS and KK to identify published CUAs with QALYs as the outcome measure. Excluded studies were non-English language, reviews, editorials, protocols, or other types of economic studies (e.g., cost-benefit, cost-minimization, cost-effectiveness). References of included articles were also reviewed. A data extraction form was developed in Microsoft Excel (Version 16.52) to extract study characteristics of CUA, including publication year, country, intervention, comparator, and population. Study perspective, economic model type, time horizon, and discount rate were also extracted to describe the characteristics of the study design (SUPPLEMENT 2). Base case and sensitivity analyses, results were extracted to summarize the methods of accounting for heterogeneity of CUA results. Methods of describing glaucoma health states and corresponding utilities were extracted. The original source of the utility value was identified and reviewed to extract author and year, the description of health staes, the sample size, the method of measuring preferences, and the health utilities.

Checklist items from the Systematic Review of Utilities for Cost-Effectiveness (SpRUCE checklist)[28] were used to assess the reporting and quality of health utilities in glaucoma CUA. Briefly, SpRUCE items describe the process by which health utilities are identified and any quality criterion applied for inclusion and to assess if health utilities selection for the CUA model was transparent and appropriate.

CUA characteristics were reported as a range of years, n (%) for country, with summary descriptions of interventions and comparators. The patient population of study described the target age and demographics. The study perspective includes either societal or payer perspectives. Sensitivity analyses were described as deterministic or probabilistic. The type of economic model was assigned as Markov state transition, decision analytic, discrete event simulation, or microsimulation. Time horizons and discount rates were described as a range and mode. ICUR values were described as minimum and maximum values, in cases where specific ICUR values were not reported, the values were calculated relative to the next most-effective intervention strategy in cases where multiple strategies were evaluated. The glaucoma health states and method of assigning utilities to health states to the CUA were described. The unique health utility values that described HPA states across CUA were summarized with plots grouped by the scale of measurement and the source of preferences with point size reflecting the sample size. Health utilities that did not describe HPA health states were described separately. Selected SpRUCE checklist items were reported as a frequency n (%). Rates of SpRUCE items checked were summarized in a figure illustrating the proportions. The count of total SpRUCE items were summed and reported as a range and mode and illustrated with a histogram.

2.4 Results

2298 articles were identified by the search terms. 403 duplicates were excluded from the review, resulting in 1895 records that were screened by title and abstract. 198 records met the criteria for full text review. 155 full text articles were excluded for non-QALY outcome (n=118), conference abstract (n=19), other language (n=8), and review article (n=10). A total of 43 CUA studies met inclusion criteria and were included in the review (Figure 1).

These studies were published between 2005 and 2022 with 9 from the United States (21%), 9 from Great Britain (21%), and 3 from Canada (7%). Interventions included early clear lens extraction, primary selective laser trabeculoplasty, phacoemulsification (for patients with cataract and glaucoma), increasing frequency of monitoring, changing medication ordering, preventative treatment, combinations of medications and SLT, and population/community targeted screening programs. The standard of care was frequently used as the comparator, however the definition of standard care varied between studies. The target patient population of study reported often reflected patients older than 40 years old with glaucoma in one or both eyes. 24 (56%) studies reported the societal perspective. 32 (74%) used a Markov state transition model, and 5 (14%) microsimulation or discrete event simulation model. The time horizon specified ranged from a minimum of 3 years to lifetime with 34 (79%) using a time horizon of 10 years or longer. A discount rate was applied in 40 studies (93%) and ranged from 2% to 5% with the most common rate of 3% applied in 20 studies (50%). All studies conducted sensitivity analyses, where 26 (61%) applied probabilistic sensitivity analysis. ICUR values ranged from reporting that the intervention increased QALY with savings of $42,769 USD to increased benefit with an increased cost with ICUR of $89,072 USD (Table 1).

2.4.1 Glaucoma health state

Glaucoma health states were often grouped by Hodapp-Parrish-Anderson (HPA) stages of glaucoma. HPA staging consists of six stages: OHT (ocular hypertension, pre-glaucoma), mild (0db to -6db), moderate (-6db to -12db) advanced (-12db to -20db), blindness in one eye, and blindness in both eyes.[7] Visual acuity from 20/20 to 20/800 health states were other glaucoma health states. Additionally, authors have included health states based on treatment status such as the type of medication or intervention assigned in a clinical trial (Table 2).

2.4.2 Measures of health utility

11 unique sources of health utilities were referenced across the 43 CUA studies, with 59 values that described HPA glaucoma states[15,30–38]. 31 of the 59 values (52%) reflected preference scoring from the general population on the perfect health to death scale, of which 23 were measured using the EQ-5D-3L and 8 were measured by the HUI-3. The remaining 28 (48%) utilities were elicited from patient preferences, with 19 Time-Trade-Off, 8 Standard Gamble, and 1 Visual Analog Scale. 12 of the 28 direct patient health utilities reported a sample size that ranged from 32 to 365. 8 patient preference utilities were elicited

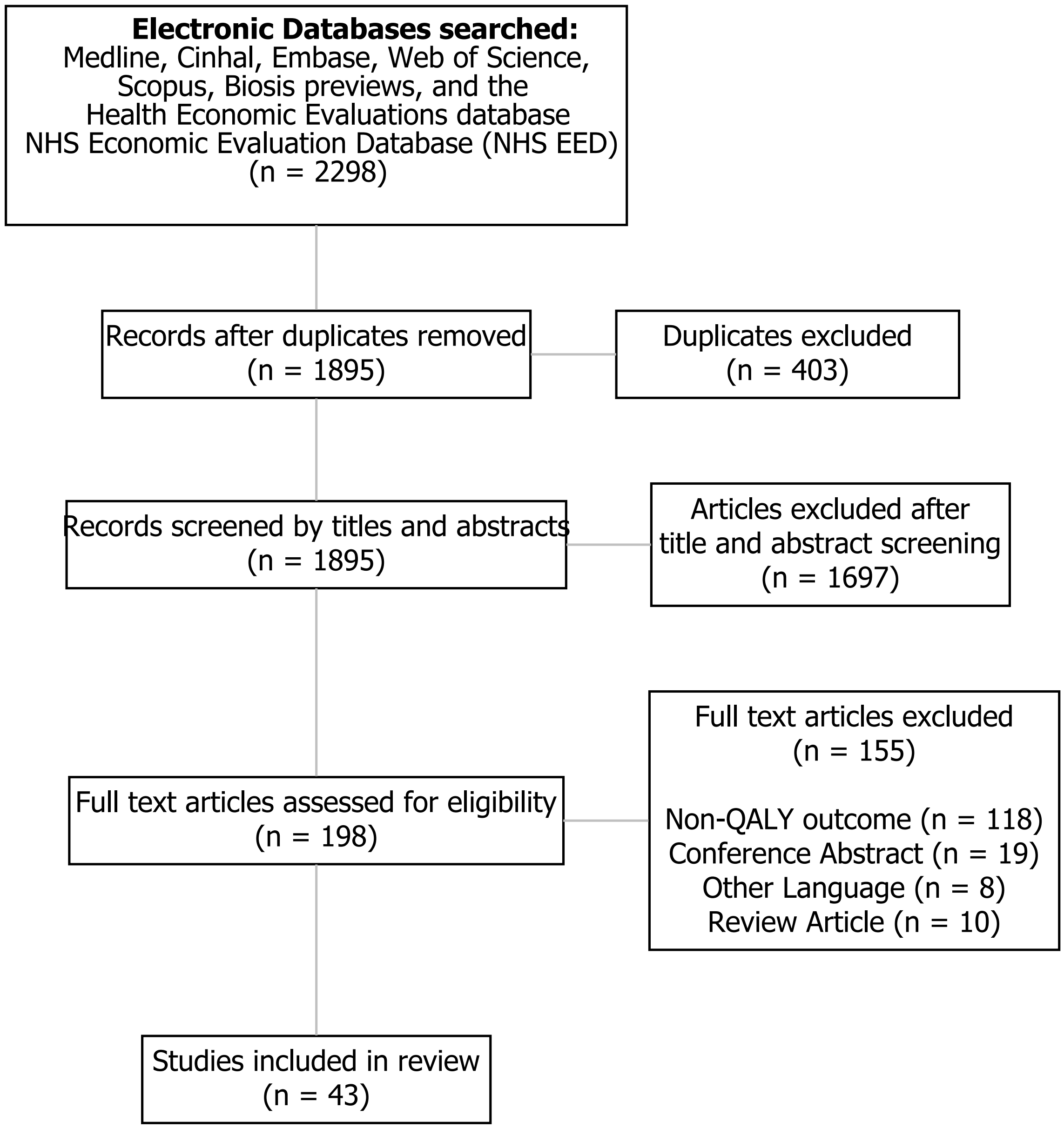


Figure 1: PRISMA flow diagram.

Table 1: Study characteristics of publications included in the review.

Study	Year	Country	Intervention	Comparator	Study Population	Perspective	Economic Model	Time Horizon (years)	Discounting	Sensitivity Analysis	ICUR	Utility Method	Utility Source Author (Year)
Ahmed et al.[166]	2019	CAN	iStent trabecular bypass stent during Phacoemulsification	Phacoemulsification	OAG with visually significant cataract	Societal	Markov	15	2%	Probabilistic	-$ 16,691.00	EQ-5D-5L	Burr et al.[31]
Barlelt-Hofer et al.[172]	2020	GER	Phacoemulsification and one MIGS or Phacoemulsification and two MIGS or Intracanalicular scaffold and Phacoemulsification	Phacoemulsification alone	POAG	Payer	Markov	Lifetime	3%	Probabilistic	€ 10,955.00	EQ-5D-3L	Wolfram et al.[171]
Blumberg et al.[147]	2014	USA	Spectral Domain (SD)-OCT screening	Standard of care	General population	Societal	Markov	10	3%	Deterministic	$ 46,416.00	Direct elicitation	-
Boodhna et al.[49]	2016	GBR	3 VFT in first year, 3 VFT Second Year, then 1 a year	Standard of care (1 VFT/year)	OHT	Societal	Markov	25	-	Probabilistic	£ 21,392.00	HUI-3	Van Gestel et al.[104]
Brown et al.[159]	2019	USA	Bimatoprost or Latanoprost or Tafluprost or Travoprost or Timolol	Observation	POAG	Societal	Decision Analytic	20	3%	Deterministic	$ 88,836.92	HUI-3	Wolfram et al.[171]
Burr et al.[179]	2007	GBR	GO or Technician	Opportunistic case finding	General population	Payer	Markov	Lifetime	4%	Probabilistic	£ 20,571.00	QALYs = 0.98991 + 0.0022 dBs - 0.00080518 dBs^2	Rein et al.[35], Brown et al.[30]
Burr et al.[145]	2012	GBR	SOH hospital or SOH primary care or NICE conservative or NICE intensive	Standard of care	General population	Societal	Discrete Event Simulation	20	4%	Deterministic	£ 85,312.00	EQ-5D-3L	Burr et al.[31]
Burr et al.[142]	2014	GBR	Glaucoma Screening Platform Study or GPS 1, instead of ONP, use VFT or (IOP + ONP) – optometrist or (IOP + VFT) – optometrist	Standard of care	General population	Payer	Markov	10	4%	Probabilistic	£ 68,718.00	Direct elicitation	Brown et al.30
Choi et al.[165]	2019	KOR	Trabeculectomy or Laser trabeculoplasty	Medication therapy	POAG	Societal	Markov	25	3%	Probabilistic	KRW 30,885,179.00	EQ-5D-3L	Personal communication, Gupta et al.[79]
Crane et al.[174]	2013	AUS	4 month VFT + 1 year or 4 month VFT, laser trabeculoplasty as line 1 + 70/84 days to review or 6 month laser trabeculoplasty first, medication, laser if none previous or surgery	Standard of care	POAG	Payer	Discrete Event Simulation	5	-	Probabilistic	$ 19,866.00	EQ-5D-3L	Personal communication, Gupta et al.[79]
Gazzard et al.[156]	2019	GBR	Primary SLT followed by topical medications as required.	Standard of care (Medication first treatment)	POAG	Societal	Markov	3	-	Probabilistic	$ 20,000.00	EQ-5D-3L	Personal communication, Gupta et al.[79]
Guedes et al.[151]	2016	BRA	Medical therapy or SLT	Observation	POAG	Payer	Markov	Lifetime	5%	Deterministic	$ 2,811.39	EQ-5D-3L	Choi et al.[32], Lee et al.[15], Brown et al.[30]
Hernandez et al.[21]	2008	GBR	Screening for OAG by Optometrist or Technician	Standard of care	POAG	Societal	Markov	Lifetime	5%	Probabilistic	£ 65,924.00	QALYs = 0.98991 + 0.0022 dBs - 0.00080518 dBs^2	Rein et al.[35]
Hernandez et al.[155]	2016	GBR	Biennial monitoring (secondary care) or Bienniel monitoring (primary care) or NICE conservative or NICE intensive	Treat all (annual IOP only)	OHT	Societal	Discrete Event Simulation	20	4%	Deterministic	£ 85,312.00	Direct elicitation	Lee et al.[15], Brown et al.[30]
Javanbakht et al.[146]	2017	GBR	Lens extraction	Standard of care - Laser iridotomy followed by medical therapy and glaucoma surgery	PACG	Societal	Markov	5	4%	Probabilistic	£ 15,223.00	EQ-5D-3L	Burr et al.[31], Gupta et al.[79]
John et al.[180]	2017	IND	Population-based community screening and subsequent treatment program	Opportunistic case finding	General population	Societal	Decision Analytic	10	3%	Deterministic	₹ 7,292.30	EQ-5D-3L	Burr et al.[31], Gupta et al.[79]
John et al.[167]	2018	IND	Population-based community screening and subsequent treatment program	Opportunistic case finding	POAG and/or PACG	Societal	Decision Analytic	10	3%	Deterministic	₹ 7,292.30	Direct elicitation	Gupta et al.[79]
Kaplan et al.[152]	2015	USA	Surgical trabeculectomy, baedvert implant	Routine maximal medication therapy	POAG	Societal	Markov	5	3%	Deterministic	$ 8,289.00	Direct elicitation	Gupta et al.[79]
Kymes et al.[148]	2010	USA	Treat patients with a 5%, 4%, 3%, 2%, or greater annual risk of developing glaucoma	Treat no one until evidence of glaucoma-related nerve damage	OHT	Societal	Markov	Lifetime	3%	Deterministic	$ 50,000.00	QALYs = 0.98991 + 0.0022 dBs - 0.00080518 dBs^2	Lee et al.[15], Rein et al.[35], Brown et al.[30]
Li et al.[173]	2013	USA	Treatment (Medical: Timolol or Latanoprost; Surgical: Laser trabeculoplasty or Trabeculectomy)	Observation	NTG	Payer	Markov	10	3%	Probabilistic	$ 34,225.00	Direct elicitation	Jampel et al.[33], Brown et al.[30], Lee et al.[15]
Newman-Casey et al.[154]	2019	USA	Improved adherence to medication regime	Standard of care	POAG	Payer	Markov	Lifetime	3%	Probabilistic	$ 29,600.00	EQ-5D-3L	Kobelt et al.[34]
Nordmann et al.[175]	2009	FRA	Line 1 increase 50% or Line 2 increase 50% or Switch 1st line to 2nd line	Standard of care	OHT or POAG	Societal	Markov	Lifetime	3%	Deterministic	-€ 2,083.00	Direct elicitation	Brown et al.[30], Lee et al.[15]
Ordonez et al.[160]	2019	COL	Trabecular micro-bypass stents + Timolol or Laser trabeculoplasty + Timolol + Dorzolamide or Latanoprost + Timolol + Dorzolamide or Bimatoprost + Timolol + Dorzolamide	Travoprost + Timolol+ Dorzolamide	General population	Societal	Markov	Lifetime	5%	Probabilistic	-$ 104.00	U =(0.374)(VA)+ 0.514	Sharma et al.[44], Brown et al.[30]
Patel et al.[157]	2019	CAN	Trabecular microbypass stents	Standard of care	POAG	Societal	Markov	15	2%	Probabilistic	-$ 42,769.10	EQ-5D-3L	Lee et al.[15]
Pen et al.[177]	2005	EU	Latanoprost or Travoprost	Timolol	POAG	Societal	Markov	5	5%	Probabilistic	€ 23,828.00	HUI-3	Van Gestel et al.[104]
Rein et al.[149]	2009	USA	Screening program	Opportunistic case finding	General population	Payer	Microsimulation	Lifetime	3%	Deterministic	$ 11,000.00	U =(0.374)(VA)+ 0.514	Sharma et al.[44]
Stein et al.[181]	2012	USA	Laser Trabeculoplasty or Timolol or Brimonidine	Observation	POAG	Payer	Markov	25	3%	Probabilistic	$ 14,179.00	QALYs = 0.98991 + 0.0022 dBs - 0.00080518 dBs^2	Rein et al.[35]
Stewart et al.[158]	2008	USA	timolol or latanoprost or bimatoprost or travoprost or brimonidine	Observation	OHT	Payer	Markov	5	3%	Deterministic	$ 89,072.00	Direct elicitation	Lee et al.[15], Brown et al.[30]
Stewart et al.[150]	2009	GBR/NOR/ DNK/SWE	Latanoprost and timolol as monotherapy in the treatment of open-angle glaucoma	Observation	POAG	Payer	Markov	5	4%	Deterministic	$ 8,175.00	Direct elicitation	Tengs et al.[36]
Tang et al.[178]	2019	CHN	Community-level glaucoma screening (Wenzhou glaucoma screening program)	Opportunistic case finding	General population	Payer	Markov	30	4%	Deterministic	$ 9,060.00	Direct elicitation	Kobelt et al.[34], Tengs et al.[36]
Thomas et al.[169]	2015	CAN	Tele glaucoma screening device	In person screening	General population	Payer	Markov	30	3%	Probabilistic	-$ 29,749.00	EQ-5D-3L	Burr et al.[31], Brown et al.[30], Sun et al.[144]
Vaahtoranta-Lehtonen et al. [176]	2007	FIN	5-year interval screening program	Opportunistic case finding	General population	Societal	Markov	Lifetime	4%	Probabilistic	€ 32,602.00	EQ-5D-3L	Burr et al.[31], Tengs et al.[36]
van Gestel et al.[37]	2012	NLD	(Start on timolol or Start on Latanoprost) or Start Timolol + Target 15 mmHg or Start Timolol + VF every 6 months or Start Timolol + VF every 245 months	Standard of care	POAG	Societal	Discrete Event Simulation	Lifetime	2%	Deterministic (Structural)	£ 12,931.00	15D	Van Gestel et al.[104]
van Gestel et al.[170]	2014	NLD	Immediate pressure-lowering treatment for OHT with 21 mmHg target pressure – after converting to POAG target at 18 mmHg, with reduce target if further VFL progression (Direct Treatment) or Incremental Treatment	Standard of care	OHT	Societal	Decision Analytic	Lifetime	2%	Probabilistic	€ 33,645.00	HUI-3	Van Gestel et al.[104]
Wittenborn et al.[38]	2011	GHA	Population level glaucoma screening	Opportunistic case finding	General population	Societal	Markov	Lifetime	3%	Deterministic	$ 12,108.00	HUI-3	Van Gestel et al.[104]
Xu et al.[143]	2016	CHN	Phacoemulsification	No treatment	POAG with cataract	Payer	Decision Analytic	13	3%	Probabilistic	$ 727.00	WHO DALY	-
Lazzaro et al.[39]	2022	Germany	STN1013001	Other latanoprost formulations	POAG or OHT with OSD	Payer	Markov	5	3%	Probabilistic	-€ 573.77	Algorithm and EQ-5D-3L	van Gestel et al.[104]
Lazzaro et al.[40]	2022	France	STN1013001	Other latanoprost formulations	POAG or OHT with OSD	Payer	Markov	5	3%	Probabilistic	€ 21.26	Algorithm and EQ-5D-3L	van Gestel et al.[104]
Healey et al.[141]	2021	Australia	Trabecular bypass devices	Standard of care	POAG with cataract	Payer	Microsimulation	20	5%	Deterministic	$2,430	Direct	Brown et al.[30]
King et al.[164]	2021	GBR	Trabeculectomy	Medical management	POAG	Societal	Markov	Lifetime	4%	Probabilistic	£ 9,596.43	Algorithm	-
Teus et al.[161]	2021	ESP	iStent and Phacoemulsification	Phacoemulsification alone	POAG with cataract	Payer	Markov	Lifetime	3%	Probabilistic	$ 12,293.86	Algorithm	van Gestel et al.[104]
Fea et al.[163]	2021	ITA	iStent and Phacoemulsification	Phacoemulsification alone	POAG with cataract	Payer	Markov	Lifetime	4%	Probabilistic	€ 13,037.00	Algorithm	van Gestel et al.[104]
Nieland et al.[162]	2021	France	iStent and Phacoemulsification	Phacoemulsification alone	POAG with cataract	Societal	Markov	Lifetime	4%	Probabilistic	€ 1,154.00	Algorithm	van Gestel et al.[104]

Table 2: Health utilities, values, and sources not reported by HPA health states used to inform CUA model parameters.

CUA first author (year)	Utility Measure	Utility Scale	Preference Source	Health State	Value	Utility first author (year)
Ahmed et al.[166]	HUI-3	Policy (PH-D)	General population	Trabeculectomy side effect or complication - disutility	0.007	van Gestel et al.[104]
				Medication-related side effects - disutility	0.101	van Gestel et al.[104]
Brown et al.[159]	TTO	Vision (PV-B)	Patient population	Bimatoprost 0.01	0.9273	Brown et al.[30]
				Travoprost 0.004	0.9081	Brown et al.[30]
				Latanoprost 0.005	0.895	Brown et al.[30]
				Tafluprost 0.0015	0.9265	Brown et al.[30]
				Timolol 0.5	0.9043	Brown et al.[30]
				20-20 VA	0.97	Brown et al.[30]
				20-25 VA	0.902	Brown et al.[30]
				20-40 VA	0.8	Brown et al.[30]
				20-70 VA	0.722	Brown et al.[30]
				20-800 VA	0.52	Brown et al.[30]
John et al.[180]	SG	Vision (PV-D)	Patient population	Medical treatment	0.88	-
				Surgical treatment	0.87	-
Kaplan et al.[152]	TTO	Vision (PV-B)	Patient population	20-60 VA	0.74	-
Patel et al.[157]	HUI-3	Policy (PH-D)	General population	Medication-related side effects - disutility	0.101	van Gestel et al.[104]
Stein et al.[181]	TTO	Vision (PV-B)	Patient population	Side effect or complication	0.75	-

on the perfect health to death scale, with the remaining 20 health utilities on the perfect vision to blind/death scale. Unique health state values were grouped by HPA health states, measure scale, and elicitation method and are plotted in Figure 2. Utilities estimated by EQ-5D alongside clinical trials with intervention defined states are illustrated in Figure 3. Table 2 reports the utility values of health states different from HPA health states or applied utilities collected alongside clinical trials to inform a CUA. A wide range of health utilities for the same HPA glaucoma health states were reported; OHT (0.84 - 0.95), Mild (0.68-0.94), Moderate (0.57-0.92), Advanced (0.58-0.88), Severe/blind (0.46-0.76), and Bilateral blindness (0.26-0.5). Health states described with visual acuity were reported one primary source from 0.97 (20/20) to 0.52 (20/800).[30] Table 2 describes the treatment status health states ranged from 0.74 to 0.93 with preference weighting from both the general population (generic measure) and patients (direct elicitation). Most studies used health utilities from directly estimated patient health utilities and generic measures of health state with preferences of a general population across health states. Health utilities were often combined from multiple sources that reported different elicitation strategies (e.g., Time-Trade Off [TTO], Standard Gamble [SG], Visual Analog Scale [VAS]) on scales of both perfect health to death and perfect vision to blindness. Health utilities were also assigned to hypothetical cohorts (Markov state transition, decision tree) or to individual simulated patients (microsimulation, discrete event simulation). Individual assignment of health utilities depended on the simulated clinical features including visual field loss (VFL) or visual acuity (VA). Clinical features were input as parameters into 2 linear functions to convert a hypothetical clinical presentation to health utilities (1) from best-eye mean deviation (dBs) and (2) from best-eye visual acuity.

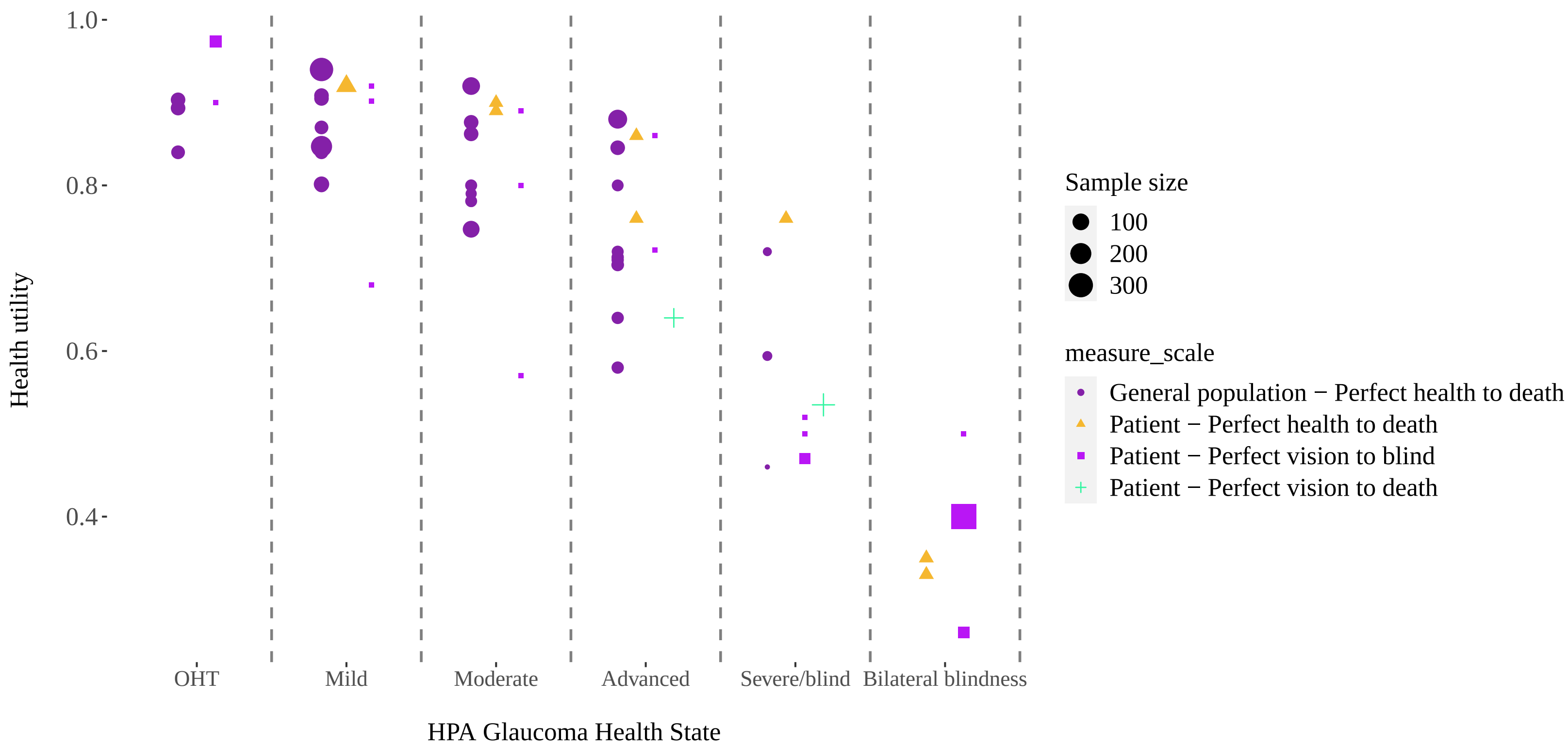


Figure 2: Unique health utilities values organized by HPA glaucoma health states and grouped by the scale of preference measurement.

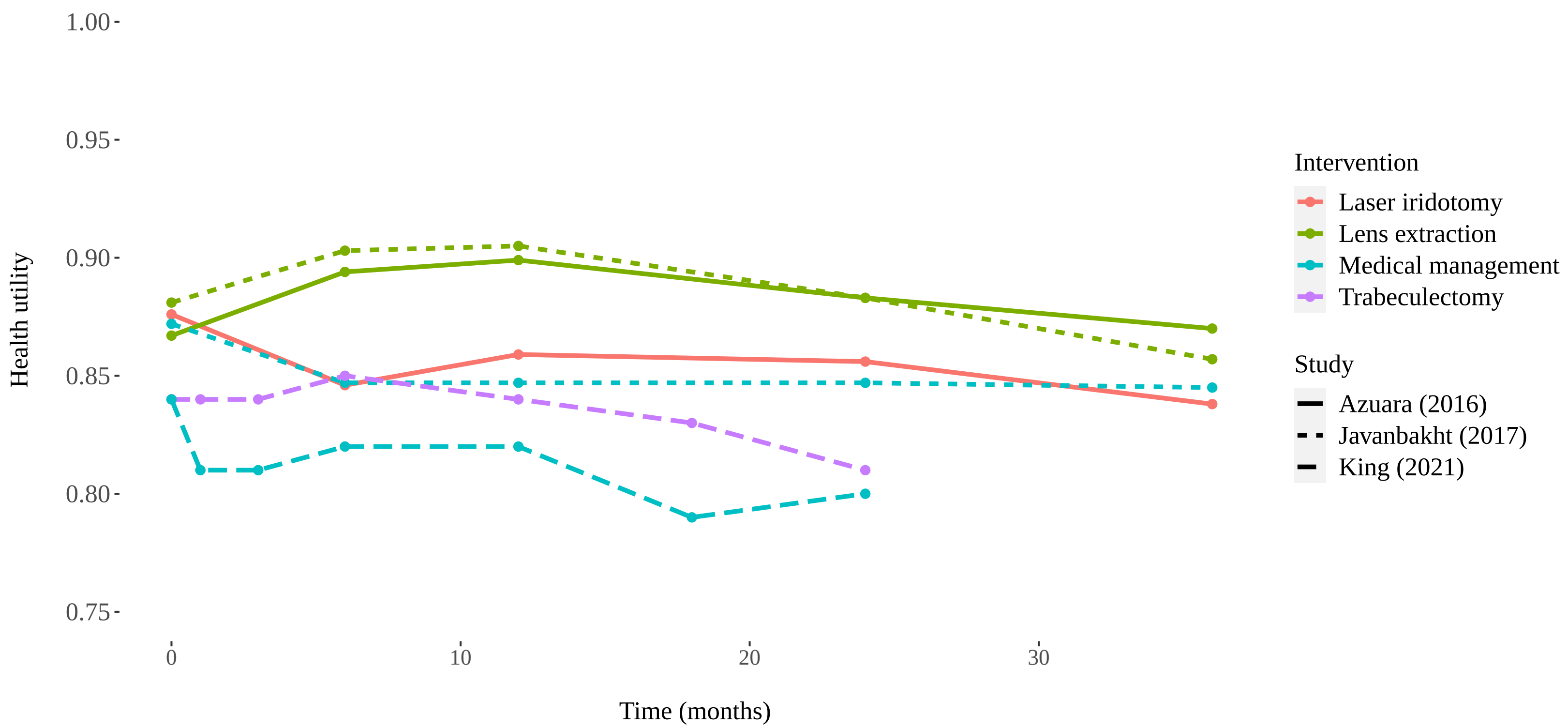


Figure 3: EQ-5D-3L Health utilities of treatment condition states collected alongside clinical trials.

$$(1) QALYs = 0.98991 + 0.0022dBs - 0.00080518dBs^2$$

$$(2) U = (0.374)(VA) + 0.514.$$

2.4.3 Health utility reporting quality

Of the 43 CUA studies included, the majority (33, 77%) reported the actual health utilities used, the basis for using them (34, 79%), the original reference (35, 81%), and any assumptions or adjustments applied to the health utilities (22, 51%). Few studies reported assessing the relevance of health utilities to a decision context (8, 19%). Even fewer (3, 7%) applied a systematic search strategy for health utilities with search terms, scope, study selection criteria, and a structured assessment of quality. CUA were assessed to report between 0 (none) and 15 (all) SpRUCE items. 23 of 43 CUA (54%) reported between 4 and 7 items of the checklist (4).

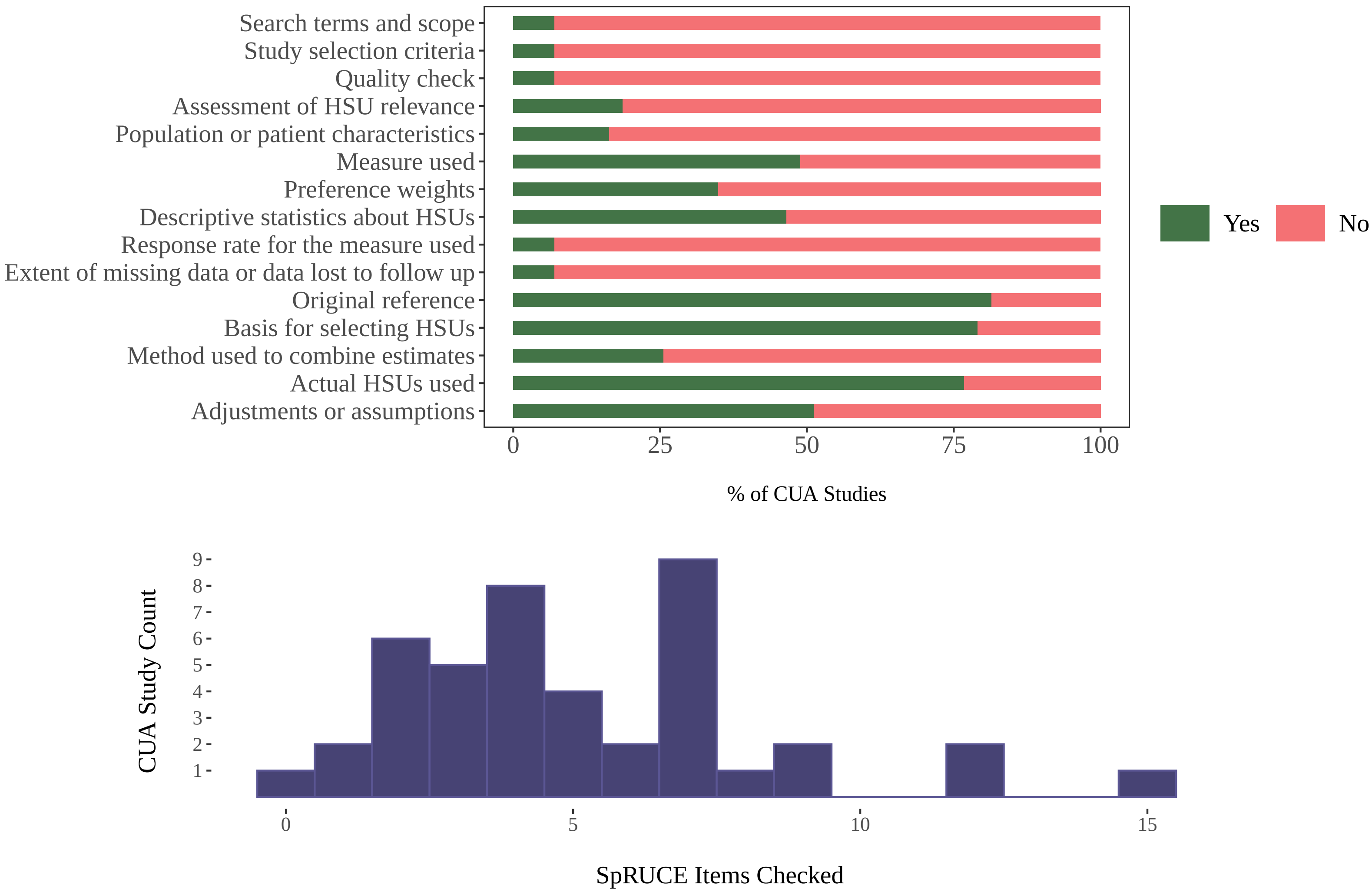


Figure 4: Proportion of glaucoma CUA meeting SpRUCE Checklist criteria by item with a histogram describing the total number of checklist items met.

2.5 Discussion

In this review, we summarized the methods of assigning health utilities to glaucoma health states, described the use of different measurement scales and preference sources within CUA, and applied the SpRUCE checklist to assess the reporting quality of using health utilities. Few glaucoma cost-utility studies adhere to reporting standards for use and selection of health state utility values.

2.5.1 Glaucoma health state

The design of the economic model has an impact on how health utilities are assigned to health states. Studies combined definitions of blindness with utilities describing the health states of patients with various ocular conditions as equivalent to one-eye blindness and bilateral blindness. Assigning utility values to patients based on interventions can be problematic, particularly in glaucoma, as patients may receive interventions across all levels of severity, often in combination, limiting the precision of measuring the health state and therefore the conclusions of the CUA.

2.5.2 Measures of health utility

Most studies reported using health utilities directly elicited from patients for at least some of glaucoma health states. Often, preferences were elicited on a scale of perfect vision to blindness. The scale of the measure and the frame of which the question is asked influences the response pattern and ultimately the value assigned to the health state. Preference tasks designed with these scale anchors are not directly comparable to health utilities on a perfect health and death scale. Use of alternative anchors in health utilities elicitation limits the usefulness of these values towards informing health policy, where it is common to consider the relative societal value of a therapy relative to therapies for other diseases. Gambling or trading values with anchors of perfect vision and blindness measures health utilities from a different context, not on the health utility scale. This affects the validity of claims made for incremental value of interventions. A previous review of vision-related health utilities has identified that the perfect vision blind scale results in truncated values relative to utility scale, overestimating health utilities for ophthalmological diseases and overstating cost-effectiveness of the treatment/prevention of vision-related disease relative to other medical conditions.[15] Of note, two of the most recent studies by the same author relied on expert opinion to estimate health utilities[39,40]. Utility values for a medicated state were derived from setting maximum and minimum values, leaving in between health states to be assigned from the opinions of an expert. This should be avoided as health utilities are not observable from a third party through routine clinical examinations as they are a function of preference and related to trade-offs. Limitations of clinicians assessing patient health utilities and quality of life have been strongly advocated against elsewhere.[41,42]

Special care is required when using a deterministic linear function to estimate utility from clinical characteristics. The health utilities mapped by Rein et al.[35] were identified as 'temporary values' to facilitate glaucoma cost-effectiveness research until 'empirical' estimates are available. The Rein et al conversion function was identified to not account for visual acuity losses and asserts the limitation that these functions we not generated with consideration of best-eye or worse-eye, highlighting the importance of binocular summation method which may be related to visual sensitivity being 40% higher. The authors described that the conversion function assumed that those with complete impairment from VFL would experience the same QALY loss as those with 'no impairment', but with legally defined blindness. The health utilities source estimates that informed the Rein et al and the Sharma et al conversion functions were derived from Brown et al original investigation of patient utilities.[43,44] The health utilities that informed both linear equation were elicited from patients with various conditions, including macular degeneration, cataract, glaucoma, and diabetic retinopathy. Despite many studies endorsing use of linear equations to estimate health utilities, these linear equations only considered either mean deviation, visual acuity, and whether or not the patient had cataracts. Glaucoma is known to affect both visual field loss and visual acuity at different stages of severity, neglecting either clinical feature in ascribing health utilities to individuals in economic models leads to inaccurate results. Additionally, in models estimating health utilities over lifetime there are other known factors to affect health utilities beyond the discrete health state including age and gender.[45]

Future use of health utilities should aim to be informed by evidence obtained from studies that used the same preference-based measure and preference weights, with utilities on the same scale of measurement. When multiple appropriate health utilities are available, or it is not possible to identify all health utilities from same measure, it can be reasonable to synthesize information. However, this limitation affecting the validity of the conclusions of a CUA should be described with consideration towards the CUA generalizability.

[If someone was using paper 1 to determine which utilities to use in CUA, what should they choose?]

2.5.3 Reporting quality

Few studies used systematic reviews to identify applicable health utilities. The justifications for use of particular health utilities sources over others, or reasons not to employ a systematic search strategy were absent from the published studies that did not perform a review of health utilities. Use of different samples, estimation methods, and preference weights can result in different health utilities for the same health state. Selecting evidence in an ad hoc manner affects justification of conclusions of the CUA where health utilities are 'cherry picked'. This concern can be alleviated through systematic review of the literature for health utilities and setting clear criteria for including health utilities.[46–48] A consistent comment within the CUA included in this review is a lack of available health utilities that reflect the preferences of the countries and guidance towards optimal study design for generating health utilities. Overall, many studies cite a lack of health utilities in general.[49] This may be due to the methods used to search for health utilities. Studies with the same authors over time often referenced their previous work to describe the identification and use of health utilities. Subsequent CUA with the same health utility data is efficient to conduct related investigations. However, as the evidence base grows and new appropriate health utilities may become available, subsequent reports should be held to the same standard of rigor and description to inform the reader with an unbiased and transparent analysis. We encourage authors to include a SpRUCE checklist as a supplementary file with future glaucoma economic evaluation to encourage adherence to good practice guidelines in reporting the selection and use of health state utilities.

2.5.4 Strengths and limitations

To our knowledge, this is the first use of systematic review to investigate methods applied and reporting of health utilities in the available glaucoma CUA literature. This study offers unique insights into the decisions clinical investigators and health economists make when modeling incremental value of treatments and programs in glaucoma. We believe that this methodology can be applied in other disease areas to identify the state of practice and inform future studies of strategies to improve the transparency and validity of CUA. Through this methodological review, we aggregated approaches in a variety of decision contexts, time horizons, and model structures to identify specific areas to improve the transparency and generalizability of future CUA of glaucoma interventions.

This review is limited in that SpRUCE checklist items has a large variation between users as it directs investigators to perform a quality check.[50] We attempted to account for this by having multiple reviewers apply the checklist. This review analyzes the identification and use of health utilities, however does not comprehensively review the literature for published glaucoma health state utilities. This review does not consider the impact of health utilities on the conclusions of ICUR estimates.

2.6 Conclusion

This methodological review describes the state of practice among published glaucoma CUA studies. Descriptions were provided for the characteristics of glaucoma CUA, the health utilities sources, and assess the health utilities reporting standards against good practice guidelines with the intention of improving the reporting of future glaucoma CUA. We suggest including a SpRUCE checklist[28] as a supplementary file with future glaucoma economic evaluation to justify decisions that are beyond the scope of what is required to be reported by journal specifications.

3 DISEASE-SPECIFIC PREFERENCE-BASED MEASURE OF GLAUCOMA HEALTH STATES: HUG-5 PSYCHOMETRIC VALIDATION

Published as: Kennedy, K., & Podbielski, D., & Nanji, K., & Muratov, S.M., & Ahmed, I.I.K., & Xie, F. (2019). *Journal of Glaucoma.* (Published July, 2019), DOI: 10.1097/IJG.0000000000001267

3.1 Foreward

The second paper of this dissertation builds on the need for reliable estimates of health utility by investigating the psychometric properties of the HUG-5 when compared to a frequently used preference-based generic measure of health state (EQ-5D-5L) and vision-specific measure of health related quality of life (NEI-VFQ-25). The individual attributes of the EQ-5D-5L and the sub scales of the NEI-VFQ-25 were compared to HUG-5 attributes for discriminant and convergent construct validity. The test-retest reliability after a 2-week period and an additive composite score was used to describe the sensitivity of the HUG-5 in measuring health states in mild/moderate and advanced glaucoma patients.

124 patients with glaucoma were administered the NEI-VFQ-25, the HUG-5, and the EQ-5D. The HUG-5 demonstrated construct validity, with convergent and discriminant support for visual discomfort, mobility, daily life activities, emotional distress, social activities. The HUG-5 concurrently measured HRQoL associated with best-eye visual field loss $r = 0.63$, $p < 0.001$. The HUG-5 measured health state consistently with test-retest reliability Intraclass Correlation = 0.91, $p < 0.001$. The HUG-5 was established to be sensitive in detecting differences between mild/moderate and advanced patients with a rank-sum test with continuity correction ($W = 693.5$, $p < 0.001$).

Overall, this study demonstrates that the HUG-5 can effectively measure glaucoma health states and relationship with the NEI-VFQ and best-eye visual field loss in a sample of glaucoma patients.

3.2 Introduction

Glaucoma is a clinical term for a group of optic neuropathies characterized by the progressive degeneration of retinal ganglion cells and increased intraocular pressure. It is associated with damage to the optic nerve, generally damaging peripheral vision in its early stages, eventually progressing to central vision loss.[5,51] Glaucoma is the second leading cause of irreversible blindness in the world with 60 million people worldwide estimated to have glaucoma[52], a figure expected to rise to 76 million by the year 2020.[53] Consequently, there is a rising economic burden associated with glaucoma. In 2004, the direct costs in the United States were estimated at $1450 per person.[54]

Although glaucoma progression or treatment have no sizable impact on mortality, both have a significant effect on the health-related quality of life (HRQoL) of patients.[55–58] HRQoL is a measure of how well individuals are able to function and how they perceive the quality of their physical, mental, and social dimensions of life.[59] As glaucoma progresses, patients often experience troubles with daily activities and poor emotional well-being, it is important to measure these changes.[60] HRQoL is an integral component that informs health economic evaluations and maximizing HRQoL is an important goal of patient centered healthcare service.[61]

Value of treatments to society and financial costs should be considered for the appropriate allocation of scarce healthcare resources. Economic evaluations synthesize evidence on costs and outcomes to inform resource allocation decisions. The quality adjusted year (QALY) is a common outcome that allows comparisons across disease and clinical outcomes. It is a measure of disease burden incorporating length of life and impact on HRQoL associated with an intervention.[62] Economic evaluations synthesize evidence on costs and QALY and are widely used to inform resource allocation decisions.[63] QALYs are calculated from health state utilities values; measurements of patients' health states, measuring utility between 0 and 1, where 0 indicates death and 1 indicates full health.[64] QALYs are calculated by multiplying the amount of time an individual spends in a particular health state by the HRQoL weight associated with that particular health state.[65] Health utility can be obtained directly for common measures of HRQoL with preference elicitation techniques such as the time trade off, standard gamble, or visual analogue scale. However, measuring accurate health utility with these methods has proven challenging and time consuming.[65] Alternatively, health utility can be measured indirectly from generic preference-based instruments, for example the EuroQol's 5 Dimensions (EQ-5D). The EQ-5D is generic preference-based instrument for measuring health status.[66] The generalizability of the EQ-5D comes at the expense of a loss of sensitivity to measure condition-specific characteristics.[67,68] Furthermore, generic preference-based measures have demonstrated a lack of sensitivity in the vision sciences.[69] Existing vision or glaucoma specific instruments (e.g., National Eye Institute Visual Function Questionnaire or Glaucoma Quality of Life-15) consist of multiple domains and multiple items per domain formulatig subscale scores.[70–72] Developing a mapping algorithm to convert existing non-preference-based glaucoma-specific HRQoL scores to health utility generated by statistical models is an alternative that has been explored in recent years.[73] Mapping results have been indicated as inconclusive, with studies often reporting poor model performances and limited validity.[74–76] Moreover, existing glaucoma instruments have been reported to measure dimensions that do not cover the HRQoL consequential of treatments.[77] The only available glaucoma-specific preference-based measure is the Glaucoma Profile Instrument (GPI) developed by Burr et al.[31] The GPI consists of six dimensions with four response levels per dimension. There is a limitation impacting the psychometric validity and reliability of the instrument. The GPI uses a catch-all category in the final question of the tool "Other possible effects of glaucoma or its treatment". Furthermore, the anchors used to elicit preferences are 0 (worst state) and 1 (best state) rather than 0 (dead) and 1 (perfect health). As a result of limitations in detecting differences in glaucoma patients by existing measures[73,78,79], health utilities cannot be calculated. This limits the conclusions of glaucoma economic evaluations and has consequently impacted how economic evaluations contribute to decision-making within the field of glaucoma.

The Health Utility measure for Glaucoma (HUG-5) was developed from the information collected in a systematic literature review, patient interviews and focus groups. The HUG-5 consists of 5 questions, assessing 5 levels of 5 dimensions.[80] The HUG-5 measures health states in patients with glaucoma that represents dimensions important to patients and consequential to treatment alternatives.[16] The objective of this study was to establish validity and reliability of the HUG-5 in capturing the impact of glaucoma disease severity on HRQoL.

3.3 Materials and Methods

3.3.1 Inclusion Criteria

Eligible patients were required to have a confirmed diagnosis of glaucoma and have impairment in one or both eyes. Glaucoma is typically staged as early or mild, moderate and advanced. According to the Canadian Ophthalmological Society, glaucoma clinical practice guidelines, mild glaucoma is characterized by an early optic nerve change and (or) mild visual field defect not within 10 degrees of fixation and mean deviation (MD) of visual field loss (VFL) greater than -6dB.[81] Moderate glaucoma is characterized by more disc features and a larger visual field defect not within 10 degrees of fixation with an MD from -6 to -12dB. Advanced glaucoma has more severe optic disc changes and (or) a visual field defect within 10 degrees of fixation with an MD of more than -12dB on a visual field. Previous papers specifying the planning and development of the HUG-5 describe the methodology and rationale for grouping mild and moderate patients to be compared with advanced glaucoma patients.[16,80] To participate, patients were required to be at least 18 years of age, able to speak and write English, provide informed consent and not belong to a vulnerable group (e.g., people with mental illness, memory problems, learning difficulties, etc.). Patients were identified to meet study criteria by their attending ophthalmologist specialized in glaucoma care.

3.3.2 Instruments

The package administered to participants contained three questionnaires; the HUG-5, the EQ-5D-5L and the NEI-VFQ-25. The three instruments were ordered randomly within the order of presentation to avoid any experimental order effect. Participants were asked to provide information on their characteristics in a demographic questionnaire.

3.3.2.1 HUG-5 The Health Utility measure for Glaucoma (HUG-5) is a five question, disease-specific, preference-based measure. Each question has 5 response options reflecting none, slight, moderate, very much, severe impact of glaucoma on five HRQoL dimensions. The composite score for the HUG-5 was calculated as unit-weighted, where each of the five dimensions were equally weighted and summed. For example, if a patient selects none for all five dimensions, the composite score for that patient is 5. Alternatively, if a patient selects severely for all five dimensions, the composite score for that patient is 25. This measure took patients less than two minutes to complete. After the validation, the next step is to develop HUG-5's preference-based scoring algorithm which allows the conversion of health state responses to health utility.

3.3.2.2 EQ-5D-5L The EQ-5D-5L is a five-dimension, generic preference-based instrument for measuring health status.[66] Similar to the HUG-5, there is one question for each dimension. This measure records responses on scales from having no problems (1) to being (5) unable or experiencing extreme levels of the dimension. Patients are asked on their general state describing their health that day with regard to: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression and to rate their current health state on a visual analogue scale (VAS) from 1 to 100. Similar to the HUG-5 composite score, the EQ-5D-5L composite was also calculated using unit weighting, with possible scores ranging from 5 to 25. This measure took patients less than two minutes to complete on average.

3.3.2.3 NEI-VFQ-25 The National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) measures vision-related quality of life with 10 scales and a composite score. The NEI-VFQ-25 is not specific to glaucoma and each scale is informed from multiple items. The NEI-VFQ records responses on three types of adjective scales. The first ranges from 1- no difficulty to 4 - extreme difficulty. The second adjective scaling captures the extent a statement applies to the patient; ranging from 1 - all of the time) to 5 - none of the time. The third adjective scaling captures the extent a statement is true for the patient; ranging from 1 definitely true to 5 definitely false. Patients are asked on their experiences relating to ocular pain, driving, near activities, mental health, social function, peripheral vision, distance activities, role difficulties, general health, general vision, and color vision. Numeric values associated with each of the NEI-VFQ-25 questions are re-coded following the scoring rules outlined in Mangione's scoring algorithm. In a brief summary of the scoring algorithm, items comprising a subscale are averaged for a subscale scores. The composite score was calculated by averaging together all sub-scales.[70] The NEI-VFQ scoring was flipped and rescaled to be directly comparable to the EQ-5D and the HUG-5 (e.g. a score of 100 would be rescaled to 5 and a score of 0 would be rescaled to 25).

3.3.3 Data Collection

Following their clinical encounter, each patient was identified for meeting inclusion criteria and given a short verbal introduction to the study by their attending ophthalmologist. If the patient expressed an interest in participating, the ophthalmologist directed the patient to the onsite study administrator stationed in another clinical office. The study administrator reviewed the contents of the study consent form, describing the purpose and rights of the patient in detail before obtaining the informed consent. All study materials were approved by the Trillium Health Partners Ethics Review Board in August 2017. Patients were randomly selected to receive a follow up phone call after two weeks where the study assistant administered the HUG-5 over the phone. Patients were monitored by the study administrator while completing all instruments, in the event a patient could not read the instruments themselves, the administrator would dictate the questions to the patient and record their responses. In these cases where written consent was not possible, verbal consent was obtained. All participating patients were reimbursed with a 10$ gift card to a national retailer. Demographic data and patient tracking were stored within a protected spreadsheet separate from survey data that used unique identifiers to index patient data. Patients who had elected to cease study participation were removed from record and their physical survey responses were destroyed.

3.3.4 Psychometric Properties

This study uses indices of: validity, reliability, and sensitivity as evidence for use of the HUG-5 as the standard instrument to assess HRQoL in patients with glaucoma across broad dimensions for use in clinical trials that include economic components in the evaluation of glaucoma therapies.

3.3.4.1 Score Distributions Differences of shape between composite score distributions of the HUG-5, EQ-5D-5L and subscales of the NEI-VFQ-25 were considered. Patient response patterns to the HUG-5 dimensions are described with response frequencies comparing mild/moderate and advanced patients. The score distributions of the three measures were investigated for skewness, ceiling and floor effects. Patient response patterns were further explored to provide additional insight into which components of the HUG-5 contributed to skew or ceiling effects.

3.3.4.2 Validity Construct validity was assessed by evaluating three types of evidence that support use of the HUG-5 in measuring the HRQoL of glaucoma patients; concurrent, convergent, and discriminant validity. Concurrent validity was investigated by contrasting the HUG-5 with the rescaled NEI-VFQ-25 composite scores, the standard for non-preference based HRQoL tools in ophthalmology[74] and the EQ-5D-5L composite scores. Rescaled NEI-VFQ scores and EQ-5D-5L composite scores were investigated relative to the HUG-5 composite score in their relationship with best-eye visual field loss.[51] Patient best-eye VFL was determined by identifying the greater value of bilateral eye oculus uterque (OU) in mean deviation (dB). Mean deviation (MD) is a global index available on the visual field that describes the average value of overall deviation from the expected results within the same age group of normal visual fields. Depression in a visual field is indicated as a negative MD value. If right oculus dexter (OD) or left oculus sinister (OS) VFL in MD was greater, that value was used to describe the relationship between better eye VFL and HRQoL.

Convergent validity was assessed by correlating NEI-VFQ-25 subscales and EQ-5D-5L dimensions with HUG-5 dimensions. A strong correlation of $r >= 0.5$ demonstrated convergent validity for related constructs. The HUG-5 visual discomfort dimension was expected to demonstrate a strong relationship with general vision and ocular pain NEI-VFQ-25 subscales, and best-eye MD. The HUG-5 dimension included measures of general vision performance as criterion for validity because the question addresses difficulty seeing in darkness as a component of visual discomfort and burning or itching in and around the eye. The HUG-5 mobility dimension was expected to demonstrate a strong relationship with the NEI-VFQ-25 subscales general vision, near activities, distance activities, and peripheral vision; best-eye MD; and the mobility and usual activities dimension of the EQ-5D-5L. It was rationalized that the measurement of glaucoma's effect on free movement would significantly effect activities and overall visual function. The HUG-5 daily living dimension was expected to demonstrate a strong relationship with general vision, near activities, distance activities, role difficulties, and dependency subscales of the NEI-VFQ-

25; a relationship with best-eye MD and the usual activities dimension of the EQ-5D-5L. Subscales of the NEI-VFQ and dimensions of the EQ-5D were chosen to reflect the capacity of this dimension to capture glaucoma's effect on routine activities that people tend to do every day without requiring assistance. The HUG-5 emotion dimension was expected to demonstrate a strong relationship with the mental health subscale of the NEI-VFQ-25 and the anxiety/depression dimension of the EQ-5D-5L. These relationships were expected to capture the extent that glaucoma effects the range of symptoms and experiences of a person's internal life that are troubling, confusing, or out of the ordinary. In one question attempting to capture the effect glaucoma has on an individual's current state of subjective well-being. The HUG-5 social activities dimension was expected to be strongly related to the NEI-VFQ-25 subscales general visions, near activities, distance activities, social function, and role difficulties; best-eye MD, and the usual activities dimension of the HUG-5. These relationships between subscales and dimensions were expected to reflect the negative effect of glaucoma on activities relating to, or having the purpose of, promoting companionship and participating in communal activities.

Discriminant validity was assessed by correlating HUG-5 dimensions with NEI-VFQ-25 subscales, EQ-5D-5L dimensions and best-eye MD that were expected to be unrelated to the construct measured. A poor correlation of 0.3 demonstrated discriminant validity for unrelated constructs. Few predictions were made for discriminant validity, and most HUG-5 dimensions were expected to follow atleast a mild correlation between HUG-5 dimensions and other measures. One important hypothesis was tested to evaluate if the HUG-5 dimensions reflected similar properties as the NEI-VFQ-25 as a vision related measure of HRQoL. Similar to the NEI-VFQ's initial validation effort, all HUG-5 dimensions were expected to be discriminant from the general health subscale of the NEI-VFQ-25.[82] The HUG-5 emotion dimension was expected to discriminate from the NEI-VFQ-25 subscale for general vision, best-eye MD, as well as the mobility, self-care, usual activities, and pain/discomfort domains in the EQ-5D-5L. All HUG-5 dimensions were excepted to be discriminant of the EQ-5D-5L dimension of pain/discomfort.

3.3.4.3 Test-Retest Reliability Test-retest reliability refers to the repeatability of a measurement administered on two occasions during which there is no significant change in the patient's status. The HUG-5 was administered two weeks after the baseline to a random group of patients (n=24). Reliability was measured by determining the Intraclass Correlation Coefficient (ICC) of the two administrations of the HUG-5 using the composite score. An ICC 0.4 was considered acceptable, between $0.6 < x < 0.74$ good, and excellent $0.75 < x < 1.00$.[83]

3.3.4.4 Sensitivity Sensitivity of the HUG-5 was evaluated by a cross-sectional comparison among patients with two disease severity levels (mild and advanced) with a Wilcox nonparametric test of independent group differences[84], contrasting the two groups composite scores. The HUG-5 was considered sensitive if the measure rejects the null that there is no difference between severity groups mild/moderate and advanced.

3.4 Results

Demographic and clinical characteristics of the sample are summarized by OU best-eye severity in Table 3. 124 patients diagnosed with glaucoma participated in the study. After the initial data collection period, 4 patients opted to remove their data from the study, leaving a final sample of 120 patients. The mean age of the overall sample was 67.0 (16.2) years, where most patients were employed (32.5%) or retired (58.3%), females (43.4%) and white (40.0%) or asian/pacific islander ethnicity (31.7%). Advanced patients were more likely to report being single than mild-moderate patients (50% > 29%), were retired (68% > 55%), and less frequently had a documented treatment change on the day of their participation in the study (9% < 23%). Comparable proportions were observed between advanced and mild/moderate patients for age categories and type of glaucoma.

3.4.1 Score Distributions

Figure 5 describes the distribution of HUG-5, NEI-VFQ-25, and EQ-5D composite scores. All three measures of HRQoL present with a positive skew, reflecting a tendency for most participants to score relatively high ratings of HRQoL. The range of composite scores reported by patients was altered by the generic nature of the EQ-5D. In this sample of patients with glaucoma, the EQ-5D response distribution ceiling effect captured a ceiling effect, where the instrument failed to capture health above no problems. This EQ-5D ceiling effect is illustrated in Figure 5. In contrast with the EQ-5D, the HUG-5 avoided this ceiling effect. HUG-5 composite scores ranged from 5 (minimum possible) to 25 (maximum possible); compared to the EQ-5D-5L which ranged from 5 to 19 and the rescaled NEI-VFQ-25 which ranged from 5.4 to 21.3. Patient response patterns for the HUG-5 were further explored in Table 4, describing response frequencies of each level of all dimensions for mild/moderate and advanced patients. The most common concern ailing advanced patients with their HRQoL above slightly is visual discomfort, followed by mobility and daily life activities. Mild/moderate patients most commonly identified moderate concerns of visual discomfort (16.3%), emotion (14.0%) and mobility (11.6%).

3.4.2 Validity

3.4.2.1 Concurrent Validity All measures were significantly correlated with best-eye MD. The HUG-5 composite demonstrated a strong correlation $r = 0.63(0.51, 0.73), p < 0.001$ with VFL. The NEI-VFQ-25 composite scale also demonstrated a similar correlation with VFL, capturing a more precise estimate of $r = 0.70, (0.59, 0.78), p < 0.001$. The EQ-5D-5L was marginally less correlated with VFL and less precise relative to the HUG-5 and the NEI-VFQ-25, where $r = 0.54, (0.40, 0.66), p < 0.001$. Figure 6 depicts the relationships between HRQoL composite scores of the NEI-VFQ-25, HUG-5, EQ-5D-3L

Table 3: Sample characteristics.

	Mild/Moderate n (%)	Advanced n (%)	Total n (%)
Age Category			
18 to 40	2 (2.33)	2 (5.88)	4 (3.33)
41 to 60	20 (23.26)	5 (14.71)	25 (20.83)
61 to 80	47 (54.65)	22 (64.71)	69 (57.5)
>80	17 (19.77)	5 (14.71)	22 (18.33)
Gender			
Female	40 (46.51)	12 (35.29)	52 (43.33)
Ethnicity			
White	38 (44.19)	10 (29.41)	48 (40)
Hispanic or Latino	3 (3.49)	1 (2.94)	4 (3.33)
Black or African American	15 (17.44)	9 (26.47)	24 (20)
Asian/Pacific Islander	24 (27.91)	14 (41.18)	38 (31.67)
Other	5 (5.81)	0 (0)	5 (4.17)
Education			
Highschool Diploma or less	26 (30.23)	11 (32.35)	37 (30.83)
College	23 (26.74)	7 (20.59)	30 (25)
Undergraduate Degree	19 (22.09)	9 (26.47)	28 (23.33)
Graduate Degree	17 (19.77)	7 (20.59)	24 (20)
Marital Status			
Prefer not to answer	1 (1.16)	0 (0)	1 (0.83)
Single	25 (29.07)	17 (50)	42 (35)
Married or domestic partner	59 (68.6)	17 (50)	76 (63.33)
Employment Status			
Employed	31 (36.05)	8 (23.53)	39 (32.5)
Unemployed	7 (8.14)	2 (5.88)	9 (7.5)
Student	0 (0)	1 (2.94)	1 (0.83)
Retired	47 (54.65)	23 (67.65)	70 (58.33)
Glaucoma Type			
Open-Angle	76 (88.37)	28 (82.35)	104 (86.67)
Angle Closure	2 (2.33)	3 (8.82)	5 (4.17)
Mixed Mechanism	6 (6.98)	1 (2.94)	7 (5.83)
Juvenile Open Angle	1 (1.16)	2 (5.88)	3 (2.5)
Normal Tension	1 (1.16)	0 (0)	1 (0.83)
Cataracts	24 (27.91)	3 (8.82)	27 (22.5)
Treatment Change	20 (23.26)	3 (8.82)	23 (19.17)

Note: Demographics table reporting the characteristics of mild/moderate (at least one eye) and advanced patients (both eyes)

Table 4: HUG-5 item response frequencies by glaucoma severity.

	Level response frequency (%)				
	Not at all	*Slightly*	*Moderately*	*Very Much*	*Severely*
	Visual Discomfort				
Mild/Moderate (OS/OD)	50	25.58	16.28	4.65	3.49
Advanced (OU)	29.41	8.82	20.59	23.53	17.65
	Mobility				
Mild/Moderate (OS/OD)	68.6	15.12	11.63	3.49	1.16
Advanced (OU)	32.35	11.76	17.65	23.53	14.71
	Daily Life Activities				
Mild/Moderate (OS/OD)	75.58	16.28	5.81	1.16	1.16
Advanced (OU)	29.41	14.71	20.59	20.59	14.71
	Emotional Well-Being				
Mild/Moderate (OS/OD)	52.33	29.07	13.95	2.33	2.33
Advanced (OU)	26.47	29.41	20.59	11.76	11.76
	Social Activities				
Mild/Moderate (OS/OD)	87.21	4.65	3.49	2.33	2.33
Advanced (OU)	52.94	11.76	20.59	2.94	11.76

Note: HUG-5 responses of mild/moderate (at least one eye and advanced patients (both eyes) by dimension and level.

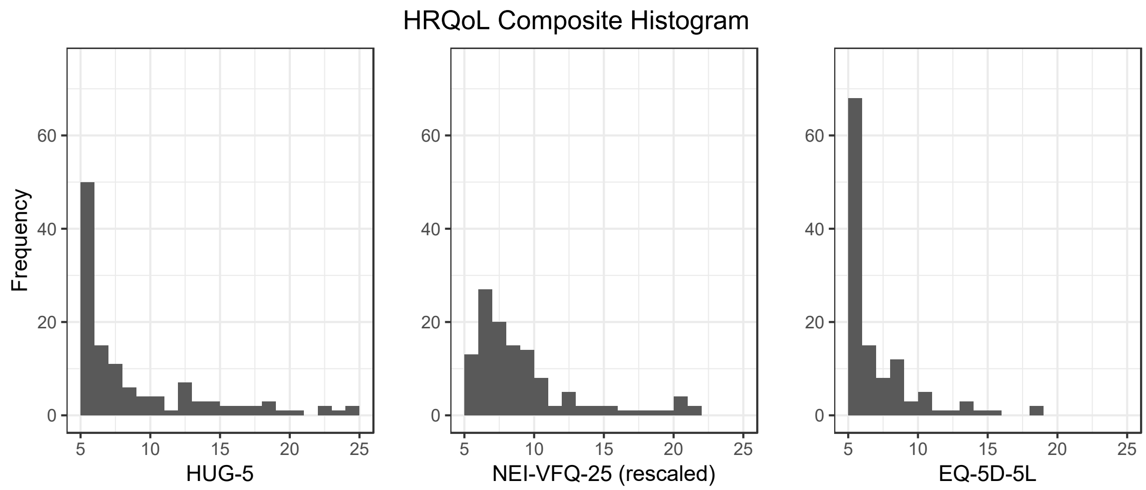


Figure 5: Histogram plots of HUG-5, NEI-VFQ-25 (rescaled) and EQ-5D-5L; illustrating the distribution of scores among this sample of mild, moderate and advanced glaucoma patients collectively.

and best-eye VFL. The complete correlation matrix describing the relationships between all dimension and subscales are reported in Table 5.

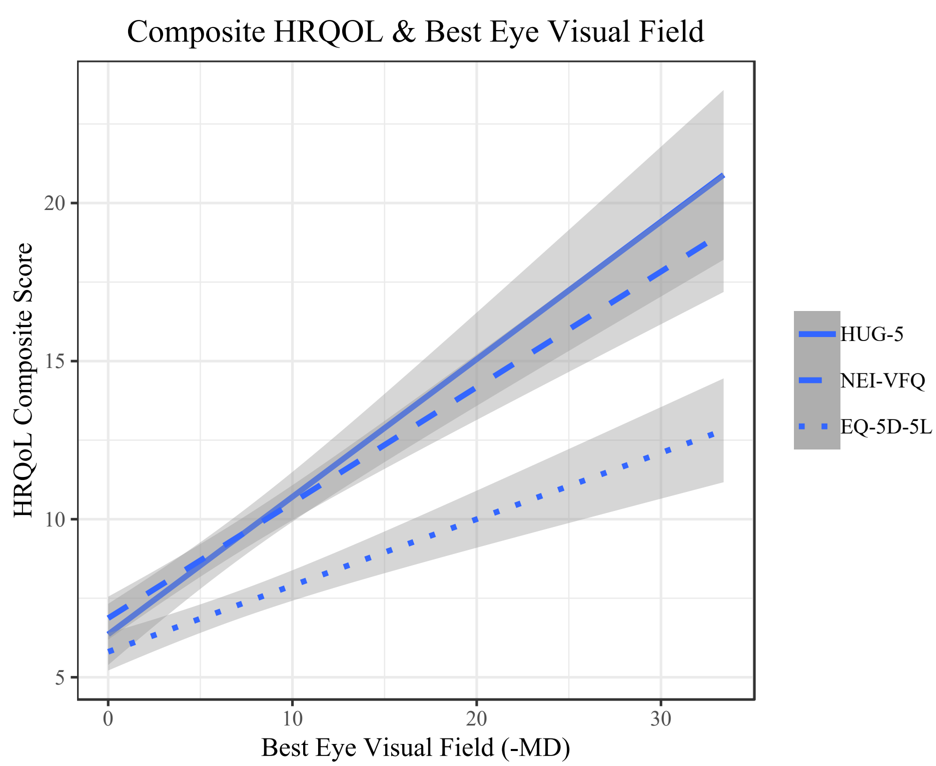


Figure 6: Relationships between HRQoL composite scores of the NEI-VFQ25, HUG-5, EQ-5D-3L and best-eye VFL

Table 5: Correlation Matrix of HUG-5, NEI-VFQ-25 and EQ-5D-5L.

	HUG-5					NEI-VFQ-25														EQ-5D-5L				
	Visual Discom fort	Mobility	Daily Living	Emotion	Social Activity	General. Health	General. Vision	Ocular Pain	Near. Activities	Distance. Activities	Social. Function	Mental. Health	Role. Difficulties	Dependency	Driving	Color. Vision	Peripheral. Vision	Comp	Best Eye MD	MOB	SEL	DA	P/D	A/D
HUG-5																								
Visual Discomfort	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mobility	0.71	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Daily Living	0.73	0.77	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Emotional	0.63	0.56	0.59	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Social Activities	0.72	0.76	0.74	0.59	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NEI-VFQ-25																								
General.Health	-0.31	-0.24	-0.25	-0.27	-0.26	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
General.Vision	-0.59	-0.57	-0.68	-0.41	-0.66	0.28	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ocular.Pain	-0.48	-0.45	-0.53	-0.45	-0.48	0.22	0.43	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Near.Activities	-0.73	-0.65	-0.81	-0.54	-0.72	0.29	0.75	0.53	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Distance.Activities	-0.67	-0.71	-0.79	-0.51	-0.7	0.17	0.7	0.48	0.82	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Social.Function	-0.63	-0.66	-0.77	-0.5	-0.76	0.15	0.66	0.4	0.81	0.8	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Mental.Health	-0.71	-0.71	-0.76	-0.73	-0.83	0.26	0.68	0.5	0.74	0.7	0.73	1	-	-	-	-	-	-	-	-	-	-	-	-
Role.Difficulties	-0.76	-0.68	-0.77	-0.61	-0.77	0.3	0.69	0.54	0.79	0.72	0.75	0.81	1	-	-	-	-	-	-	-	-	-	-	-
Dependency	-0.63	-0.71	-0.76	-0.61	-0.79	0.1	0.67	0.48	0.79	0.76	0.84	0.76	0.77	1	-	-	-	-	-	-	-	-	-	-
Driving	-0.47	-0.48	-0.47	-0.34	-0.46	0.2	0.48	0.22	0.45	0.54	0.49	0.45	0.44	0.47	1	-	-	-	-	-	-	-	-	-
Color.Vision	-0.49	-0.5	-0.64	-0.37	-0.64	0.12	0.57	0.4	0.7	0.68	0.83	0.6	0.66	0.75	0.41	1	-	-	-	-	-	-	-	-
Peripheral.Vision	-0.62	-0.74	-0.75	-0.47	-0.69	0.22	0.62	0.47	0.74	0.75	0.72	0.69	0.65	0.69	0.46	0.63	1	-	-	-	-	-	-	-
Composite	-0.77	-0.77	-0.86	-0.63	-0.83	0.36	0.81	0.6	0.9	0.87	0.87	0.85	0.87	0.86	0.65	0.78	0.82	1	-	-	-	-	-	-
Progression																								
Best Eye MD	-0.5	-0.6	-0.71	-0.41	-0.52	0.1	0.53	0.24	0.65	0.72	0.72	0.53	0.53	0.68	0.56	0.6	0.59	0.7	1	-	-	-	-	-
EQ-5D-5L																								
MOB	0.35	0.43	0.46	0.24	0.46	-0.19	-0.42	-0.25	-0.42	-0.48	-0.46	-0.46	-0.44	-0.46	-0.34	-0.41	-0.44	-0.52	-0.41	1	-	-	-	-
SEL	0.41	0.43	0.46	0.39	0.6	-0.24	-0.43	-0.37	-0.51	-0.54	-0.57	-0.52	-0.48	-0.55	-0.33	-0.54	-0.45	-0.59	-0.42	0.56	1	-	-	-
DA	0.56	0.65	0.67	0.45	0.74	-0.14	-0.6	-0.35	-0.64	-0.7	-0.78	-0.68	-0.69	-0.74	-0.46	-0.7	-0.62	-0.76	-0.68	0.59	0.65	1	-	-
P/D	0.3	0.38	0.29	0.29	0.36	-0.3	-0.17	-0.32	-0.25	-0.28	-0.25	-0.33	-0.31	-0.26	-0.25	-0.2	-0.27	-0.35	-0.21	0.5	0.35	0.33	1	-
A/D	0.41	0.45	0.37	0.54	0.49	-0.2	-0.37	-0.29	-0.46	-0.46	-0.47	-0.49	-0.42	-0.54	-0.15	-0.32	-0.39	-0.48	-0.37	0.32	0.43	0.5	0.4	1

3.4.2.2 Convergent Validity This sample of patients often identified concerns with the HUG-5 visual discomfort dimension as their best-eye MD increased $r = 0.50$ and concurrently elevated the NEI-VFQ-25 subscale for general vision $r = 0.59$. The HUG-5's measure of visual discomfort was correlated with, but did not meet convergent validity criteria with the NEI-VFQ subscale for ocular pain $r = 0.48$. When patients identified general vision concerns as measured by the NEI-VFQ subscale, the HUG-5 mobility dimension was also frequently elevated $r = 0.57$. When mobility was reported above none or slight concerns due to glaucoma, patients also tended to report difficulties with their peripheral vision $r = 0.74$ (VFQ) and best-eye MD $r = 0.6$, altering the free movement important to engaging in usual activities $r = 0.67$ (EQ-5D), near activities $r = 0.65$ (VFQ), and distance activities $r = 0.71$ (VFQ). Surprisingly, the EQ-5D-5L mobility dimension $r = 0.43$ did not meet convergent validity criteria with the HUG-5 mobility dimension. The HUG-5 daily living dimension was assessed to have a strong, positive relationship with all NEI-VFQ-25 subscales and EQ-5D-5L dimensions specified. Patients reporting daily living difficulties on the HUG-5 also tended to have greater VFL where the best-eye MD was strongly correlated $r = 0.71$. Not so surprisingly, these patients also tended to report general vision difficulties $r = 0.68$ (VFQ), problems engaging in near activities $r = 0.81$ (VFQ), and distance activities $r = 0.79$ (VFQ). These patients also reported needing more support of others to perform basic functions or usual activities $r = 0.67$ (EQ-5D), captured by the NEI-VFQ-25 subscale for dependency $r = 0.76$ and role difficulties $r = 0.77$. The HUG-5 dimension measuring emotion was assessed to have a strong, positive relationship with the NEI-VFQ subscale for mental health $r = 0.73$ and the EQ-5D dimension of anxiety/depression $r = 0.54$. When patients described their experience with how glaucoma has impacted their ability to engage in social activities, as measured by the HUG-5 dimension of social activities, patients reported reduced VFL best-eye MD $r = 0.52$ and general vision problems $r = 0.66$ (VFQ). Patients reported difficulties engaging in communal activities; specifying difficulties with near activities $r = 0.72$ (VFQ), distance activities $r = 0.70$ (VFQ), social function $r = 0.76$ (VFQ), and role difficulties $r = 0.77$ (VFQ). These patients reported an effect of their glaucoma progression or treatment on social activities also reported challenges with usual activities $r = 0.74$ measured by the EQ-5D-5L.

3.4.2.3 Discriminant Validity This sample of glaucoma patients did not concurrently elevate HUG-5 dimensions mobility $r = 0.24$, daily life activities $r = 0.25$, emotion $r = 0.27$, or social activities $r = 0.26$ when elevating the NEI-VFQ-25 measure of general health, meeting the criteria for discriminant validity. The fifth HUG-5 dimension, visual discomfort had a weak correlation with general health $r = 0.31$ (VFQ). These are similar results as the reporting in the NEI-VFQ-25 psychometric validation study.[70] HUG-5 dimension emotion did not meet criteria for discriminant validity with best-eye MD $r = 0.41$, general vision $r = 0.41$ (VFQ), self-care $r = 0.46$ (EQ-5D), and usual activities (EQ-5D) $r = 0.45$. However, emotion met criteria for discriminant validity with mobility $r = 0.24$ (EQ-5D) and pain/discomfort $r = 0.29$ (EQ-5D). The HUG-5 dimensions of daily living ($r = 0.29$) and emotion ($r = 0.29$) met discriminant validity criteria in their correlation with pain/discomfort dimension (EQ-5D-5L). Social Activities ($r = 0.36$), mobility ($r = 0.38$), and visual discomfort ($r = 0.3$) demonstrated a weak correlation, however these subscales did not meet criteria for discriminant validity. A correlation plot mapping convergence and divergence predictions while identifying relationships that did not meet criteria with an X is illustrated as Figure 7.

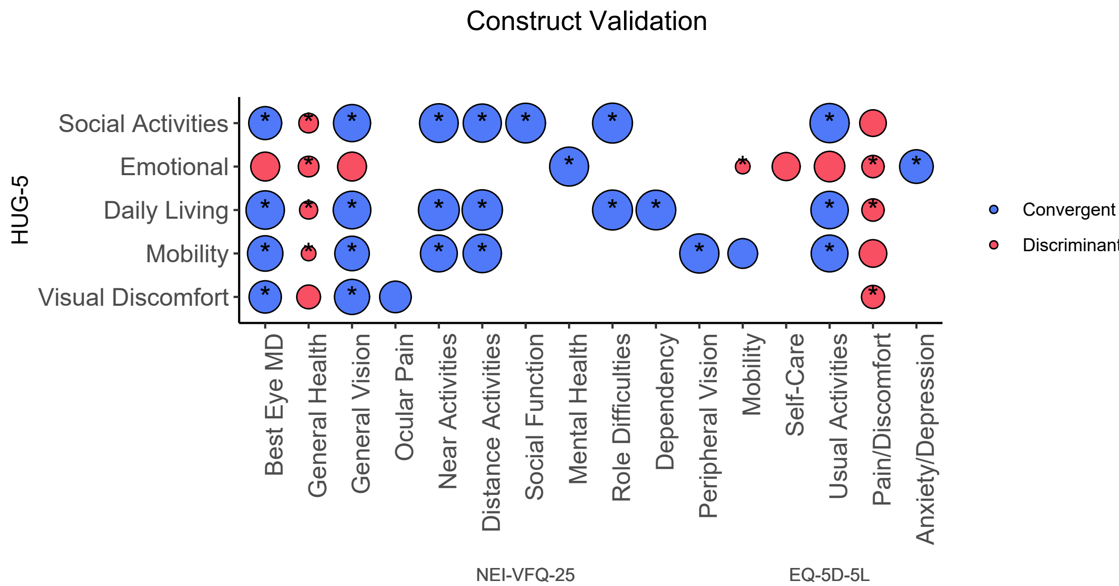


Figure 7: Correlation plot of HUG-5 dimension relationships between NEI-VFQ-25, Best-Eye MD, and EQ-5D-5L. The size of each circle is proportionate to the magnitude of the correlation. Significant convergent relationships where $r >= 0.5$ and significant discriminant relationships where $r <= 0.3$ are annotated with a $*$. Unmarked circles were apriori hypotheses that did not meet criteria.

3.4.3 Reliability

After a 2-week period, a random sample of 39 patients was administered the HUG-5 over the phone. The second HUG-5 composite scores were compared to the first administration, demonstrating excellent reliability, $F(38, 38) = 22, ICC = 0.91, (0.84, 0.95), p < 0.001$. The relationship between time 1 and time 2 HUG-5 administrations was evaluated using a 2-way random-effects model for single raters and conclude that 91% of the variability in the scores captured by the HUG-5 represented the constuct of glaucoma-specific HRQoL and 9% represented random variation. The HUG-5 dimensions were individually evaluated for an index of item reliability (rii). The most reliable dimension of the HUG-5 is daily activities with $rii = 0.88$, followed by social activities with $rii = 0.84$ and mobility $rii = 0.80$. The least reliable dimensions of the HUG-5 are visual discomfort with $rii = 0.77$ and emotion $rii = 0.73$.

3.4.4 Sensitivity

Patients classified as mild/moderate and advanced were compared using a Wilcox test, where the distribution of HUG-5 composite scores are illustrated in the Figure 8 density plot. The rank-sum test with continuity correction rejected the null, accepting the true location shift is not equal to 0 ($W = 693.5, p < 0.001$). This confirms that the HUG-5 is sensitive in detecting group level differences between mild/moderate and advanced patients with glaucoma.

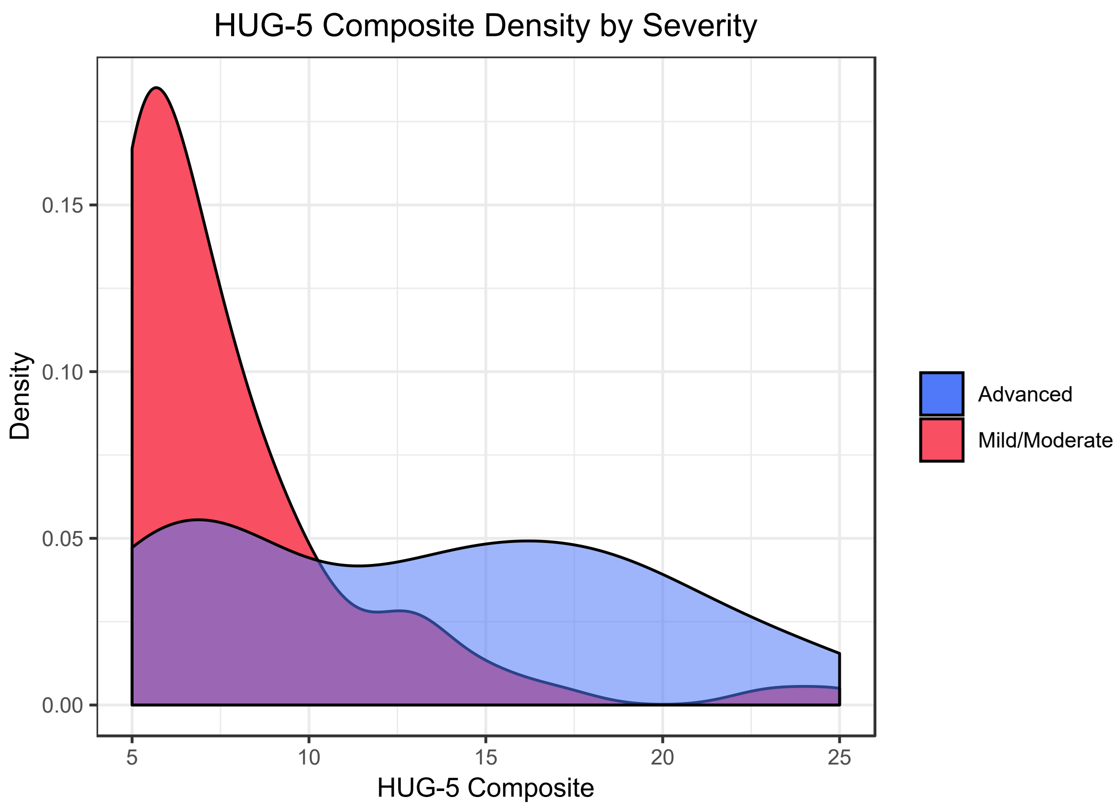


Figure 8: Distributions of HUG-5 composite score by best-eye glaucoma severity

3.5 Discussion

The present study sought to determine if the HUG-5 preference-based measure met criteria for validity, reliability and sensitivity in a population of patients followed for glaucoma management. This methodology was chosen for the purpose of assisting clinicians and researchers and informing their decisions regarding the use of the HUG-5 as an outcome measure. Since glaucoma is a slow progressive disease, excellent HUG-5 reliability indicate that the tool provides very consistent results for individuals in a stable disease state within 2 weeks of measurement. The HUG-5 can effectively discriminate between individuals in a mild to moderate stage of glaucoma versus advanced glaucoma, an important characteristic identified in the planning stages of measure design.[80] From the results of this study, it is clear that the HUG-5 demonstrated sufficient dimensional consistencies between NEI-VFQ-25, EQ-5D, and best-eye VFL to classify the HUG-5 as valid questions to measure visual discomfort, mobility, daily life activities, emotion and social activities as they pertain to a HRQoL state in a glaucoma patient population.

The NEI-VFQ-25 composite score had the most consistent relationship with best-eye MD (Figure 2), this is most likely reflective of the instruments use of subscaling and multiple items informing that composite. Additionally, the NEI-VFQ was designed with an intent to measure vision-specific HRQoL. Despite the imprecision in predicting best-eye MD, the HUG-5 has a strong, significant relationship with best-eye MD and met criteria for concurrent validity. Of the 35 construct validity comparisons, 31 met a priori criteria. The strong associations with NEI-VFQ and best-eye VFL demonstrate that the HUG-5 measures key elements in specifying HRQoL states in glaucoma, providing quantitative evidence to support our previous work in the development of the measure from patient perspectives.[16] A proposed explanation as to why the HUG-5 dimension of visual discomfort did not have a substantial correlation with the NEI-VFQ-25 subscale is that the HUG-5 embodies both physiological and psychological discomfort, whereas the NEI-VFQ-25 subscale measures physiological discomfort such as burning or itching of the eye exclusively. Ultimately there is imperfect yet acceptable evidence for the construct validity of the visual discomfort dimension. Visual discomfort should be further explored in it's ability to capture the components of psychological discomfort with vision (difficulty seeing in darkness) and physiological discomfort associated with glaucoma progression or therapies (burning or itching in and around the eyes). There seemed to be strong evidence supporting the mobility dimension from how patients report their HRQoL and ability to engage in autonomous free movement. For the dimensions of daily living, emotion, and social activities, all anticipated relationships between NEI-VFQ-25 subscales and EQ-5D-5L dimensions were observed, demonstrating excellent evidence for construct validity.

In a direct comparison with the internationally accepted generic measure EQ-5D-5L, the HUG-5 consistently captured health states that were not detectable by the EQ-5D-5L. In the case that both measures were administered to patients alongside a clinical trial to evaluate cost-effectiveness, patients would report a higher quality of life with the EQ-5D than the HUG-5 because the questions are designed to capture general quality of life. Measuring relevant dimensions to glaucoma patients allows for a more concise estimation of utility between individuals, providing the conditions necessary for precise estimates of patient health utility. Relative to the NEI-VFQ-25, the HUG-5 covers a wider range of possible health states and requires a fifth of the questions and time to administer. The HUG-5 was able to consistently measure key elements of health states for glaucoma patients, with excellent test-retest reliability across the 5 dimensions. On an item level, the means of all dimensions slightly decreased from the initial administration. This over-reporting effect may be attributed to the differing environments of administration (at clinic after appointment versus at home over the phone). In our measurement of HUG-5 sensitivity, the HUG-5 could differentiate HRQoL distributions of mild/moderate and advanced patients. This should be interpreted to the extent that the HUG-5 can detect differences in groups, but it will not consistently detect important differences as a clinical assessment of an individual patient.

An important piece of information regarding measurement quality is the responsiveness of a quality of life instrument. Responsiveness assesses the value created by treatment of glaucoma, if relieving glaucoma symptoms improves function, the HUG-5 should be able to measure that change. Due to sampling constraints, it was decided to evaluate the responsiveness of the HUG-5 in a longitudinal study among a homogeneous patient group, despite our initial plan to assess responsiveness. The HUG-5 will be administered alongside a randomized controlled clinical trial evaluating treatment efficacy as a secondary outcome measure and reviewed for responsiveness among patients with differing characteristics. The HUG-5 is intended to replace the EQ-5D or other multi-attribute utility instruments for measurement of utilities in glaucoma to inform cost-effectiveness analysis of introducing changes to treatment alternatives or service changes. The following discrete choice study will map utility values to HUG-5 health states from the generated multi-attribute utility function and placed on the health utility scale with a time-trade-off discrete choice experiment. This study was not without its limitations. This study did not account for the effect of comorbid conditions. Three patients were included in the analysis that were diagnosed with Juvenile Open Angle Glaucoma (JOAG). These patients were included as they met the prespecified criteria for inclusion and in the interest of patient equity for participation. However, these patients may perceive quality of life different from patients diagnosed at a later age. It is unlikely that these three individuals invalidated the conclusions of this study, however future research should aim to evaluate the validity of the HUG-5 in this population of glaucoma patients with a representative sample. Data was only collected from patients at one clinical group in Southern Ontario, more patients in different regions would be important to measure to generalize results to other populations. Regarding concurrent and discriminant validity, this study made numerous comparisons from threshold criteria without employing a statistical test to determine the likelihood of obtaining a true result. There were simply too many comparisons that were identified as important relationships to describe similarities and discrepancies captured in the reporting patterns present in the NEI-VFQ-25, the HUG-5 and the EQ-5D-5L. The NEI-VFQ-25 has met criticism in recent years for its internal consistency[85] other measures such as the GlauQol-36 may have provided additional information on the properties of the HUG-5. Future studies may look to collect more patients from different populations and test the performance of the HUG-5 relative to other generic and disease specific measures of HRQoL. Additional research is required to address the factors impacting measure properties in the detection of subtle HRQoL differences between mild and moderate patients. In this respect, investigating HUG-5 properties relative to the HUI III, which includes a vision component, is a worthwhile pursuit.[86]

3.6 Conclusion

The HUG-5 has met the expectations previously outlined[80] and presents a collection of evidence that supports its use as a preference-based measure of glaucoma health states. The final step of this 4-step process is to utilize a discrete choice experiment in conjunction with a time trade-off exercise with a glaucoma patient population to determine a preference-based scoring algorithm to measure health states on a health utility scale.

4 RESURRECTING MULTI-ATTRIBUTE UTILITY FUNCTION; DEVELOPING A VALUE SET FOR HEALTH UTILITY FOR GLAUCOMA (HUG-5)

Submitted for publication as: Kennedy, K., & Pickard, A.S., & Tarride, J.E., & Xie, F. (2022). *Value in Health.*

4.1 Foreward

In 2019 and at the point of publishing the HUG-5 validation study, I was still developing an understanding of preference-elicitation methodology and had ended the validation study with a statement towards using DCE to generate a value function for the HUG-5. In the design of this study, time trade-off (TTO) and discrete choice experiments (DCE) were considered as viable methods for generating a preference-based value algorithm for the HUG-5. However, we ultimately decided to adopt the multi-attribute disutility function (MADUF) approach.

The rationale for this decision was that despite TTO being a popular value-based framework used in defining preference-based value functions of the general population for other, similar, measures such as the EQ-5D, we considered that TTO foundationally was based on limited theoretical underpinnings.[87,88] Discrete choice experiments have grown in popularity over the past decade, with additional specifications to the experimental method including max-diff (best-worst scaling).[89–91] DCE questions require direct comparison of select health states, where the cardinal preferences of the sample are ranked on aggregate and fitted to a logit function between 0 and 1. DCE is designed to elicit choice within a random utility framework. We did not use the DCE method due to our belief that the preferences for attributes described in the HUG-5 health states would not be additive, expecting an interaction between attributes. The MADUF represented a framework that allowed for single attributes combined as discrete health states with interactions between attributes/levels. In addition, the MADUF utilities are elicited using Standard Gamble, a method design to identify indifference points between alternative health states and is closely design to reflect the tenants of expected utility framework. The SG-VAS MAUF approach allows for additive relationships (if observed) in calculating the preference-scoring algorithm, yet also retains capacity to specify multiplicative relationships in utility scoring for non-single attribute health states.

In the initial design of this experiment, we considered using SG only to develop a multiattribute utility function directly from elicited single attribute utilities (for all attribute levels). Unfortunately, in the pilot testing, we found that the range of SG derived utilities was insufficient to generate a reliable scoring function. We concluded that the iterative nature of the SG ping-pong approach did not allow for respondents to qualify precise differences between incremental single attribute health states. The consequence of eliciting all single state attributes was a lack of discrimination in aggregate estimates between attributes. We subsequently pilot tested a sample with both VAS and SG for single attribute states and found that the precision afforded by VAS to assign values at the upper end of the scale (without equivocating perfect health) supported reliable estimation of the function. We used an online data collection process in concert with a sampling service to collect the preferences of the United States general population. The online application was iteratively tested through piloting to ensure quality data collection with the final sample. Quota sampling was used to recruit a representative sample of the US general population in terms of age, sex, and race/ethnicity.

MADUF utilities were estimated for 3125 HUG-5 health states. The correlation between mean elicited and estimated values for marker states was strong ($R^2 = 0.97$) with $MAE = 0.11$. The HUG-5 value set offers improvements over other glaucoma-specific preference-based measures. First, it was developed from the preferences of a representative sample of the US general population and second, the HUG-5 health state values are on the scale of perfect health and death which can be used to estimate QALYs for economic evaluations.

4.2 Introduction

Cost-utility analyses (CUAs) use a generic or condition-specific, preference-based measure of health state to facilitate comparisons across diseases and clinical outcomes. Health utility is a single index measure anchored at 0 for dead and 1 for full/perfect health, with negative values for health states worse than dead.[92] Health utility has been widely used to calculate quality-adjusted life year (QALY) in CUAs and inform health resource allocation policy making. Alternatively, in the United States, CUAs have been used by Institute for Cost Effectiveness Review (ICER) to provide guidance on pricing under a value assessment framework that applies cost per QALY thresholds.

Health utility can be measured directly using preference elicitation techniques or indirectly using predeveloped, preference-based instruments. Direct preference-elicitation methods ask patients to assign value to their current health state. Preference-based instruments are scored with a preference-weighted function. This scoring function assigns each health state defined by the instrument to an index of health utility.[93,94] The preference-based scoring function is derived by selecting a subset of health states of interest, identifying a group of respondents, measuring preferences for the selected states, and modeling a value function for all health states described by the instrument.[95–97] A representative sample of the general population is a recommended source of preferences when developing a scoring function for a few reasons.[98] First, evidence has consistently shown that patients who have experienced an impaired health state may adapt to the state and thus assign a higher utility value than the general population who has no prior experience or knowledge of the state.[99] Secondly, in a publicly funded health care system it is the general public who sponsors the system through taxation, and is eventually affected by any resource allocation decision. Therefore, any such decision should be made based on the general public preferences.[100,101]

In certain conditions, there is evidence that suggests generic health utility measures inadequately capture the impact of treatment and outcomes. This is particularly an issue for vision-related diseases, where generic measures are unable to reliably detect differences between disease stages.[78] Glaucoma is different from acute or emergent ophthalmic conditions is that it occurs as a gradual progression from peripheral to central vision loss. Glaucoma is a progressive disease that occurs more often in aging populations and impacts the lives of 1.86% of the global population.[102] In the most common type of glaucoma, open-angle, vision loss progresses from peripheral visual field loss to central vision loss and possibly total blindness if untreated.[103] Patients, on average, are nearing blindness in both eyes before a generic instrument detects differences in health-related quality of life.[67,104]

In recent years, there has been growing interest in developing condition-specific preference-based measures (CS-PBMs) for conditions that involve deficits to sensory systems such as vision, hearing, and cancer.[105–109] CS-PBMs are developed as a utility-based alternative to generic preference-based measures when the latter is demonstrably limited for detecting change and/or differences in health states. Multiple valid and reliable condition-specific instruments exist for glaucoma such as the NEI-VFQ-25, and the GPI which measure glaucoma health states and assign utility values by either a mapping function or a value set.[31,77,110] However, limitations have been described for preference-based scoring previously developed for the GPI (GUI) and NEI-VFQ-25 mapping algorithm to EQ-5D utility scores.[75] The GUI's function was generated from the preferences of patients, not the general population, a significant association was observed between the patient's disease state and their preferences for health states. The NEI-VFQ-25 mapping algorithm reported low predictive power, where the authors highlighted that the mapping function may lead to inaccurate utility values.[75]

We have previously developed and validated the Health Utility for Glaucoma (HUG-5).[16,111] To address the need for an improved preference-based utility measure, the primary objective of this valuation study was to develop a scoring function to calculate health utility for health states described by the HUG-5 based on the preferences of the general population in the United States.

4.3 Methods

This was a valuation study aimed at developing a preference-based scoring function for the HUG-5. Two preference elicitation methods, visual analog scale (VAS) and standard gamble (SG) were used to elicit preferences for a subset of health states described by the HUG-5. A representative sample of the US general population was recruited through an online platform. A multiple attribute disutility function (MADUF) was adopted to develop the scoring function for the instrument.[86] The HUG-5 describes patient reported glaucoma health states as a composite of 5 possible levels of 5 attributes: visual discomfort (physiological and psychological), mobility concerns (limitations to spatial awareness), daily life activities (close up work, household chores), emotional health (negative thoughts, worries and fears), and social activities (getting out with friends, large social gatherings, incidents of social embarrassment).[16,80,111] These dimensions of health state focus on physical, mental, and social functions which are related to, but not defined in exclusivity by the progressive loss to the field of vision. In total, the HUG-5 describes 3125 health states.[16,80,111] A subset of the HUG-5 health states was selected to develop the value set.[94,112]

4.3.1 Health state selection

Single attribute health states refer to the health states in which only one attribute is elevated, and all other attributes are held constant at level 1 (e.g., 21111, 31111, 41111, 51111). This allows respondents to consider each attribute and level independent of other attributes. For each attribute, there are 4 single attribute health states (i.e., slight, moderate, very much, and severe) and hence the total number across five attributes are 20. The single attribute health state at the most severe level of the attribute is called corner state. There are five corner states for the HUG-5, namely, 51111, 15111, 11511, 11151, and 11115. PITS health state is defined by all attributes at the most severe level, describing the worst possible health state. For the HUG-5, the PITS state is 55555. Marker states are health states that reflect different levels of severity. Marker states facilitated a comparison between health utilities directly derived from SG and the utilities estimated from the scoring function. We selected five marker states based on common health states observed in the HUG-5 psychometric validation studies for patients with mild/moderate and severe glaucoma. Based on health state frequency of occurence and describing different combinations of attribute levels, we selected two mild/moderate health states (31232, 31221) and three severe health states (11545, 14314, 42512).

4.3.2 Preference elicitation methods

Visual analog scale (VAS) was used for deriving health utilities for single attribute health states. The anchors for the VAS are 0 for death and 100 for perfect health. The VAS was represented as a vertical line with a slider button, where the respondents can slide the button and view the current value of the slider to assign value to the health state. The slider was placed at 50 as the default value for each health state to avoid position bias. The VAS values were used to determine the distance between single attribute health states. Standard gamble (SG) is a preference elicitation method developed based on the expected utility theory.[113–115] SG tasks constitute respondents being offered a choice between an uncertain option of with perfect health and death with probability p and 1-p, respectively, and a certain option of living in a (sub-optimal) HUG-5 health state. The probability p started with 1 and then was iteratively varied using a ping-pong approach with 0.1 increments until either the respondent indicated indifference between the two options at which the utility for the HUG-5 health states was calculated as p, or the value reaches an end by design (e.g., between 0.8 and 0.9, therefore the utility value of 0.85 was assigned). The SG tasks were presented using an icon grid visualization of the probabilities[116], beginning with 100% chance of being in perfect health. Respondents assigned values for a warmup health state, the 5 single attribute corner health states, the 5 marker states, and death or PITS depending on which they deemed to be worse. A functional transformation was applied to estimate single attribute health state utilities with information from SG utilities and VAS values.

4.3.3 Sampling

A sample representative by age, sex, and race of English-speaking Americans was recruited through an online survey platform. The sample size required in health utility valuations of preference-based measures requires considering the level of precision sought for estimates and the degree of balance in alternatives presented to respondents.[94] For this study, we aimed at reaching a standard error of mean estimates at 0.05. To reach this level of precision for the sampled mean, we assumed an average standard deviation of 0.3.[96] The total sample size required to reach a minimum level of precision was n = 180. Based on previous work[117,118], we expected 32% of the total sample would not meet exclusion criteria, adjusting our minimum effective sample to n = 263. A priori exclusion criteria were established to ensure the quality of the scoring function estimated. Respondents were excluded if their subjective numeracy score was less than 2.5 and a self-assessed understanding of the tasks less than

2 on a scale of 1 to 5 (where 1 was defined as ‘no understanding’, and 5 was defined as excellent understanding).[119] Respondents were also excluded if they used less than 10% of the VAS or health utility scale or failed more than 2 VAS monotonicity checks across attributes.[118] A failure of monotonicity within an attribute was defined as VAS values not incrementally lower or the equal for less severe states. In addition to reaching the minimum effective sample, respondents were recruited to meet quotas for representation of the US general population.

4.3.4 Data Collection and Management

The online survey application was developed in R 4.1.1 and the shiny package.[120,121] SUPPLEMENT 4 to 6 illustrate the user interface used to elicit preferences. The application followed a prescribed sequence that informed respondents about glaucoma, encouraged reflection on what glaucoma would mean for them, and asked a series of preference-elicitation tasks data quality questions and demographics (Figure 9). The application source code is open access and is available for download on github.

Initially, respondents were briefed on the purpose of the study and provided an overview of the study tasks before being asked to consent to participate. Following a confirmation of their consent to participate, we provided information on the nature of glaucoma, common therapies, and described the 5 dimensions of the HUG-5. This description included the definition for each dimension and a narrative summary of how each health dimension is related to the experiences faced by patients with glaucoma. Providing a context of the health states valued is important for an effective valuation.[122] Respondents participated in a reflective framing activity to internalize basic knowledge of glaucoma, under the impacts of the disease and some considerations related to receiving ongoing treatment and generate a framework to support valuation of glaucoma health states. The reflexive framing task consisted of a series of open-ended questions where respondents described how the key attributes of HUG-5 health states would hypothetically affect them at 70 years old.

Respondents may decide that living with PITS is worse than dead. To account for this, respondents chose between PITS and death to determine group membership between Group A: prefer PITS to death and Group B: prefer death to PITS. Respondents were asked to complete 20 single attribute health states on the VAS. Each attribute was presented in random order. However, within each attribute, the single attribute health states appeared from slight problems to severe problems before moving on the next attribute. After completing the VAS questions, respondents were briefed on the purpose and sequence of the SG task. The series of SG tasks consisted of a warm-up health state (31121), five HUG-5 corner health states in random order followed by 5 marker health states also in random order. The final SG task was to use the least preferred health state (PITS or death) as the lower anchor to derive utility for the other state (death or PITS) according to the answer to the first question of the survey. Groups A and B assignment determined the last SG task. Group A respondents were assigned to a SG task for PITS between perfect health and death. Group B respondents were assigned to a SG task for death between perfect health and PITS.

Respondents were then asked to respond to a series of data control questions and demographic questions. The Subjective Numeracy Scale (SNS) is a self-report measure of perceived ability to perform mathematical tasks and preference for using numerical information.[119] A single question was asked to the participant to reflect identify the level of their perceived understanding of the tasks. Respondents were asked to identify their sex, age, overall health, marital status, if they knew someone with vision loss, if they have experienced vision loss, level of education, employment status, and annual household income. This study

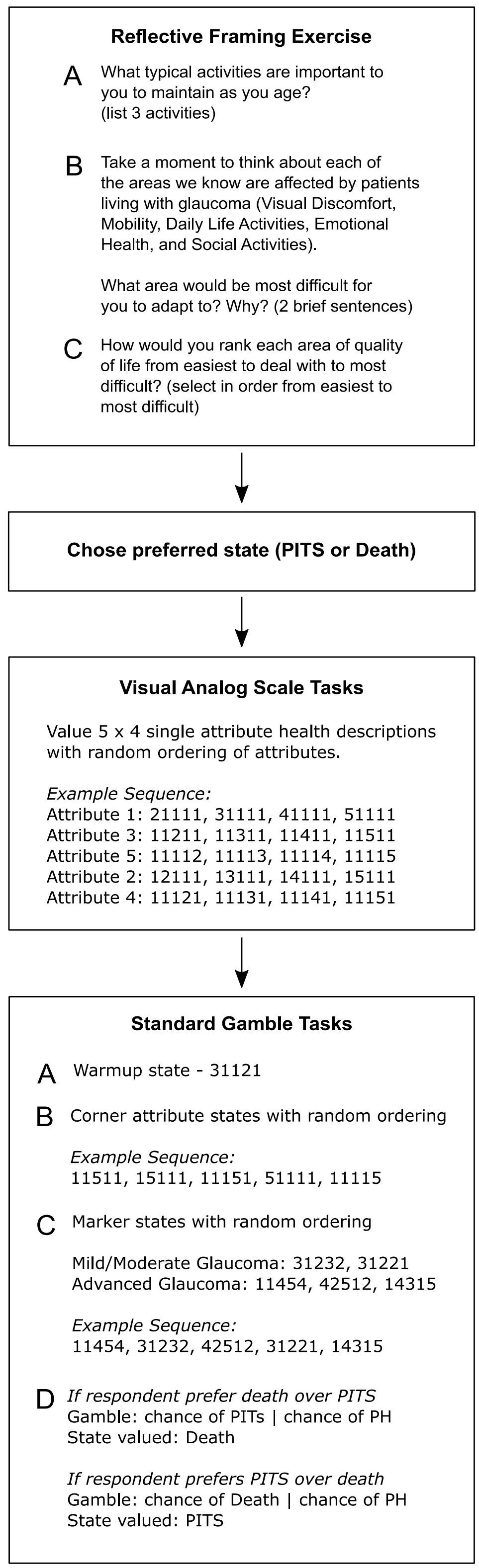


Figure 9: Experimental design and task sequencing.

was evaluated and approved by the Hamilton Integrated Research Ethics Board (HiREB, ref 2020-8247-GRA).

4.3.5 Multi-Attribute Disutility Function

The value set for the HUG-5 was generated from a ‘person-mean’ approach that models utilities from single attribute states, corner states and the PITS in a MADUF. The MADUF method estimates the health utility of multiple attribute outcomes as a function of the utility of each constituting attribute. The MADUF integrates the directly estimated corner state utilities with the estimated single attribute utilities converted from VAS values to produce an equation that calculates the disutility of all health states measured by the HUG-5 from 1 to 0 (PITS to perfect health).

4.3.6 Estimating utility from rescaled VAS

Respondents were assigned to one of the two groups according to their answer to the first question, Group A indicating death was the least preferred and Group B the PITS state. For each group, mean single attribute VAS scores were adjusted for the end of scale bias (EOSBA) using a positive linear transformation with the adjustment factor of 1.78.[123] The EOSBA rescaling only applies to the least severe level health state of each attribute (e.g. 21111), the more severe health states (e.g. 31111, 41111) are rescaled to maintain proportional distances between attribute levels. Proportional differences are calculated from the range of values from the original vector, determining the percentile that each state represents among that range. Then find the new range of values from most severe state to the mildest EOSBA adjusted state and multiply each state’s original percentile by that range and add to the most severe state. This adjusts the values while maintaining the original spacing of states relative to one another on a continuous number scale (see Supplementary File).

(EQ. 1)

$$VAS_{EOSBA} = 100 - (\frac{100 - VAS}{1.78})$$

After adjusting for EOSBA, mean single attribute VAS scores were scaled to health utilities from the corner states (e.g., 11151) and perfect health (0) using a power function.[124] Where $v = 1 - VAS/100$, u = disutility for single attribute health states, p is optimized to result in mean u scores from 0 to 1 and estimated from the relationship between mean SG derived disutility and the VAS EOSBA score for the corner states.

(EQ. 2)

$$u = v^p$$

The single attribute disutility for Group A and Group B were combined using a weighted mean, adjusted for the proportion of respondents in each group, for each single attribute state. Single attribute disutility values were then rescaled from 0 (no problems) to 1 (severe problems) within each attribute.[96]

The MADUF (EQ. 3) specifies how the corner state disutility values are combined with single attribute disutility to estimate the overall disutility of a health state $u(x)$. The x term represents values of j single attributes states x_j with u_j single attribute disutilities and k_j disutility of corner state for attribute j. The term $u_j(x_j)$ describes the single attribute disutility of a level of a single attribute state (e.g., for health state $j = 32415$, $u_j(x_j)$ represents the vector of disutility of single attribute states 31111, 12111, 11411, 11111, and 11115). The single attribute values are multiplied with corner state's disutility (k_j) which is calculated from $1 - u_j$, where u is the mean utility for the corner state directly estimated from responses to the SG tasks.

(EQ. 3)

$$u(x) = (1/k)\prod_{j=5}^{n}(1 + kk_j u_j(x_j)) - 1$$

To calculate disutilities using the MADUF of all HUG-5 health states, we solve for an interaction term k which captures the interaction in preferences among attributes and scales the MADUF to PITS (weighted disutility directly elicited from the death/PITS SG). Group A utilities for PITS state were summarized with a mean value and Group B values for death were summarized with a mean value. Group B mean disutilities for death state were elicited on a scale from 1 PITS to 0 perfect health and were rescaled to 1 death and 0 perfect health by calculating a scaling factor for a linear transformation. Group A and B mean disutility for PITS were subsequently combined in a weighted mean. (EQ. 4) substitutes the variable $u(x)$ for PITS, allowing a simplified expression without $u_j(x_j)$ where the single attribute disutility values for corner states are equal to 1 for all 5 attributes. The disutility values generated from the MADUF for all health states were converted to utility values by taking the difference of 1-MADUF to place health states on a scale from perfect health (1) to death (0).

(EQ. 4)

$$PITS = (1/k)\prod_{j=5}^{n}(1 + kk_j) - 1$$

4.3.7 Model fit

The model fit was assessed by comparing directly elicited utilities with the estimated utility values converted from the MADUF . The fit statistics included the overall square of the person product-moment correlation between observed and predicted (R^2), calculated across 5 marker health states. Additionally, the mean absolute error (MAE) was calculated to capture the magnitude of error between the predicted and observed utilities. We hypothesized similar error the observed MAE of the HUI3 ($MAE = 0.067$) and the observed MAE of the PORPUS utility weighting function ($MAE = 0.10$).[96,125]

4.4 Results

A total of 634 respondents completed the online valuation tasks between September and November 2021 Among them, 29 reported a subjective numeracy score less than 2.5, 7 reported a task understanding less than 3, 94 failed more than 2 tests of VAS monotonicity, 89 used less than 10% of the scale. 218 of the 634 respondents were excluded for failing to utilize the VAS scale, lack of discrimination between health states in SG task, low task understanding, and subjective numeracy (Figure 10). The analysis sample included n = 416 respondents. Respondents reported a mean (standard deviation) age of 42.34 (16.82) years, 57% identified themselves as female and 78% as White. The median duration of the time taken to complete all valuation tasks was 22.6 minutes (Q1, Q3: 16.8, 31.4). 28% of the sample had high school or lower education, 18% a college diploma, 31% a bachelor university degree, 15% graduate studies, and 11% other post-graduate programs. 47% of the effective sample reported full-time employment, 17% part time, 11% student, 7% on leave, and 18% reported being retired. Most respondents reported no chronic health problems (68%), with the remaining 32% reporting 1 or more chronic health problems. 54% of respondents reported having family members with vision related problems, and 4% having a diagnosis of glaucoma. 13% of the analytical sample reported 'Fair' or 'Poor' general health, 87% reported 'Good', 'Very good', or 'Excellent' general health. 4% of the respondents reported a 3 for task understanding, 22% reported 4, and 73% reported 5 out of 5. The median subjective numeracy was 6 (4.33, 5.67) out of 6 (Table 6).

4.4.1 Constructing the MADUF

Of 416 respondents, 260 (62.5%) chose death as worse than PITS and were assigned 'Group A', 156 (37.5%) were assigned to 'Group B'. Mean utility for PITS in Group A was 0.414. Mean utility for death in Group B was measured to be 0.362 (on a scale of perfect health and PITS). The scaling factor to transform the value of PITS on the perfect health to death scale for Group B was 1.57. The PITS mean disutilities for Group A and B were weighted by sample size n, resulting in $u_{pits}(55555) = 0.95'$, informing constant $k = -0.929$. The MADUF values were applied to 3125 health states with (EQ.3) and converted to utility values (1-disutility), ordered from the highest to lowest utility (SUPPLEMENT 7).

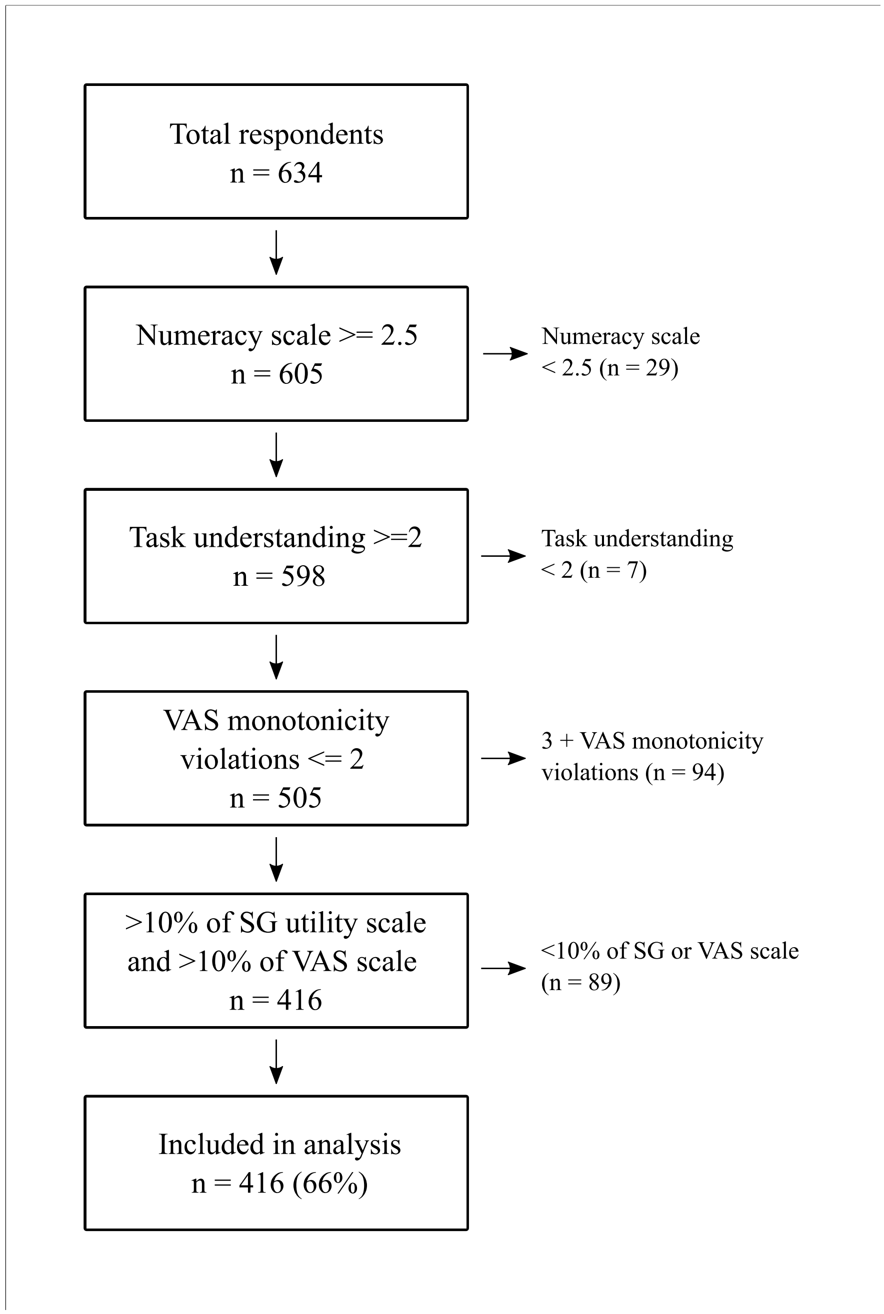


Figure 10: Exclusion flow diagram.

Table 6: Descriptive characteristics of analytical sample, stratified by preference groups (death worse than PITS state - Group A and PITS state worse than death - Group B).

	All respondents (n =416)	Death worse state Group A (n =260)	PITS worse state Group B (n =156)
Age, n (%)			
18-30	135 (32.5)	60 (23.1)	75 (48.1)
31-45	94 (22.6)	59 (22.7)	35 (22.4)
46-60	104 (25)	77 (29.6)	27 (17.3)
61+	83 (20)	64 (24.6)	19 (12.2)
Gender, n (%)			
Female	238 (57.2)	148 (56.9)	90 (57.7)
Male	177 (42.5)	111 (42.7)	66 (42.3)
Race, n (%)			
White	323 (77.6)	210 (80.8)	113 (72.4)
Black	23 (5.5)	14 (5.4)	9 (5.8)
Latin American	30 (7.2)	15 (5.8)	15 (9.6)
Asian	19 (4.6)	10 (3.8)	9 (5.8)
Mixed	7 (1.7)	2 (0.8)	5 (3.2)
Other	5 (1.2)	3 (1.2)	2 (1.3)
Native American	7 (1.7)	5 (1.9)	2 (1.3)
Education, n (%)			
High school or lower	99 (23.8)	55 (21.2)	44 (28.2)
College diploma	76 (18.3)	52 (20)	24 (15.4)
Bachelor University	130 (31.2)	84 (32.3)	46 (29.5)
Graduate Program	64 (15.4)	41 (15.8)	23 (14.7)
Post-Graduate Program	45 (10.8)	27 (10.4)	18 (11.5)
Employment Status, n (%)			
Full-time	194 (46.6)	116 (44.6)	78 (50)
Part-time	69 (16.6)	41 (15.8)	28 (17.9)
Student	46 (11.1)	21 (8.1)	25 (16)
On leave	28 (6.7)	22 (8.5)	6 (3.8)
Retired	76 (18.3)	59 (22.7)	17 (10.9)
Household Income (USD), n (%)			
< 15,000	50 (12)	27 (10.4)	23 (14.7)
15,000 - 45,000	136 (32.7)	81 (31.2)	55 (35.3)
45,000 - 75,000	99 (23.8)	64 (24.6)	35 (22.4)
> 75,000	129 (31)	87 (33.5)	42 (26.9)
Diagnosed with glaucoma, n (%)			
Yes	17 (4.1)	9 (3.5)	8 (5.1)
No or not sure	397 (95.4)	250 (96.2)	147 (94.2)
Family members with vision related problems, n (%)			
None	193 (46.4)	117 (45)	76 (48.7)
1	117 (28.1)	74 (28.5)	43 (27.6)
2	69 (16.6)	44 (16.9)	25 (16)
3 or more	34 (8.2)	23 (8.8)	11 (7.1)
Chronic health problems, n (%)			
None	283 (68)	180 (69.2)	103 (66)
1	78 (18.8)	44 (16.9)	34 (21.8)
2	32 (7.7)	21 (8.1)	11 (7.1)
3 or more	22 (5.3)	15 (5.8)	7 (4.5)
General health, n (%)			
Poor	10 (2.4)	6 (2.3)	4 (2.6)
Fair	46 (11.1)	34 (13.1)	12 (7.7)
Good	124 (29.8)	77 (29.6)	47 (30.1)
Very good	146 (35.1)	87 (33.5)	59 (37.8)
Excellent	88 (21.2)	56 (21.5)	32 (20.5)
Task understanding, n (%)			
3/5	18 (4.3)	13 (5)	5 (3.2)
4/5	91 (21.9)	57 (21.9)	34 (21.8)
5/5	305 (73.3)	189 (72.7)	116 (74.4)
Median subjective numeracy (Q1, Q3)			
Score	5 (4.3, 5.7)	5.3 (4.3, 5.7)	5 (4.3, 5.7)
Median monotonic value violations (Q1, Q3)			
Count	0 (0, 1)	0 (0, 1)	1 (0, 1)

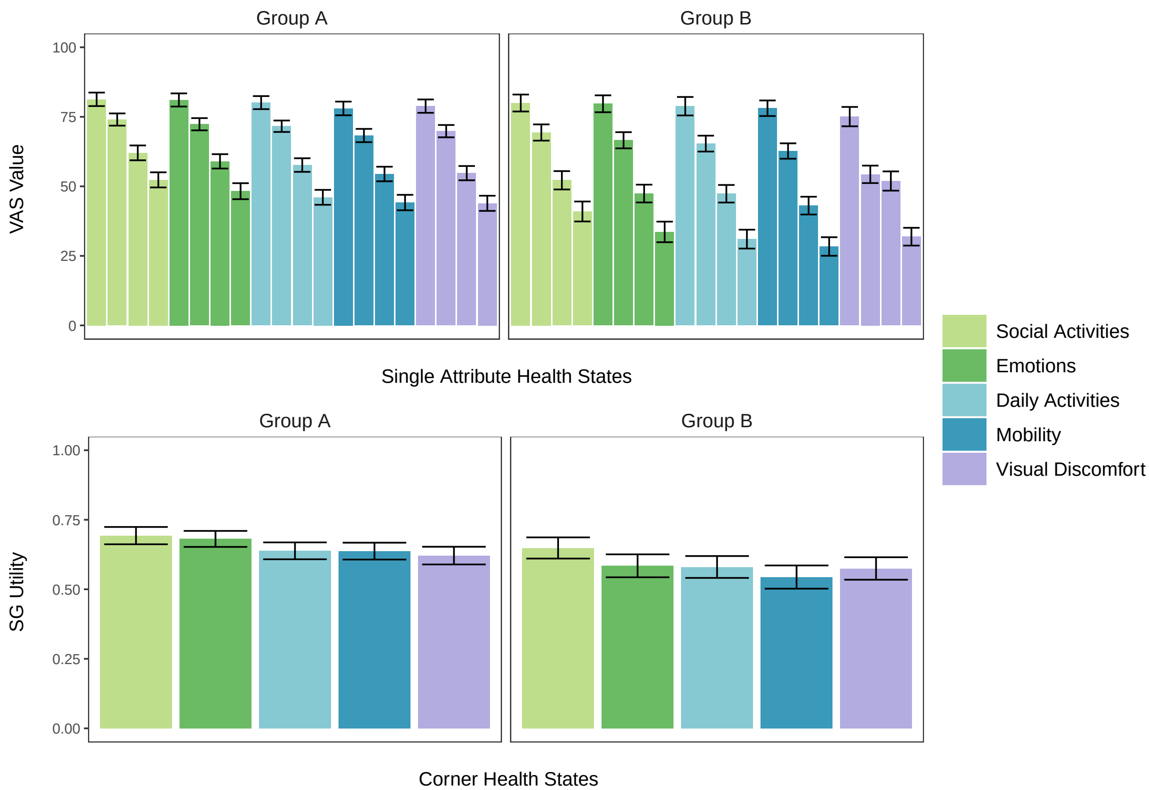


Figure 11: (a) Directly estimated VAS respondents for each single attribute health state with a reference for the default value stratified by preference groups. (b) Directly estimated utilities elicited from SG task.

Table 7: Single attribute scaling properties by measure dimension.

	Visual Discomfort	Mobility	Daily Activities	Emotions	Social Activities
$p(u = v^p)$					
Group A	0.60	0.57	0.60	0.55	0.60
Group B	0.49	0.49	0.47	0.50	0.49
Single Attribute Disutility Scaling (u_j)					
1	0.00	0.00	0.00	0.00	0.00
2	0.18	0.17	0.16	0.16	0.18
3	0.43	0.39	0.35	0.37	0.37
4	0.66	0.71	0.67	0.70	0.72
5	1.00	1.00	1.00	1.00	1.00
Disutility (k_j)					
Corner States (Level 5)	0.40	0.41	0.39	0.37	0.33

Table 8: Mean SG Value, estimated utility, and mean absolute difference by marker state.

Marker State	Mean SG Value	Estimated Utility	SD SG Value	Mean Difference	Mean Absolute Difference
11545	0.52	0.28	0.27	-0.23	0.23
14314	0.55	0.46	0.25	-0.10	0.10
31221	0.68	0.72	0.23	0.05	0.05
31232	0.65	0.62	0.23	-0.03	0.03
42512	0.54	0.38	0.25	-0.16	0.16

4.4.2 Model fit

Overall, the correlation between mean elicited and estimated values for marker states was strong ($R^2 = 0.97$) with $MAE = 0.11$. The most significant contributors to the differences elicited and estimated for marker states were the states selected to reflect advanced glaucoma (11545 and 42512). The largest MAE of 0.23 was observed for 11545 (Table 8).

4.5 Discussion

The primary objective of this study was to generate a preference-based value set for HUG-5 health states using the multi-attribute utility function method. The value set can be used to calculate all HUG-5 health state utility values from best possible (11111) to worst possible HUG-5 state (55555) on the perfect health (0) to death (1) scale. We found that despite challenges with SG task complexity, eliciting preferences from an online application was an efficient method for generating the HUG-5 scoring function for the US. The HUG-5 and the value set offer two major improvements over other glaucoma-specific preference-based measures. First, it was developed from the preferences of a representative sample of the US general population and second, the HUG-5 health state values are on the scale of perfect health and death which can be used to estimate QALYs for economic evaluations.

Relative to currently available utility estimates, the value set for the HUG-5 was developed based on the preferences of the US general population. This is in direct contrast to the only other glaucoma utility instrument Glaucoma Utility Index (GUI) that utilized a max-diff discrete choice methodology to elicit preferences from patients for disease states. In this study of the GUI, the authors identified a significant limitation to policy makers (who evaluate HTA), where respondent preferences were associated with levels of disease severity. The consequence of the relationship between disease severity and preferences for the GUI was an upward bias of utility values for more severe states. Studies that have collected utilities of patients with glaucoma through direct elicitation have elicited preferences on a scale between perfect vision and blindness. Therefore, they cannot be directly used as quality weights in the calculation of QALY and to subsequently inform insurance policy making. Preferences for health state on a vision scale reduces the effect of glaucoma to a single dimension of health. The HUG-5 describes health states with 5 dimensions, encompassing aspects of patient experience beyond their visual sensory ability. The HUG-5 measures broader dimensions of patient experience that are directly attributed to glaucoma by patients.[16] This preference weighting algorithm for the HUG-5 supports CUA of glaucoma interventions alongside clinical trials to improve resource allocation decisions in the United States.

4.5.1 Limitations

There are a few important limitations to consider in the interpretation of our study. First, the MAE goodness of fit was influenced by the decision to include more severe health states as markers. From the mean difference results between estimated and elicited utilities of marker states, the MADUF assigned greater values to mild/moderate glaucoma marker states and assigned lower utility values to advanced glaucoma marker states. Retrospectively, including an additional mild/moderate state to evaluate MAE would have led to a less biased estimate of error and allow further insight into model fit. Future studies should include additional marker states, sampled from within quantiles of HUG-5 utility values. Second, iterative tasks common in health state preference research have been associated with fatigue effects and learning effects.[126] Decisions of rules set for iteration sequence and limits on stopping can bias respondents and prevent them from expressing their 'true' indifference point.[127] The 'ping-pong' approach was employed in this study, however, with online administration and use of computer programs affords the possibility of optimal iterative sequences which may result in more efficient and reliable estimates of indifference points. Third, the respondents included in the analysis sample was not representative of Latin and Black Americans, oversampling White Americans. Fourth, the SG approach has been criticized for scale-comparability bias (where respondents inflate utility estimates from focusing on the bad-outcome probability opposed to the good-outcome probability)[128], poor internal consistency[129], and sensitivity to the gambling effect.[130]

4.6 Conclusion

This study generated a value set for the HUG-5 based on the US general population preferences, revitalizing use of the MADUF method in health state utility valuation. The value set allows for converting the health states described by the HUG-5 into health utilities, which can be used in the calculation of QALY for economic evaluations in gluacoma.

5 CONCLUSION

The issue of choosing which utility or value for what health state to inform cost-utility analyses does not have a silver bullet. No single preference-based measure can completely capture the clinical presentation of the physical state, the perceived emotional/social experience of the patient, the level of function, and the interactions that affect attributes of patient health.[131] This dissertation presents one approach to measuring glaucoma health states scored with utilities that reflect the preferences of the US general population. This final chapter discusses the implications of this work on economic models of glaucoma interventions and future research applications that builds on the foundation of these three studies.

In the following section of this chapter, the contributions of these works to health research methods and health economics are discussed in more detail. Subsequently, the implications of these works on cost-utility studies of glaucoma interventions and preference-elicitation methods are presented and aims for future research are considered.

5.1 Research contributions

The primary goal of this dissertation was to describe how we can improve the quality of evidence for utilities and values of glaucoma health states. Thus, the studies included in this dissertation extract from and make significant contributions to several research areas including: cost utility analyses of glaucoma interventions, measurement of glaucoma health states, preference elicitation with MADUF methodology, and valuing condition-specific health states with a sample from the general population.

5.1.1 Cost utility analyses of glaucoma interventions

In the first study, we summarized the growing literature of cost-utility analyses of glaucoma interventions. Building on previous studies, we found that most studies included in the review applied a wide range of utilities and values within glaucoma health states. Utilities and values were elicited on multiple scales and reflected the preferences of both the general population and patients. We described how authors choose to include utilities elicited anchored on 2 or more of: perfect vision/blindness, perfect health/blindness, and the policy scale perfect or full health/death. Within the reviewed studies we observed a clear pattern of authors continuing to apply early utility estimates without reviews or expansive searches for relevant utilities. This observed practice does not align with modern good practices of health utility use in cost-utility analyses.[28] major limitation of the reporting in most, if not all studies included was transparency. Justifications for use of certain utilities were rarely incorporated into the published reports and use of systematic search strategies to identify alternative estimates to inform model parameters was infrequently applied. The utility studies sourced by the CUA conveyed limited reporting on sequencing or iteration to determine the point of indifference that is characteristic of HSU and most often elicited utilities from a single question posed to patients. Within the utility publications that form the basis of mapping algorithms and utility value estimates, reporting is severely limited on important technical, ethical and practical obstacles performing TTO or SG with patients. In the review article we describe why utilities with 'health policy' anchors (perfect health and death) are important and how changing the anchor points can have a significant impact on the validity and conclusions of cost-utility analyses. This is a significant contribution to the literature as this article describes where improvements can be made to improve the transparency of reporting and support improved rigour of economic models of glaucoma interventions by paying special attention to the primary outcome quality-adjusted life years, where the utilities come from and if they are appropriate for the decision context of the study at hand. This study design can also be applied to other domains to identify how CUA in other disease areas identify and apply utility values.

5.1.2 Assigning value to glaucoma health states

In the third study we observed that when comparing the PITS state (worst possible state described by the HUG-5), general population respondents assigned utilities on average much lower than estimates from the available literature for 'severe glaucoma' (also described as blindness in both eyes) posits (from the first study). We believe that this observation is a consequence of the two key components that the HUG-5 addresses. First, the preferences are collected from the general population, in comparison with preferences directly elicited from patients. There is evidence that supports that patients consistently report higher utilities than the general population.[132] Second, the attributes of visual discomfort, emotional health, and social activities, measured by the HUG-5, are not described by the HPA definition of health state progression from mild glaucoma to blindness. HPA health states are exclusively defined by changes to visual field and eventual central vision loss. The HUG-5 attributes describe aspects of life that are complex and not fully accounted for by blindness alone.

There is ongoing debate over whether to collect preferences of patients who are affected by the disease or the preferences of the people who are paying for the care. This becomes more complicated depending on the reimbursement system. Preference source has been identified as a significant shortcoming of the quality adjusted life-year (QALY) measurement, especially in the United States.[133] There are important differences in application and use of patient versus general population preferences. Patients have lived experience and can acutely reflect on aspects of disease progression difficult to capture in a vignette. The general population without first-hand experience with a condition often form preferences at the time of preference valuation and therefore rely on information from the elicitation study to decide their preferences.[132] The participants in the third study were required to learn and understand consequences of glaucoma quickly. Participants were required to integrate the future consequences of living with glaucoma with their values and activities before assigning values and determining utilities for the study health states. An important contribution of the third study is the method introduced to support participants in understanding glaucoma health states. Within the preference elicitation study of a glaucoma-specific measure, activities and vignettes were incorporated to qualify the condition and ensure participants have an understanding of how the condition would impact them and their lives. This study offered a novel approach to assist in participant comprehension of glaucoma health states.

The primary goal of clinical trials is to assess whether a therapy or treatment is effective, with the treatment benefit being caused by the therapy.[134] To detect changes, measurements must be sensitive to health state changes and meaningful to how a patient with a condition feels or functions. Without being able to detect meaningful change, the conclusions of the trial are limited, and further economic evaluations become inconsequential. Decision maker's preference for generic measures (e.g. EQ-5D) require use of measures that lack sensitivity to detect differences in the feeling and function important to patient's daily lives in glaucoma over time. The impacts of side effects and adverse events related to glaucoma therapies are not detectable by generic measures of glaucoma health states.[78] This dissertation makes substantial contributions to glaucoma health state measurement as we have assessed the validity of the HUG-5 and assigned a preference-weighted algorithm to score health states between 0 (dead) and 1 (perfect/full health). The papers included in this dissertation build on the available literature and describe a measure that can be adopted in both clinical practice and trials to measure glaucoma health states. For future CUA models of glaucoma interventions with published utilities, investigators should consider the decision context and available values that reflect the preferences of their region. If none are available, HUI-3 utilities are preferred over EQ-5D-5L until such time as a bolt-on vision dimension is added to the EQ-5D. If investigators are able to collect utilities from a sample of patients, HUG-5 should be administered with a generic instrument. Investigators should also take care to report the rationale for using the utilities and, at a minimum, identify the scale of measurement and preference source before assigning to glaucoma health states.

5.1.3 Preference elicitation with MADUF methodology

The third study revisits the MADUF method developed over 20 years ago with a modern data collection strategy of online sampling. It is common today for preference weighting of generic measures to adopt the econometric approach. However, the multiple attribute utility function was developed to align preference-elicitation with expected utility theory.[101,135–137] This study makes significant contributions to the area of research in applied MADUF methodology, where the HUG-5 is one of few instruments that have chosen to take this approach.[125,135,138] During the development of the application, an alternative 'standard gamble only' method was attempted to (1) better align the value function with directly elicited utilities and (2) simplify analyses. However, this method was limited, where respondents were unable to differentiate between incremental levels of attributes (e.g., gamble for single attribute state 11211, followed by 11311). This resulted in estimating an MADUF that did not differentiate between health states, failing to impose cardinal ordering, a key axiom of von Neumann expected utility theory.

Whether or not to apply inclusion criteria in preference-based utility weighting investigations is a hotly debated topic. On one hand, setting inclusion criteria helps to identify and exclude respondents who may have not understood the tasks or engaged faithfully with the experiment tasks. On the other, there are reasonable explanations for preferences for ‘worse’ health, such as likelihood to receive remuneration from insurance or time off from work. We decided to use an online sampling platform to elicit preferences of the US general population for glaucoma health states. The sampling platform represents an income for participants, where they are reimbursed for their time at a pre-specified hourly rate. The platform itself was not infallible, where respondents could create multiple accounts, extend experiment duration to increase chance of receiving larger reward, and complete multiple studies at once, with divided attention. We found that while an online platform improves the quantity of data, increasing access to people of various background, ethnicity, and age, the inherently impersonal engagement with respondents can also increase the chance of respondents engaging in exploitative behaviours. For others considering use of online platforms to generate a sample of the general population, carefully monitoring data quality and applying inclusion criteria are effective means at minimizing the impact of these behaviours on estimating a preference-weighting function using the MADUF approach. In supplementary file 11, the impacts of these inclusion criteria on the final form of the HUG-5 are illustrated in a series of plots, where the MADUF was generated with respondent data that met the inclusion criteria of each run. Ultimately, a better fit for the MADUF model would have been achieved had we only applied a stringent VAS monotonicity, improving the MAE fit of the function from 0.114 to 0.102 and improving the representation of respondents with the effective sample size increasing from 416 to 469.

5.2 Practical implications

This dissertation was designed to build on the available evidence to improve the practice of measuring glaucoma health states. Therefore, there are two primary practical implications to consider as a consequence of these three studies.

First, we outlined the issues of health utility reporting in glaucoma cost-utility studies and directed clinical investigators and economists to ISPOR good practices resources aims to have a positive impact on the transparency, confidence, and comparability of future cost-utility studies. Second, the HUG-5 meets the need of providing a uniform, consistent measure of glaucoma health states with a preference algorithms that describes the preferences of the US general population. The preference-algorithm of US general population was chosen to encourage uptake in the use, where the US is by far the largest purveyor of clinical trials that measure the efficacy of glaucoma interventions. The utilities of the HUG-5 are intentionally anchored to be relevant to policy decision makers (perfect health and death). The HUG-5 can detect important differences in glaucoma patient health state among patients with differing levels of severity. The HUG-5 can be administered in clinical trials to facilitate cost-utility comparisons of trial-based data and describe preference weighted utilities that can be used in future economic models.

5.2.1 CS-PBMs in decision making contexts

Condition-specific preference-based measures have existed in glaucoma and other disease areas. However, few HTA authorities have recognized their relevance beyond as values to be used in sensitivity analyses of the ICUR in CUA. Generic measure of health state allows for comparisons between different disease states to inform resource allocation decisions. However, if a generic measure cannot detect major differences between patient glaucoma health states, it is exceedingly unlikely to detect more subtle changes that improve the quality of patients lives, while also reducing chance of progressive vision loss. The HUG-5 is unique as each attribute highlights the impact of the disease while asking patients to determine the degree of impact on each attribute. This quality of the HUG-5 is reinforced and guided by the descriptive examples of common, related, experiences that were reported by patients when reflecting on each attribute. Adoption and subsequent use of the HUG-5 in HTA decision processes will require going against accepted practice for HTA reporting. This is necessary to support identifying effective care alternatives and determine incremental value of glaucoma interventions. Future directions to support ongoing use and adoption will require working closely with ophthalmologists, engaging them in further HUG-5 research. It will be important to identify the degree of responsiveness of the measure due to glaucoma interventions and differentiate the qualities of the HUG-5 relative to the EQ-5D for investigators in important clinical features of patients with glaucoma. Knowledge translation and discussions with HTA decision makers will be integral in encouraging use of the HUG-5 for the base case of glaucoma intervention CUA.

5.3 Future research

The next step of measure development is to evaluate the use the HUG-5 in a clinical trial and within clinical practice. Longitudinal, controlled investigations are required to 1) evaluated the responsiveness to changes and 2) determine minimum important difference (MID). A study is currently underway to assess the feasibility of measuring quality of life between clinic visits. A large scale, properly controlled, clinical trial or a longitudinal observational study with change in quality of life as the primary outcome is required to adequately assess responsiveness of the HUG-5. Further work may generate additional value functions with the online collection application for other regions. An R package 'hug5' is in development to be hosted on the R programming CRAN repository with simplified look up of utilities for all health states, an HTML interface for collecting HUG-5 responses and a pdf of the HUG-5 measure for open-source, unrestricted use.

Real world evidence is growing in use with increases in demand for monitoring performance of healthcare providers. With the growth of technological services supporting medical interventions, there is an opportunity to incentivize quality of care and associate reimbursement to performance. This can have direct consequences for the patient and support informing choice of provider with effective use of large-scale data of patient populations. An investigation is currently underway to explore the feasibility of collecting routine data between patient visits in a clinic in Southern Ontario. Based on the feasibility of implementing of this program, an additional study will be planned to evaluate the responsiveness of the HUG-5 applying this method of in between visit data collection.[139] Secondary analysis of the data from study 2 is underway to describe the associations of clinical presentation, treatments, and other important variables contrasting the linear relationships between these factors and EQ-5D-5L/HUG-5 utilities.[140] Future studies may also investigate the spillover effects related to the burden on caregivers and families in supporting patients with glaucoma, characterizing these effects attributed to HUG-5 glaucoma health states.

Future studies considering the MADUF approach should pay special attention to how to best measure Death relative to PITS state. Our unique design used a discrete choice task that determined which gamble the respondent would be presented with (value PITS or value Death). We performed a positive linear transformation to estimate the average utility of where PITS is relative to death, weighted by the sample size of those who prefer Death or PITS. Where PITS is worse than death, we found that the bisection alteration of the ping-pong approach may not be optimal for valuing health states where the utility would be expected to be on the lower end of the scale. Alternative approaches to iteration to determine the point of indifference should be considered for more robust estimates. Future studies using this approach to generating a preference-weighting function for descriptive systems should explore VAS monotonicity violations for inclusion criteria as a first stop. The online application is shared as open source. We encourage others to modify components of the application to suit their individual needs in applying the MADUF method. The goal of open sourcing the application is to reduce the burden for others considering designing an experimental application and encourage further research into the MADUF method, comparing it with other preference-based utility value frameworks.

Future research should attempt alternative methods and describe their efficacy in general population comprehension of glaucoma health states. This work would provide much needed insight into valuing glaucoma-specific preference-based measures with the preferences of a general population. Future investigations of HUG-5 preference weighting algorithms in different cultural contexts and translations of the measure would improve the relevance of HUG-5 glaucoma health state utilities to different decision contexts. Future studies should investigate the role of caregivers and population utility decrements that are a consequence of caring for patients with glaucoma.

APPENDICES

SUPPLEMENT 1. Review: Search strategy for systematic review of glaucoma cost-utility studies, applied in OVID

Search	Query
	Glaucoma-related search terms
1	"Glaucoma, Angle-Closure" [Title]
2	"Glaucoma, Neovascular" [Title]
3	"Glaucoma" [MeSH]
4	"Glaucoma Drainage Implants" [Title]
5	"Glaucoma, Open-Angle" [Title]
6	"Glaucoma Drainage Implants" [Title]
7	"Glaucoma, Open-Angle" [Title]
8	"Low Tension Glaucoma" [Title]
9	"Pseudoexfoliative Glaucoma" [Title]
10	"Mixed-mechanism Glaucoma" [Title]
11	OR/1-10
	Cost-utility economic search terms [Title]
12	"cost-benefit analysis" [MeSH]
13	"Cost-effectiveness analysis" [Title]
14	"Cost-utility analysis" [Title]
15	OR/12-14
	Limits
16	English [lang]
17	11 AND 15 AND 16

SUPPLEMENT 2. Review: Fields and descriptions that constitutes the review data extraction sheet, SpRUCE checklist items were scored with 0 (not present) and 1 (present).

Study Characteristics Extraction Fields

Study	**Year**	**Country**	**Intervention**	**Comparator**
First name of author, year and bibtex reference key.	Year of publication	Setting of CUA	Primary program, intervention, treatment conditior or combination of treatments under consideration	Reference for comparison, usually the curren standard of care

Study Population	**Perspective**	**Economic Model**	**Time Horizon - years**
The characteristics of the population under consideration for the CUA	The context that determines which outcomes (costs/utilit are accounted for within the model. The perspective i closely linked to the decision context of the CUA.	The type of model used to estima future consequences of interventi relative to the comparator. (Usual markov state transition, decisior tree, microsim, or discrete event sim)	The base case time duration under study

Discounting	**Sensitivity Analysis**	**Main Findings**	**ICUR**	**Method of Calculating Utilities**	**Tariff preference country**	**Utility Source Author (Year)**
The rate subsequent years are reduced in value for future tradeoffs applied to utilities and costs.	(Deterministic or probabilistic) The strateg employed to describe th variability in parameters used in the economic model, how changing th parameters affect mode results.	The primary conclusions the CUA	The cost per QALY relative to the comparato	Whether preferences wer elicited directly or indirectly (with generic/disease-specific preference-based measures) that assign algorithms from populati preferences.	The country population's value set used for indirect measures (e.g. EQ-5D HUI3)	Primary author of the utility values cited for model parameters and bibtex referen key.

SpRUCE Checklist Extraction Fields

Search terms and scope	Study selection criteria	Quality check	Assessment of HSU relevance	Population or patient characteristics	Measure used	Preference weights
Describe the final search strategy and ensure it cover appropriate databases.	Describe the criteria used to identify and select studies f the systematic review (eg, study sample, age range, an disease stage or severity)	Describe the quality criteria used during the review to decide whether to include o exclude studies from the analysis	Describe the relevance of HSUs to the cost-effectiveness model and th target reimbursement agency if appropriate.	Report relevant patient characteristics, such as age, sex, comorbidities, diagnosis and disease severity.	Provide the name o the measure used in the study.	Describe the technique used to value the healt state (eg, TTO or SG) and the country in whic the data were collected.

Descriptive statistics about HSUs	Response rate for the measure used	Extent of missing data or data lost to follow-up	Original reference	Basis for selecting HSUs	Method used to combine estimates	Actual HSUs used	Adjustments or assumptions
Include the mean and variance around all HSU used in the model.	Indicate whether the response rate is likel to jeopardize the validity of the measure.	Report rates of loss to follow-up and of missing data, especially if missin data could threaten the extent to which the HSUs are representative.	Cite the original published study for the HSUs and not a previous economic study that used this evidence.	Provide the rationale for selecting the HSUs used in th model.	If HSUs were combined, describe the analytic method (eg, meta-analysis) used to combine them	Report all HSUs used in the model as well as the measure from which the HSUs were calculated	Describe any adjustments or assumptions used about the HSUs in the cost-effectivenes model. Report both the raw ar final HSU values used with examples, if required, to clari the method used to adjust the data.

Health Utilities and Value Source Extraction Fields

cua_ref	util_ref	state_factor	measure_type	preference_population	measure_scale

state_measure	state	state_group	time_months	sample_size	value	sd	alpha	beta

SUPPLEMENT 3. Validation: The HUG-5 measure attributes and levels presented in paper-and-pencil form.

Does glaucoma cause visual discomfort (e.g., uneasy feeling in the eye(s) or pain, difficulty seeing in the darkness)?

Not at all	☐
Slightly	☐
Moderately	☐
Very much	☐
Severely	☐

Does glaucoma impact your mobility (e.g., feeling cautious when driving or biking, difficulty going up or downstairs, incidents of tripping or bumping when walking)?

Not at all	☐
Slightly	☐
Moderately	☐
Very much	☐
Severely	☐

Does glaucoma affect your daily life activities (e.g. difficulty with any close-up work, household chores or errands)?

Not at all	☐
Slightly	☐
Moderately	☐
Very much	☐
Severely	☐

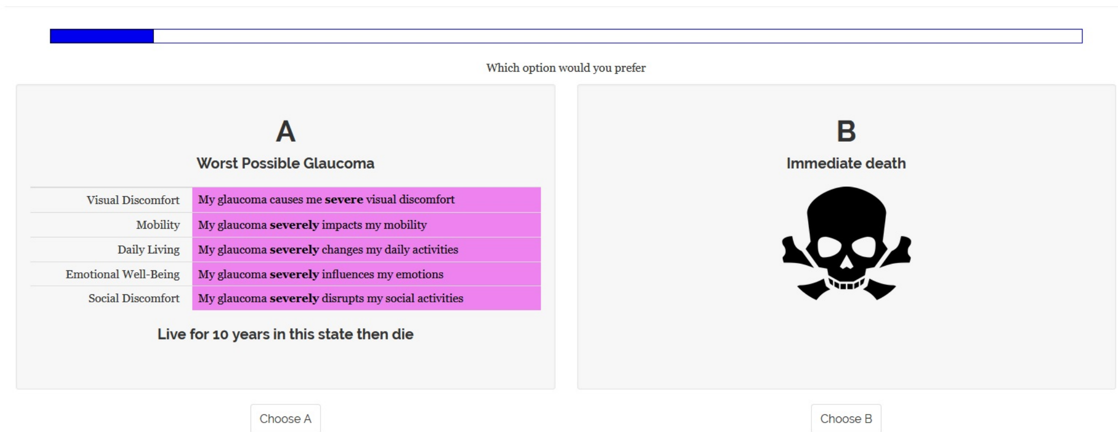
Does glaucoma affect you emotionally (e.g., frustrated with symptoms or treatment, disturbed by frequent thoughts, troubled by worries or fears)?

Not at all	☐
Slightly	☐
Moderately	☐
Very much	☐
Severely	☐

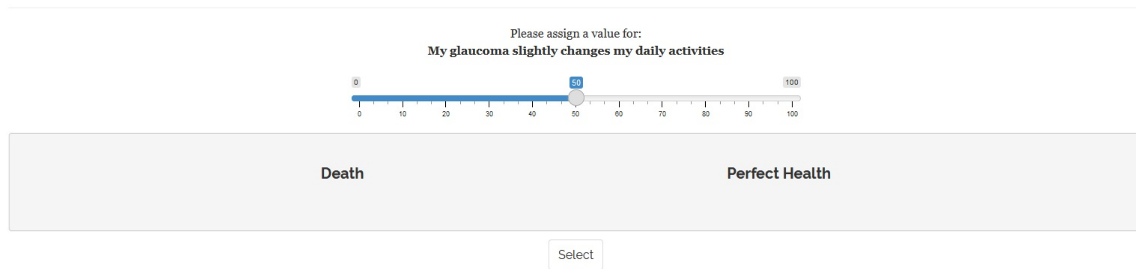
Does glaucoma disrupt your social activities (e.g., missing out on things, difficulty playing sports, discomfort attending social gatherings or crowded places, incidents of social embarrassment)?

Not at all	☐
Slightly	☐
Moderately	☐
Very much	☐
Severely	☐

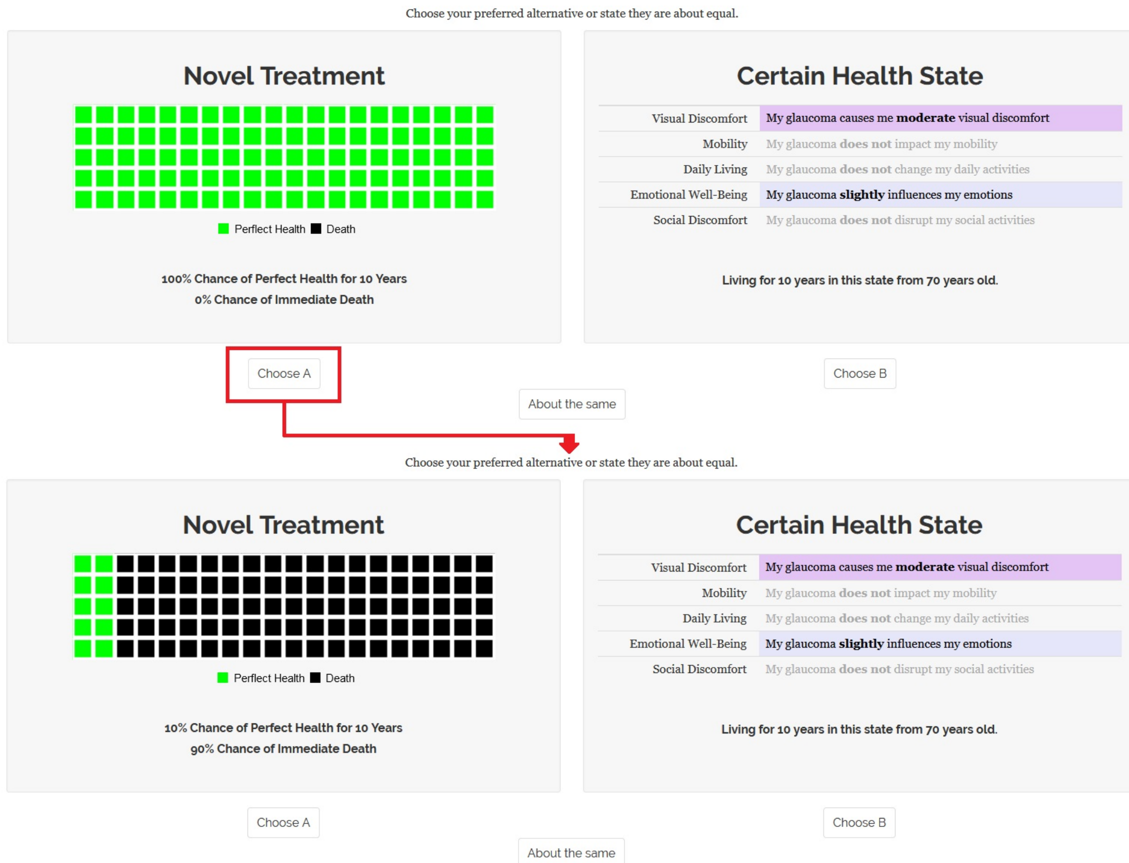
SUPPLEMENT 4. MADUF: The user interface for the online data collection application: anchor task



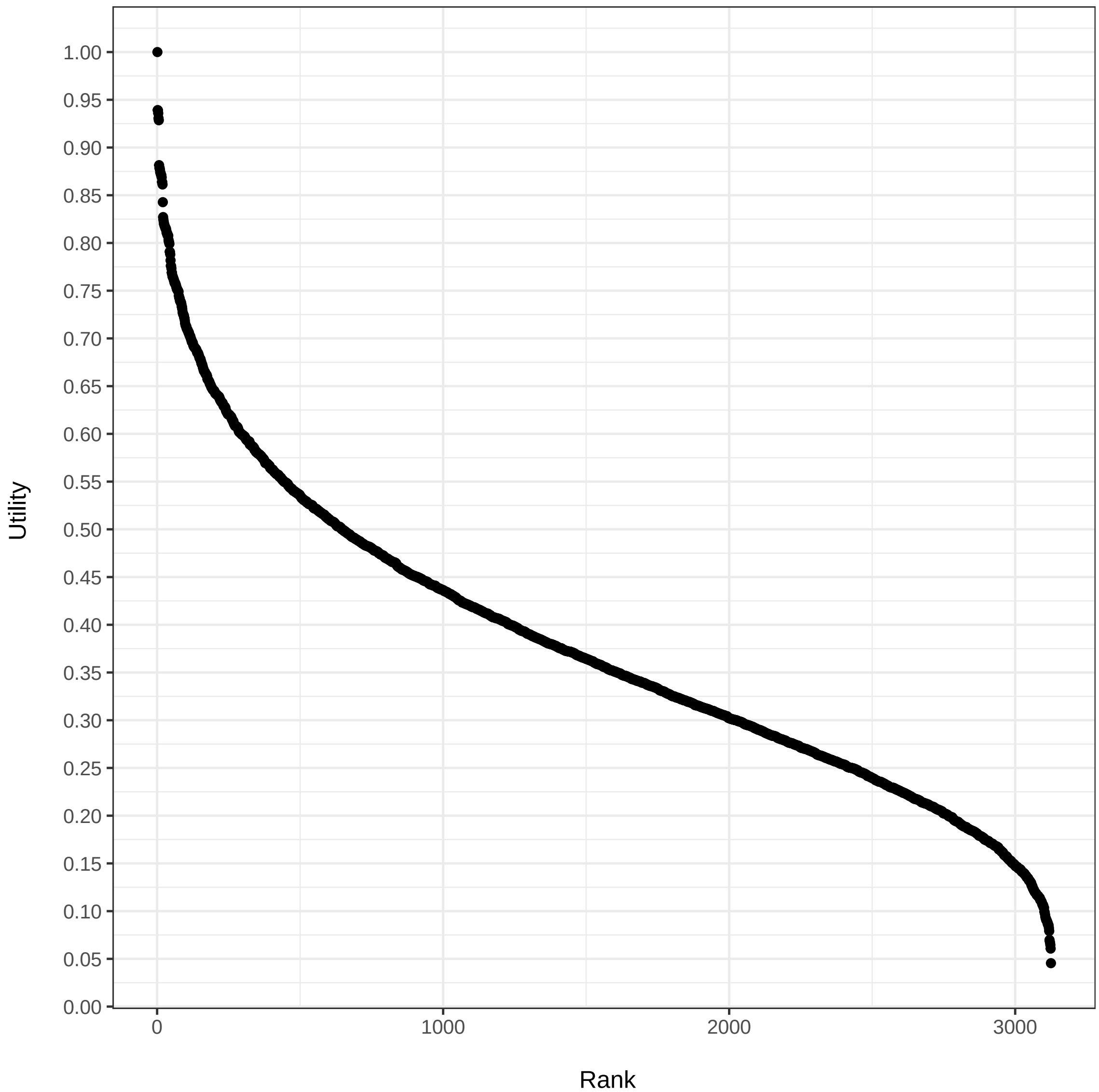
SUPPLEMENT 5. MADUF: The user interface for the online data collection application: visual analog scale task



SUPPLEMENT 6. MADUF: The user interface for the online data collection application: standard gamble task



SUPPLEMENT 7. MADUF: Utility values for all HUG-5 health states, ranked according to utility value.



SUPPLEMENT 8. MADUF: Solving for k, interaction constant:

```
0.95=(1/k)*(((1+k*0.4)*(1+k*0.41)*(1+k*0.39)*(1+k*0.37)*(1+k
    *0.33))-1)
```

SUPPLEMENT 9. MADUF: EOSBA function to adjust vector of state values

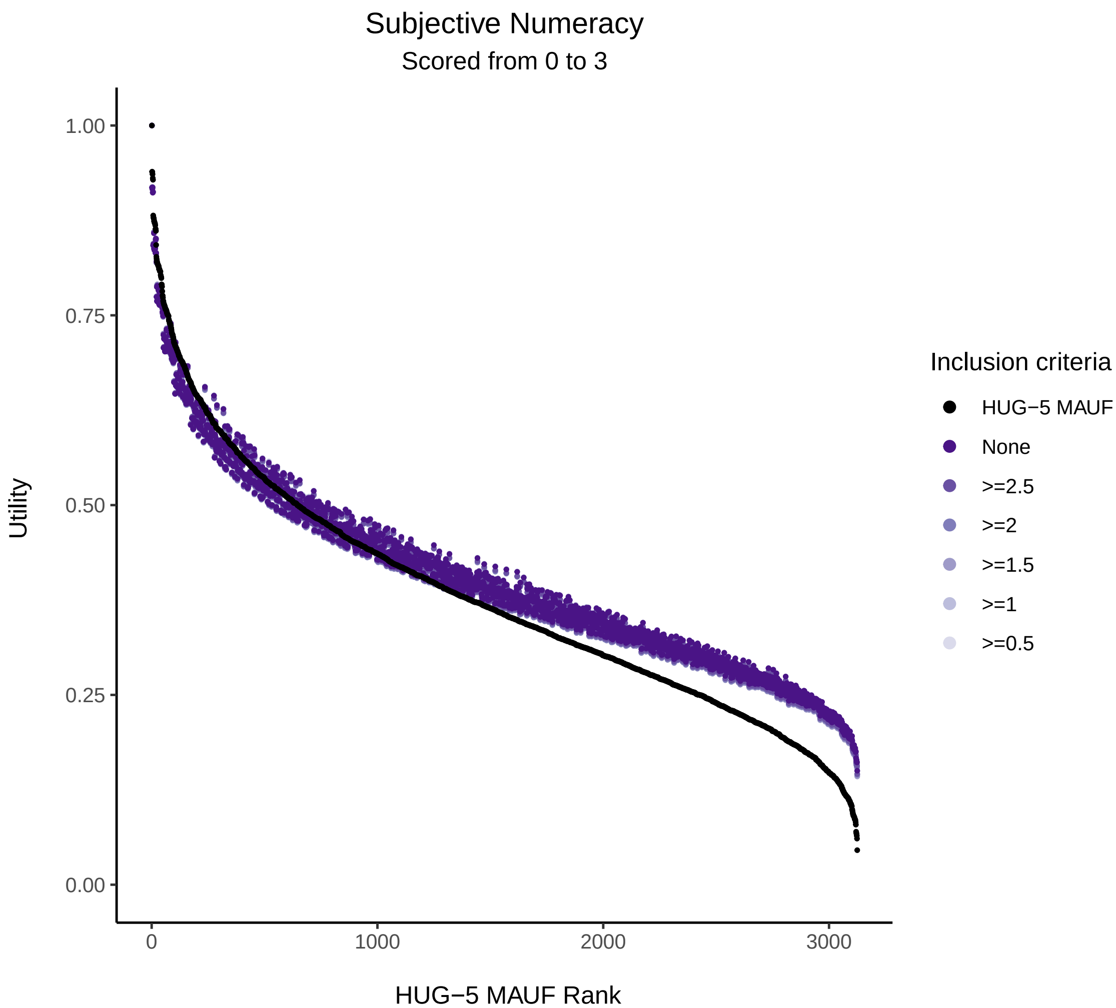
```
vector_of_state_values <- c
   (81.28462,74.01538,62.00769,52.43462)

rescale <- function(vector_of_state_values) {

  eosba_mild_state <- 100 - ((100-vector_of_state_values[1])
     /1.78)

  single_attribute_eosba_range = eosba_mild_state-
     vector_of_state_values[4]
  single_attribute_range = vector_of_state_values[1]-
     vector_of_state_values[4]

  eosba_adjusted_single_attribute_values <- c(
    single_attribute_eosba_range*(vector_of_state_values[2]-
       vector_of_state_values[4])/single_attribute_range+
       vector_of_state_values[4],
    single_attribute_eosba_range*(vector_of_state_values[3]-
       vector_of_state_values[4])/single_attribute_range+
       vector_of_state_values[4],
    vector_of_state_values[4]
    )

  return(eosba_adjusted_single_attribute_values)
}
```

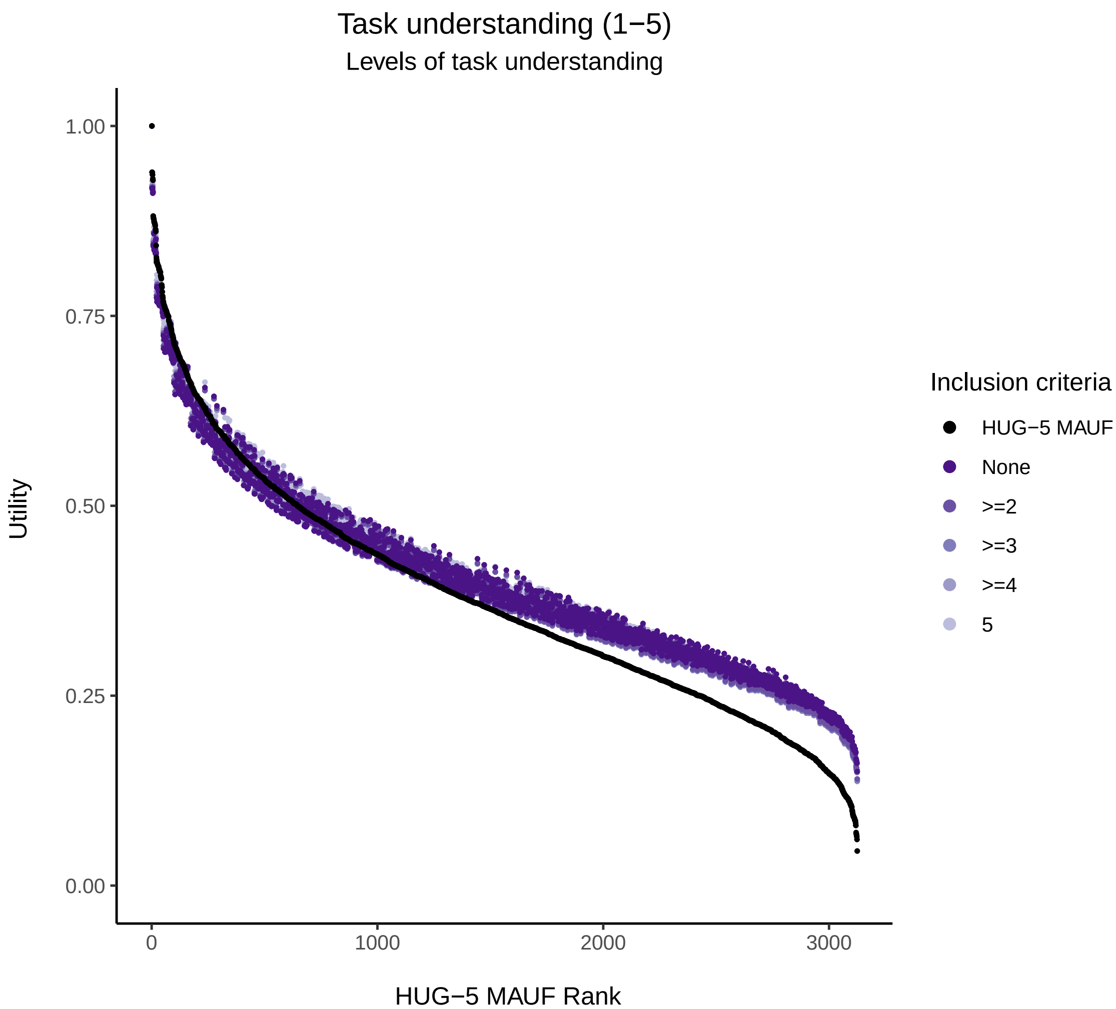
SUPPLEMENT 10. MADUF: Function to calculate all HUG-5 health states

```
maduf <- function(c, cj = vector(), ui = vector()){
  u <- (1/c) * ((
    (1 + c * cj[1] * ui[1]) *
      (1 + c * cj[2] * ui[2]) *
      (1 + c * cj[3] * ui[3]) *
      (1 + c * cj[4] * ui[4]) *
      (1 + c * cj[5] * ui[5])
  ) - 1)
  return(u)
}
```

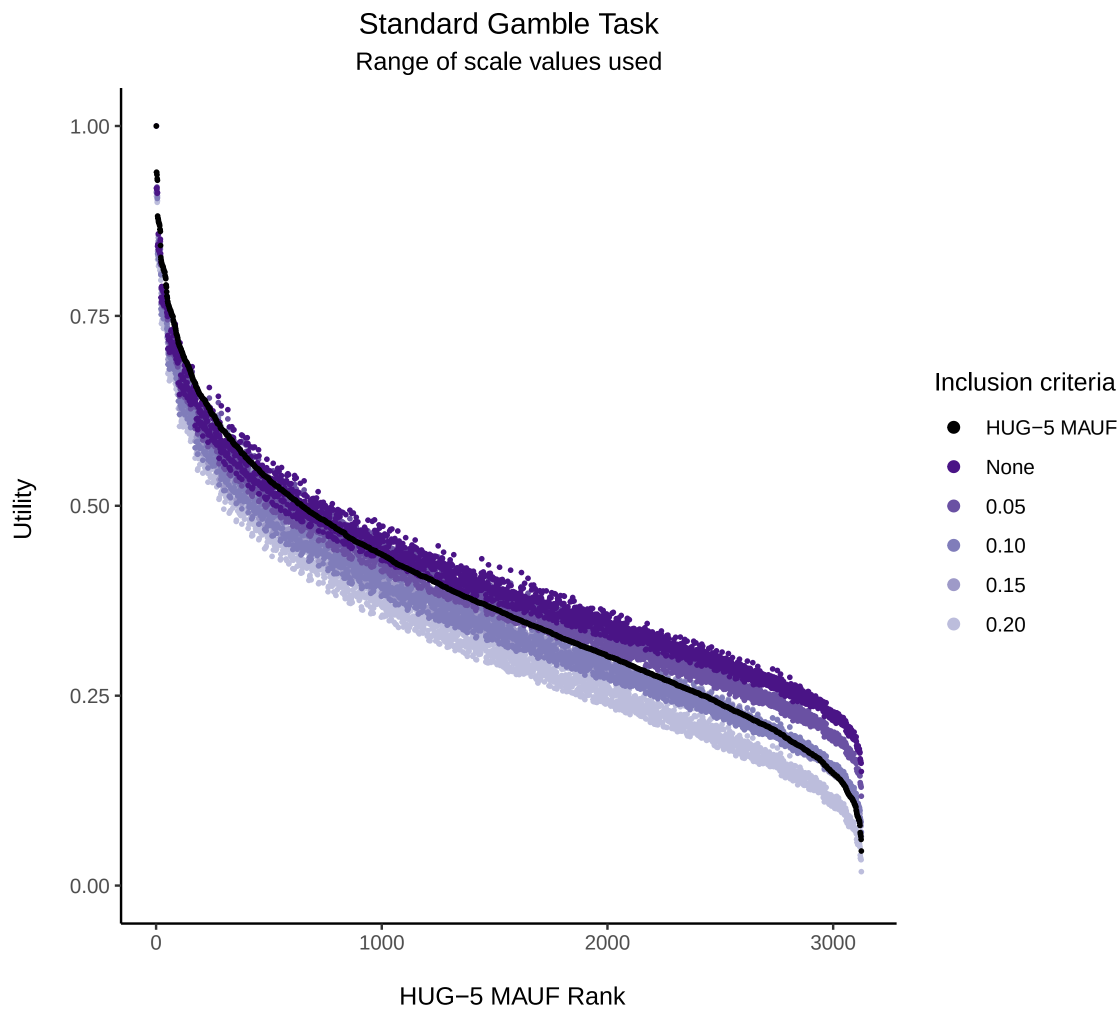
SUPPLEMENT 11. MADUF: SENSITIVITY ANALYSIS OF INCLUSION CRITERIA.

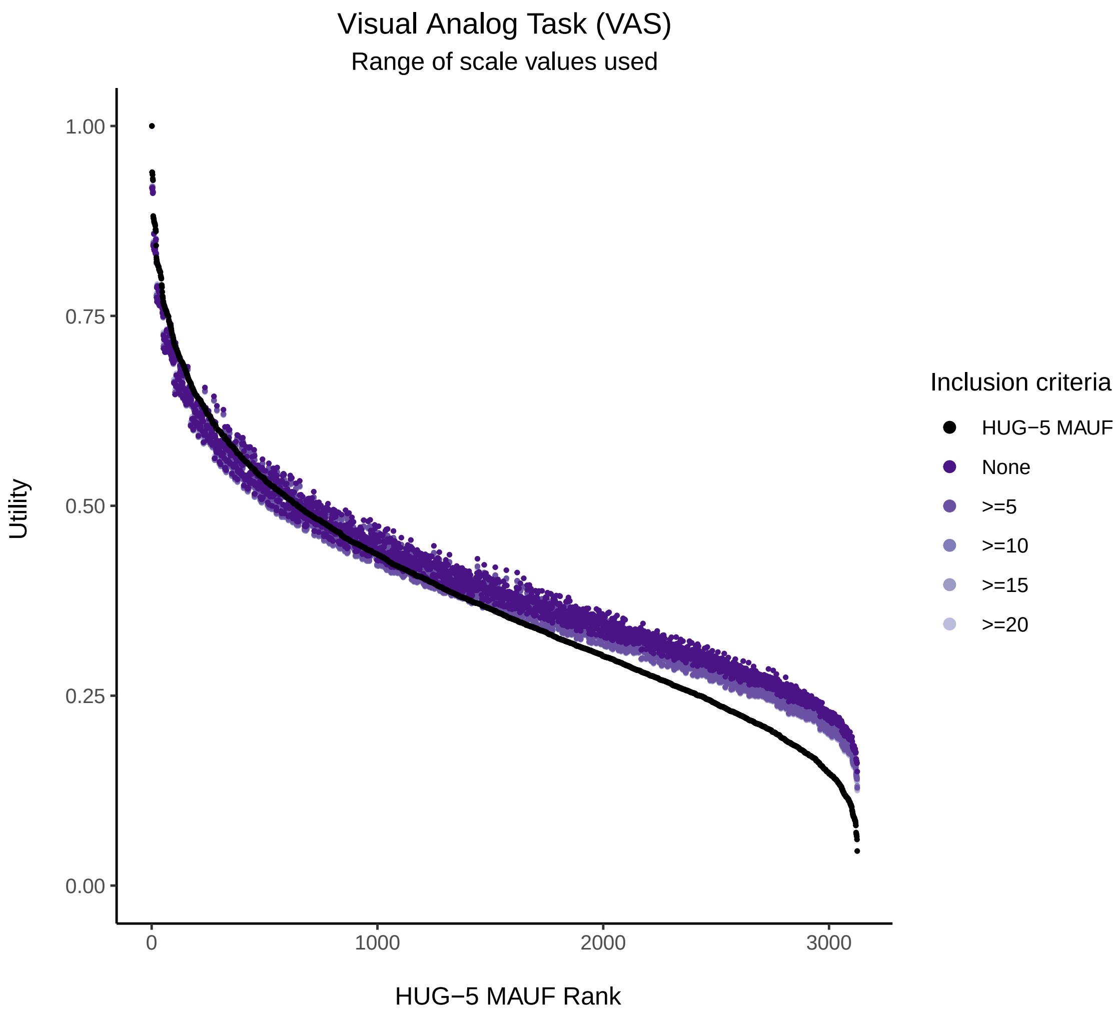
Note: The decision to set inclusion criteria was based off a series of papers that determined optimal thresholds for criteria, including subjective numeracy and task understanding, range of values in response to experiment tasks and violations of monotonicity. This supplementary information was added to illustrate understand the impact of each inclusion criteria on estimating the MADUF of the HUG-5.



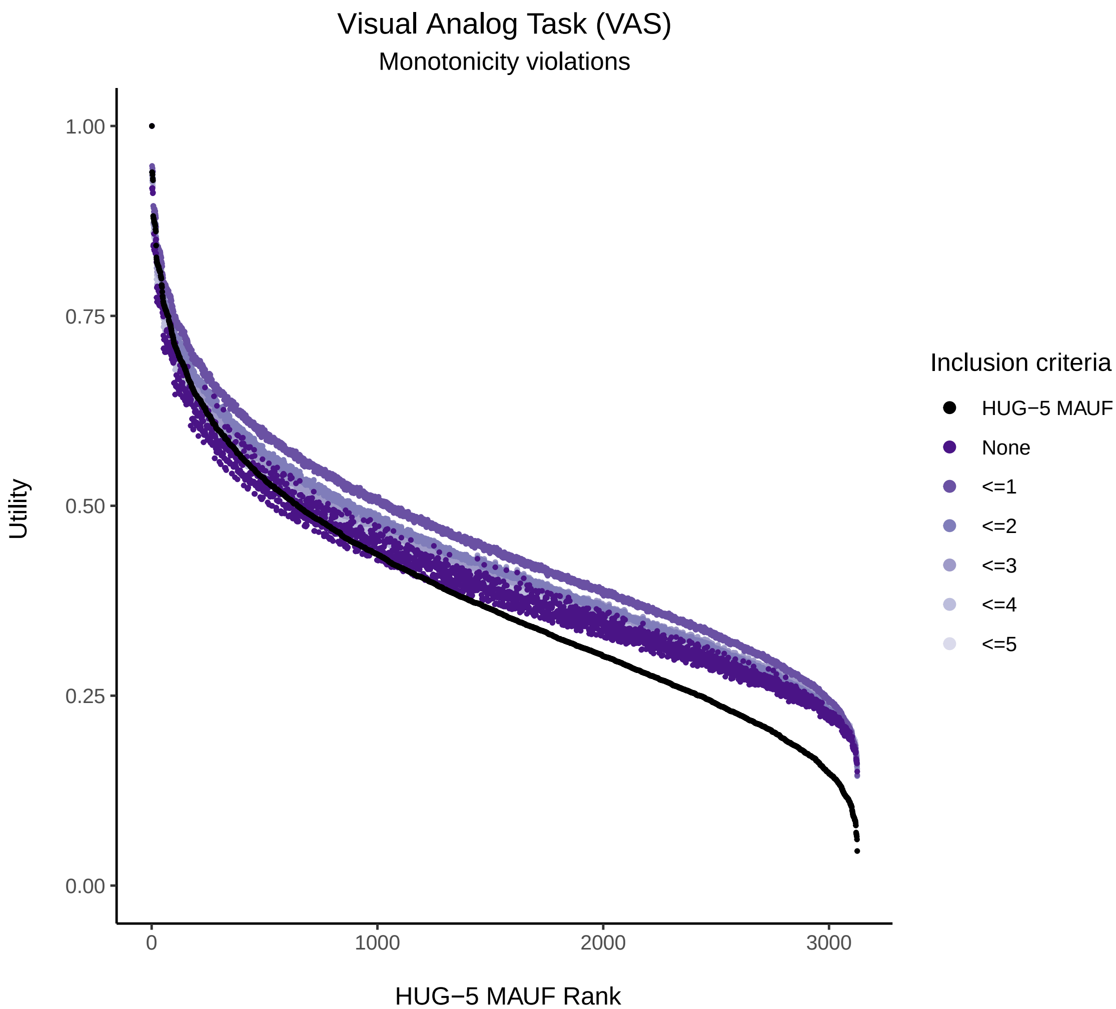


Task understanding (1−5)
Levels of task understanding
Utility
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
HUG−5 MAUF Rank
Inclusion criteria
HUG−5 MAUF
None
>=2
>=3
>=4
5

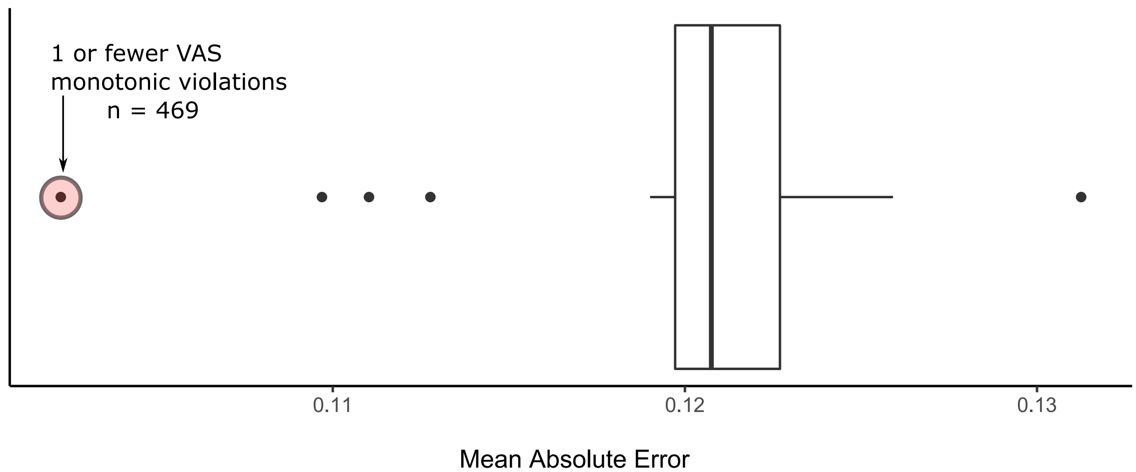

Standard Gamble Task
Range of scale values used
Utility
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
HUG−5 MAUF Rank
Inclusion criteria
HUG−5 MAUF
None
0.05
0.10
0.15
0.20



Visual Analog Task (VAS)
Range of scale values used
Utility
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
HUG−5 MAUF Rank
Inclusion criteria
HUG−5 MAUF
None
>=5
>=10
>=15
>=20


Visual Analog Task (VAS)
Monotonicity violations
Utility
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
HUG−5 MAUF Rank
Inclusion criteria
HUG−5 MAUF
None
<=1
<=2
<=3
<=4
<=5

This box plot illustrates the distribution of MAE estimates across all sensitivity analyses, where each inclusion criteria was varied for a representative range of possible values. The MAE is the average absolute error, in terms of difference between the estimated MADUF values and the directly elicited SG values for respondents included in each sensitivity sample.



	All citeria (n_events =218)	>2 VAS monotonic violations (n_events = 104)	Numeracy <2.5 (n_events = 30)	Task understanding <2 (n_events = 16)	SG Range <0.10 (n_events = 102)	VAS Range <10 (n_events = 21)
(Intercept)	0.23 *** [0.13, 0.42]	0.15 *** [0.07, 0.33]	0.03 *** [0.01, 0.11]	0.00 [0.00, Inf]	0.05 *** [0.02, 0.11]	0.00 *** [0.00, 0.02]
Age group						
18-29 (ref)	-	-	-	-	-	-
30-39	0.74 [0.42, 1.28]	0.69 [0.33, 1.45]	0.38 [0.11, 1.27]	0.42 [0.03, 7.06]	0.68 [0.31, 1.50]	1.06 [0.25, 4.44]
40-49	1.23 [0.71, 2.15]	1.48 [0.73, 2.99]	0.48 [0.14, 1.64]	0.63 [0.05, 7.62]	0.82 [0.38, 1.73]	0.46 [0.08, 2.63]
50-59	1.09 [0.61, 1.94]	0.72 [0.31, 1.68]	0.26 [0.05, 1.21]	0.44 [0.03, 5.81]	1.40 [0.69, 2.85]	1.28 [0.27, 6.01]
60 +	1.03 [0.59, 1.77]	0.99 [0.45, 2.19]	0.18 * [0.04, 0.84]	0.00 [0.00, Inf]	1.62 [0.85, 3.11]	0.12 [0.01, 1.14]
Education						
Bachelor degree (ref)	-	-	-	-	-	-
High school or lower	1.20 [0.74, 1.96]	0.58 [0.26, 1.28]	3.01 * [1.07, 8.44]	16.22 ** [2.31, 114.08]	1.78 [0.99, 3.21]	7.30 ** [1.82, 29.20]
College diploma	0.66 [0.37, 1.20]	0.50 [0.19, 1.30]	1.09 [0.25, 4.66]	0.00 [0.00, Inf]	0.96 [0.47, 1.95]	0.00 [0.00, Inf]
Graduate Program	1.40 [0.82, 2.41]	1.86 [0.94, 3.67]	1.44 [0.39, 5.28]	1.58 [0.11, 22.00]	1.02 [0.50, 2.11]	1.97 [0.37, 10.32]
Post-Graduate Program	1.90 * [1.07, 3.36]	4.40 *** [2.24, 8.65]	0.42 [0.05, 3.76]	0.00 [0.00, Inf]	0.43 [0.16, 1.10]	0.72 [0.07, 7.77]
General health						
Very good (ref)	-	-	-	-	-	-
Excellent	2.05 ** [1.29, 3.25]	1.70 [0.97, 3.00]	5.40 * [1.45, 20.13]	1080278496.11 [0.00, Inf]	1.25 [0.68, 2.31]	12.50 * [1.54, 101.63]
Fair	1.14 [0.60, 2.15]	0.67 [0.25, 1.79]	2.40 [0.38, 15.37]	1.10 [0.00, Inf]	0.83 [0.36, 1.89]	13.77 * [1.18, 161.24]
Good	1.20 [0.74, 1.92]	0.56 [0.27, 1.17]	4.90 * [1.27, 18.92]	0.00 [0.00, Inf]	1.23 [0.68, 2.21]	3.21 [0.31, 33.41]
None	2.60 [0.12, 54.27]	4.40 [0.21, 91.21]	0.00 [0.00, Inf]	1.48 [0.00, Inf]	0.00 [0.00, Inf]	0.00 [0.00, Inf]
Poor	0.82 [0.20, 3.41]	1.43 [0.22, 9.21]	0.00 [0.00, Inf]	2501403841.00 [0.00, Inf]	0.00 [0.00, Inf]	0.00 [0.00, Inf]
Has glaucoma						
No (ref)	-	-	-	-	-	-
Yes	0.58 [0.19, 1.72]	0.21 [0.02, 1.86]	0.00 [0.00, Inf]	0.00 [0.00, Inf]	1.86 [0.59, 5.88]	3.83 [0.29, 49.99]
Not sure	2.08 [0.38, 11.43]	4.36 [0.57, 33.32]	2.18 [0.18, 26.90]	24772992.24 [0.00, Inf]	0.00 [0.00, Inf]	9.10 [0.16, 517.01]
Friends or family with vision problems						
None (ref)						
1	0.75 [0.48, 1.17]	0.70 [0.37, 1.33]	0.70 [0.26, 1.91]	0.14 [0.01, 1.59]	0.99 [0.56, 1.73]	0.47 [0.12, 1.84]
2	0.46 ** [0.25, 0.82]	0.40 * [0.17, 0.96]	0.32 [0.07, 1.56]	0.00 [0.00, Inf]	0.82 [0.40, 1.69]	0.00 [0.00, Inf]
3 or more	1.26 [0.67, 2.36]	1.33 [0.59, 2.99]	0.30 [0.04, 2.47]	2.96 [0.20, 43.50]	1.12 [0.49, 2.52]	0.26 [0.02, 3.38]
Anchor choice						
Death worse (ref)	-	-	-	-	-	-
PITS worse	2.16 *** [1.43, 3.26]	1.15 [0.67, 1.98]	0.97 [0.41, 2.31]	2.05 [0.33, 12.89]	4.49 *** [2.28, 8.82]	4.81 * [1.02, 22.71]
N	621	622	622	621	622	622
AIC	775.20	495.54	234.77	96.50	527.32	158.08
BIC	863.83	584.20	323.43	185.12	615.98	246.74
Pseudo R2	0.13	0.22	0.20	0.60	0.15	0.36

Binomial GLM reported in odds ratios with [CI] of meeting each inclusion criteria. All continuous predictors are mean-centered and scaled by 1 standard deviation.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

REFERENCES

1. Petrou S, Wolstenholme J. A review of alternative approaches to healthcare resource allocation. *Pharmacoeconomics*. 2000;18(1):33-43.
2. Banta D. The development of health technology assessment. *Health policy*. 2003;63(2):121-132.
3. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health technology assessment (Winchester, England)*. 2004;8(36):iii-iv.
4. Weinstein MC, Toy EL, Sandberg EA, et al. Modeling for health care and other policy decisions: Uses, roles, and validity. *Value in Health*. 2001;4(5):348-361.
5. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: A review. *Jama*. 2014;311(18):1901-1911.
6. Wesselink C, Heeg GP, Jansonius NM. Glaucoma monitoring in a clinical setting: Glaucoma progression analysis vs nonparametric progression analysis in the groningen longitudinal glaucoma study. *Archives of Ophthalmology*. 2009;127(3):270-274.
7. Hodapp E, Parrish RK, Anderson DR. *Clinical Decisions in Glaucoma*. Mosby Incorporated; 1993.
8. Tan NY, Friedman DS, Stalmans I, Ahmed IIK, Sng CC. Glaucoma screening: Where are we and where do we need to go? *Current opinion in ophthalmology*. 2020;31(2):91-100.
9. Poitras V, Wells C, Hutnik C, et al. Optimal use of minimally invasive glaucoma surgery: A health technology assessment [internet]. Published online 2019.
10. Schmier JK, Halpern MT, Jones ML. The economic implications of glaucoma. *Pharmacoeconomics*. 2007;25(4):287-308.
11. Rowen D, Mulhern B, Banerjee S, et al. Comparison of general population, patient, and carer utility values for dementia health states. *Medical Decision Making*. 2015;35(1):68-80.
12. Matza LS, Deger KA, Vo P, Maniyar F, Goadsby PJ. Health state utilities associated with attributes of migraine preventive treatments based on patient and general population preferences. *Quality of Life Research*. 2019;28(9):2359-2372.
13. Damiano AM, Steinberg EP, Cassard SD, et al. Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. *Medical care*. Published online 1995:AS120-AS130.
14. Misajon R, Hawthorne G, Richardson J, et al. Vision and quality of life: The development of a utility measure. *Investigative ophthalmology & visual science*. 2005;46(11):4007-4015.
15. Lee BS, Kymes SM, Nease Jr RF, Sumner W, Siegfried CJ, Gordon MO. The impact of anchor point on utilities for 5 common ophthalmic diseases. *Ophthalmology*. 2008;115(5):898-903.

16. Muratov S, Podbielski DW, Kennedy K, et al. Preference-based glaucoma-specific health-related quality of life instrument: Development of the health utility for glaucoma. *Journal of glaucoma*. 2018;27(7):585-591.

17. Isenberg J, Sheybani A, Ahmed IIK. Advancing the economic assessment of microinvasive glaucoma surgery. *Journal of Glaucoma*. 2016;25(7):e721-e722.

18. Sun X, Dai Y. *Medical Treatment of Glaucoma*. Springer; 2019.

19. Zhang N, Wang J, Li Y, Jiang B. Prevalence of primary open angle glaucoma in the last 20 years: A meta-analysis and systematic review. *Scientific Reports*. 2021;11(1):1-12.

20. Lazcano-Gomez G, Angeles Ramos-Cadena M de los, Torres-Tamayo M, Oteyza AH de, Turati-Acosta M, Jimenez-Román J. Cost of glaucoma treatment in a developing country over a 5-year period. *Medicine*. 2016;95(47).

21. Hernández RA, Burr JM, Vale LD. Economic evaluation of screening for open-angle glaucoma. *International journal of technology assessment in health care*. 2008;24(2):203-211.

22. Sharafeldin N, Kawaguchi A, Sundaram A, et al. Review of economic evaluations of teleophthalmology as a screening strategy for chronic eye disease in adults. *British Journal of Ophthalmology*. 2018;102(11):1485-1491.

23. Bartelt-Hofer J, Ben-Debba L, Flessa S. Systematic review of economic evaluations in primary open-angle glaucoma: Decision analytic modeling insights. *PharmacoEconomics-open*. 2020;4(1):5-12.

24. Koleva D, Motterlini N, Schiavone M, Garattini L. Medical costs of glaucoma and ocular hypertension in italian referral centres: A prospective study. *Ophthalmologica*. 2007;221(5):340-347.

25. Orme M, Boler A. Prostaglandin analogues for the treatment of glaucoma and ocular hypertension. *Pharmacoeconomics*. 2006;24(8):743-750.

26. Ribeiro IS, Batel Marques FJ, Alves DG, Alves CMC. A systematic review of the methodological quality of economic studies evaluating ophthalmic drugs. *Expert review of pharmacoeconomics & outcomes research*. 2019;19(4):421-430.

27. Agrawal P, Bradshaw SE. Systematic literature review of clinical and economic outcomes of micro-invasive glaucoma surgery (MIGS) in primary open-angle glaucoma. *Ophthalmology and therapy*. 2018;7(1):49-73.

28. Brazier J, Ara R, Azzabi I, et al. Identification, review, and use of health state utilities in cost-effectiveness models: An ISPOR good practices for outcomes research task force report. *Value in Health*. 2019;22(3):267-275.

29. Ribeiro I, Batel Marques F, Mendes D, Alves C. A systematic review of economic studies evaluating ophthalmic drugs: An analysis of the health-state utilities. *Ophthalmic Epidemiology*. 2020;27(5):325-338.

30. Brown MM, Brown GC, Sharma S, Kistler J, Brown H. Utility values associated with blindness in an adult population. *British Journal of Ophthalmology*. 2001;85(3):327-331.

31. Burr JM, Kilonzo M, Vale L, Ryan M. Developing a preference-based glaucoma utility index using a discrete choice experiment. *Optometry and vision science*. 2007;84(8):E797-E809.

32. Choi S, Choi JA, Kwon JW, Park SM, Jee D. Utility values for glaucoma patients in korea. *PloS one*. 2018;13(5).

33. Jampel HD, Schwartz A, Pollack I, Abrams D, Weiss H, Miller R. Glaucoma patients' assessment of their visual function and quality of life. *Journal of glaucoma*. 2002;11(2):154-163.

34. Kobelt G, Jonsson B, Bergström A, Chen E, Lindén C, Alm A. Cost-effectiveness analysis in glaucoma: What drives utility? Results from a pilot study in sweden. *Acta Ophthalmologica Scandinavica*. 2006;84(3):363-371.

35. Rein DB, Wirth KE, Johnson CA, Lee PP. Estimating quality-adjusted life year losses associated with visual field deficits using methodological approaches. *Ophthalmic epidemiology*. 2007;14(4):258-264.

36. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Medical care*. Published online 2000:583-637.

37. Gestel A van, Webers CA, Severens JL, et al. The long-term outcomes of four alternative treatment strategies for primary open-angle glaucoma. *Acta ophthalmologica*. 2012;90(1):20-31.

38. Wittenborn JS, Rein DB. The cost-effectiveness of glaucoma interventions in barbados and ghana. *Optometry and vision science: official publication of the American Academy of Optometry*. 2011;88(1):155.

39. Lazzaro C, Steen C van, Aptel F, Schweitzer C, Angelillo L. Cost-utility analysis of STN1013001, a latanoprost cationic emulsion, versus other latanoprost formulations (latanoprost) in open-angle glaucoma or ocular hypertension and ocular surface disease in france. *Journal of ophthalmology*. 2022;2022.

40. Lazzaro C, Steen C van, Billeit S, et al. Cost-utility analysis of a latanoprost cationic emulsion (STN1013001) versus other latanoprost in the treatment of open-angle glaucoma or ocular hypertension and concomitant ocular surface disease in germany. *Clinical Ophthalmology (Auckland, NZ)*. 2022;16:323.

41. Slevin ML, Plant H, Lynch D al, Drinkwater J, Gregory W. Who should measure quality of life, the doctor or the patient? *British journal of cancer*. 1988;57(1):109-112.

42. Mulley Jr AG. Assessing patients' utilities: Can the ends justify the means? *Medical Care*. Published online 1989:S269-S281.
43. Brown MM, Brown GC, Sharma S, Busbee B. Quality of life associated with visual loss: A time tradeoff utility analysis comparison with medical health states. *Ophthalmology*. 2003;110(6):1076-1081.
44. Sharma S, Brown GC, Brown MM, et al. Converting visual acuity to utilities. *Canadian journal of ophthalmology*. 2000;35(5):267-272.
45. Lundberg L, Johannesson M, Isacson DG, Borgquist L. Health-state utilities in a general population in relation to age, gender and socioeconomic factors. *The European Journal of Public Health*. 1999;9(3):211-217.
46. Ara R, Brazier J, Peasgood T, Paisley S. The identification, review and synthesis of health state utility values from the literature. *Pharmacoeconomics*. 2017;35(1):43-55.
47. Ara R, Peasgood T, Mukuria C, et al. Sourcing and using appropriate health state utility values in economic models in health care. *Pharmacoeconomics*. 2017;35(1):7-9.
48. Ara R, Brazier J, Zouraq IA. The use of health state utility values in decision models. *Pharmacoeconomics*. 2017;35(1):77-88.
49. Boodhna T, Crabb DP. More frequent, more costly? Health economic modelling aspects of monitoring glaucoma patients in england. *BMC health services research*. 2016;16(1):611.
50. Zoratti MJ, Pickard AS, Stalmeier PF, et al. Evaluating the conduct and application of health utility studies: A review of critical appraisal tools and reporting checklists. *The European Journal of Health Economics*. 2021;22(5):723-733.
51. Gutierrez P, Wilson MR, Johnson C, et al. Influence of glaucomatous visual field loss on health-related quality of life. *Archives of ophthalmology*. 1997;115(6):777-784.
52. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*. 2006;90(3):262-267.
53. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090.
54. Rein DB, Zhang P, Wirth KE, et al. The economic burden of major adult visual disorders in the united states. *Archives of ophthalmology*. 2006;124(12):1754-1760.
55. Lichter P. Glaucoma therapy and quality of life. *Ophthalmol Clin North Am*. 1991;4:897-904.
56. Lichter P, Becker M, Janz N. The case for quality of life assessment in glaucoma clinical trials. In: *Glaucoma Update IV*. Springer; 1991:187-192.

57. Mills RP, Janz NK, Wren PA, Guire KE, Group CS, et al. Correlation of visual field with quality-of-life measures at diagnosis in the collaborative initial glaucoma treatment study (CIGTS). *Journal of Glaucoma*. 2001;10(3):192-198.

58. Janz NK, Wren PA, Lichter PR, et al. Quality of life in newly diagnosed glaucoma patients: The collaborative initial glaucoma treatment study. *Ophthalmology*. 2001;108(5):887-897.

59. Stenman U, Hakama M, Knekt P, et al. Measurement and modeling of health-related quality of life. *Epidem Demog Public Health*. 2010;195(1):130-135.

60. Nyman SR, Dibb B, Victor CR, Gosney MA. Emotional well-being and adjustment to vision loss in later life: A meta-synthesis of qualitative studies. *Disability and rehabilitation*. 2012;34(12):971-981.

61. Parrish RK, Gedde SJ, Scott IU, et al. Visual function and quality of life among patients with glaucoma. *Archives of Ophthalmology*. 1997;115(11):1447-1455.

62. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *International journal of technology assessment in health care*. 1989;5(4):559-575.

63. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford university press; 2015.

64. Torrance GW. Measurement of health state utilities for economic appraisal: A review. *Journal of health economics*. 1986;5(1):1-30.

65. Varma R, Lee PP, Goldberg I, Kotak S. An assessment of the health and economic burdens of glaucoma. *American journal of ophthalmology*. 2011;152(4):515-522.

66. 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(10):1727-1736.

67. 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):118-127.

68. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and quality of life outcomes*. 2003;1(1):1-8.

69. Espallargues M, Czoski-Murray CJ, Bansback NJ, et al. The impact of age-related macular degeneration on health status utility values. *Investigative ophthalmology & visual science*. 2005;46(11):4016-4023.

70. Mangione CM, Lee PP, Pitts J, et al. Psychometric properties of the national eye institute visual function questionnaire (NEI-VFQ). *Archives of Ophthalmology*. 1998;116(11):1496-1504.

71. Nelson P, Aspinall P, Papasouliotis O, Worton B, O'Brien C. Quality of life in glaucoma and its relationship with visual function. *Journal of glaucoma*. 2003;12(2):139-150.
72. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. *Journal of health services research & policy*. 1999;4(3):174-184.
73. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: An online database. *Health and quality of life outcomes*. 2013;11(1):1-6.
74. Browne C, Brazier J, Carlton J, Alavi Y, Jofre-Bonet M. Estimating quality-adjusted life years from patient-reported visual functioning. *Eye*. 2012;26(10):1295-1301.
75. Kay S, Ferreira A. Mapping the 25-item national eye institute visual functioning questionnaire (NEI VFQ-25) to EQ-5D utility scores. *Ophthalmic epidemiology*. 2014;21(2):66-78.
76. Alavi Y, Jofre-Bonet M, Bunce C, et al. Developing an algorithm to convert routine measures of vision into utility values for glaucoma. *Ophthalmic epidemiology*. 2011;18(5):233-243.
77. Severn P, Fraser S, Finch T, May C. Which quality of life score is best for glaucoma patients and why? *BMC ophthalmology*. 2008;8(1):2.
78. Bozzani FM, Alavi Y, Jofre-Bonet M, Kuper H. A comparison of the sensitivity of EQ-5D, SF-6D and TTO utility values to changes in vision and perceived visual function in patients with primary open-angle glaucoma. *BMC ophthalmology*. 2012;12(1):1-9.
79. Gupta V, Srinivasan G, Mei S, Gazzard G, Sihota R, Kapoor K. Utility values among glaucoma patients: An impact on the quality of life. *British journal of ophthalmology*. 2005;89(10):1241-1244.
80. Muratov S, Podbielski DW, Jack SM, et al. Preference-based disease-specific health-related quality of life instrument for glaucoma: A mixed methods study protocol. *BMJ open*. 2016;6(11):e012732.
81. Committee COSGCPGE et al. Canadian ophthalmological society evidence-based clinical practice guidelines for the management of glaucoma in the adult eye. *Canadian Journal of Ophthalmology*. 2009;44:S7-S54.
82. Mangione CM, Lee PP, Gutierrez PR, et al. Development of the 25-list-item national eye institute visual function questionnaire. *Archives of ophthalmology*. 2001;119(7):1050-1058.
83. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*. 1994;6(4):284.
84. Wilcoxon F. Individual comparisons by ranking methods. In: *Breakthroughs in Statistics*. Springer; 1992:196-202.

85. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*. 2015;18(1):25-34.

86. Furlong W, Feeny D, Torrance G, et al. *Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (Hui3) System: A Technical Report*. Centre for Health Economics; Policy Analysis (CHEPA), McMaster University ...; 1998.

87. Martin AJ, Glasziou P, Simes R, Lumley T. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *International Journal of Technology Assessment in Health Care*. 2000;16(1):137-147.

88. Lugnér AK, Krabbe PF. An overview of the time trade-off method: Concepzt, foundation, and the evaluation of distorting factors in putting a value on health. *Expert review of pharmacoeconomics & outcomes research*. 2020;20(4):331-342.

89. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *Journal of health economics*. 2012;31(1):306-318.

90. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press; 2015.

91. Viney R, Norman R, Brazier J, et al. An australian discrete choice experiment to value EQ-5D health states. *Health Economics*. 2014;23(6):729-742.

92. Torrance GW. Preferences for health outcomes and cost-utility analysis. *Am J Manag Care*. 1997;3(Suppl 1):S8-S20.

93. Keeney RL, Raiffa H, Meyer RF. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge university press; 1993.

94. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The health utilities index (HUI®) system for assessing health-related quality of life in clinical studies. *Annals of medicine*. 2001;33(5):375-384.

95. Fischer GW. Utility models for multiple objective decisions: Do they accurately represent human preferences? *Decision sciences*. 1979;10(3):451-479.

96. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care*. 2002;40(2):113-128.

97. Bremner K, Mitsakakis N, Wilson L, Krahn M. Predicting utility scores for prostate cancer: Mapping the prostate cancer index to the patient-oriented prostate utility scale (PORPUS). *Prostate cancer and prostatic diseases*. 2014;17(1):47-56.

98. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. *Cost-Effectiveness in Health and Medicine*. Oxford University Press; 2016.

99. Brazier J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;26(9):769-779.

100. Weinstein MC, Russell LB, Gold MR, Siegel JE, et al. *Cost-Effectiveness in Health and Medicine*. Oxford university press; 1996.

101. Pickard AS, Law EH, Jiang R, et al. United states valuation of EQ-5D-5L health states using an international protocol. *Value in Health*. 2019;22(8):931-941.

102. Group EDPR et al. Prevalence of open-angle glaucoma among adults in the united states. *Archives of ophthalmology*. 2004;122(4):532.

103. Heijl A, Bengtsson B, Hyman L, Leske MC, Group EMGT, et al. Natural history of open-angle glaucoma. *Ophthalmology*. 2009;116(12):2271-2276.

104. Van Gestel A, Webers C, Beckers H, et al. The relationship between visual field loss in glaucoma and health-related quality-of-life. *Eye*. 2010;24(12):1759.

105. Goodwin E, Green C. A systematic review of the literature on the development of condition-specific preference-based measures of health. *Applied health economics and health policy*. 2016;14(2):161-183.

106. Summerfield AQ, Barton G. Sensitivity of EQ-5D-3L, HUI2, HUI3, and SF-6D to changes in speech reception and tinnitus associated with cochlear implantation. *Quality of Life Research*. 2019;28(5):1145-1154.

107. Heintz E, Wiréhn AB, Peebo BB, Rosenqvist U, Levin LÅ. QALY weights for diabetic retinopathy—a comparison of health state valuations with HUI-3, EQ-5D, EQ-VAS, and TTO. *Value in health*. 2012;15(3):475-484.

108. Payakachat N, Ali MM, Tilford JM. Can the EQ-5D detect meaningful change? A systematic review. *Pharmacoeconomics*. 2015;33(11):1137-1154.

109. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics*. 2017;35(1):33-41.

110. Nordmann JP, Viala M, Sullivan K, Arnould B, Berdeaux G. Psychometric validation of the national eye institute visual function questionnaire—25 (NEI VFQ-25) french version. *Pharmacoeconomics*. 2004;22(3):197-206.

111. Kennedy K, Podbielski D, Nanji K, Muratov S, Ahmed I, Xie F. Disease-specific preference-based measure of glaucoma health states: HUG-5 psychometric validation. *Journal of glaucoma*. 2019;28(7):593-600.

112. Feeny D, Furlong W, Saigal S, Sun J. Comparing directly measured standard gamble scores to HUI2 and HUI3 utility scores: Group-and individual-level comparisons. *Social science & medicine*. 2004;58(4):799-809.

113. Torrance GW. Utility approach to measuring health-related quality of life. *Journal of chronic diseases*. 1987;40(6):593-600.

114. Luce RD, Raiffa H. *Games and Decisions: Introduction and Critical Survey*. Courier Corporation; 1989.

115. Morgenstern O, Von Neumann J. *Theory of Games and Economic Behavior*. Princeton university press; 1953.

116. Gresh D, Deleris LA, Gasparini L, Evans D. Visualizing risk. In: *Proceedings of IEEE Information Visualization Conference*.; 2011:1-10.

117. Craig BM, Reeve BB, Brown PM, et al. US valuation of health outcomes measured using the PROMIS-29. *Value in Health*. 2014;17(8):846-853.

118. Dewitt B, Fischhoff B, Davis AL, Broomell SB, Roberts MS, Hanmer J. Exclusion criteria as measurements i: Identifying invalid responses. *Medical Decision Making*. 2019;39(6):693-703.

119. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*. 2007;27(5):672-680.

120. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. https://www.R-project.org/

121. Beeley C. *Web Application Development with r Using Shiny*. Packt Publishing Ltd; 2016.

122. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. *The economic journal*. 2008;118(525):215-234.

123. Torrance GW, Feeny D, Furlong W. Visual analog scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making*. 2001;21(4):329-334.

124. Dolan P, Sutton M. Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. *Social science & medicine*. 1997;44(10):1519-1530.

125. Tomlinson G, Bremner KE, Ritvo P, Naglie G, Krahn MD. Development and validation of a utility weighting function for the patient-oriented prostate utility scale (PORPUS). *Medical Decision Making*. 2012;32(1):11-30.

126. Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: Addressing design and sampling issues. *Social science & medicine*. 2014;114:38-48.

127. Augestad LA, Stavem K, Kristiansen IS, Samuelsen CH, Rand-Hendriksen K. Influenced from the start: Anchoring bias in time trade-off valuations. *Quality of Life Research*. 2016;25(9):2179-2191.

128. Osch SM van, Stiggelbout AM. The construction of standard gamble utilities. *Health economics*. 2008;17(1):31-40.

129. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till J, Boyd N. The measurement of patients' values in medicine. *Medical Decision Making*. 1982;2(4):449-462.

130. Wakker P, Stiggelbout A. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Medical Decision Making*. 1995;15(2):180-186.

131. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annual review of public health*. 2000;21(1):587-611.

132. Mann R, Brazier J, Tsuchiya A. A comparison of patient and general population weightings of EQ-5D dimensions. *Health economics*. 2009;18(3):363-372.

133. Neumann PJ, Willke RJ, Garrison Jr LP. A health economics approach to US value assessment frameworks—introduction: An ISPOR special task force report [1]. *Value in Health*. 2018;21(2):119-123.

134. Henderson GE, Churchill LR, Davis AM, et al. Clinical trials and medical care: Defining the therapeutic misconception. *PLoS medicine*. 2007;4(11):e324.

135. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system: Health utilities index mark 2. *Medical care*. Published online 1996:702-722.

136. Ferreira LN, Ferreira PL, Pereira LN, Oppe M. The valuation of the EQ-5D in portugal. *Quality of Life Research*. 2014;23(2):413-423.

137. Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: The king's health questionnaire. *Medical Decision Making*. 2008;28(1):113-126.

138. Dewitt B, Feeny D, Fischhoff B, et al. Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: The PROMIS®-preference (PROPr) scoring system. *Medical Decision Making*. 2018;38(6):683-698.

139. Kennedy K, Nanji K, Patil N, et al. Responsiveness of the HUG-5 in an outpatient clinic: A 12-month randomised feasibility study protocol. *BMJ open*. 2022;12(4):e061057.

140. Nanji K, Sarohia GS, Podbielski D, Kennedy K. Health state measurement of glaucoma patients. *Investigative Ophthalmology & Visual Science*. 2019;60(9):4962-4962.

141. Healey PR, Tilden D, Jackson D, Aghajanian L. A cost-utility analysis of trabecular bypass devices versus usual care for patients with open-angle glaucoma. *PharmacoEconomics-Open*. Published online 2021:1-11.

142. Burr J, Hernández R, Ramsay C, et al. Is it worthwhile to conduct a randomized controlled trial of glaucoma screening in the united kingdom? *Journal of health services research & policy*. 2014;19(1):42-51.

143. Xu X, Ma YY, Zou HD. Cost-utility analysis of cataract surgery in advanced glaucoma patients. *Journal of glaucoma*. 2016;25(7):e657-e662.

144. Sun X, Zhang S, Wang N, et al. Utility assessment among patients of primary angle closure/glaucoma in china: A preliminary study. *British journal of ophthalmology*. 2009;93(7):871-874.

145. Burr JM, Botello-Pinzon P, Takwoingi Y, et al. Surveillance for ocular hypertension: An evidence synthesis and economic evaluation. *Health Technol Assess*. 2012;16(29):1-271.

146. Javanbakht M, Azuara-Blanco A, Burr JM, et al. Early lens extraction with intraocular lens implantation for the treatment of primary angle closure glaucoma: An economic evaluation based on data from the EAGLE trial. *BMJ open*. 2017;7(1):e013254.

147. Blumberg DM, Vaswani R, Nong E, Al-Aswad L, Cioffi GA. A comparative effectiveness analysis of visual field outcomes after projected glaucoma screening using SD-OCT in african american communities. *Investigative ophthalmology & visual science*. 2014;55(6):3491-3500.

148. Kymes SM, Plotzke MR, Kass MA, Boland MV, Gordon MO. Effect of patient's life expectancy on the cost-effectiveness of treatment for ocular hypertension. *Archives of ophthalmology*. 2010;128(5):613-618.

149. Rein DB, Wittenborn JS, Lee PP, et al. The cost-effectiveness of routine office-based identification and subsequent medical treatment of primary open-angle glaucoma in the united states. *Ophthalmology*. 2009;116(5):823-832.

150. Stewart W, Stewart J, Mychaskiw M. Cost-effectiveness of latanoprost and timolol maleate for the treatment of glaucoma in scandinavia and the united kingdom, using a decision-analytic health economic model. *Eye*. 2009;23(1):132-140.

151. Guedes RAP, Guedes VMP, Mello Gomes CE de, Chaoubah A. Maximizing cost-effectiveness by adjusting treatment strategy according to glaucoma severity. *Medicine*. 2016;95(52).

152. Kaplan RI, De Moraes CG, Cioffi GA, Al-Aswad LA, Blumberg DM. Comparative cost-effectiveness of the baerveldt implant, trabeculectomy with mitomycin, and medical treatment. *JAMA ophthalmology*. 2015;133(5):560-567.

153. Van Gestel A, Severens JL, Webers CA, Beckers HJ, Jansonius NM, Schouten JS. Modeling complex treatment strategies: Construction and validation of a discrete event simulation model for glaucoma. *Value in Health*. 2010;13(4):358-367.

154. Newman-Casey PA, Salman M, Lee PP, Gatwood JD. Cost-utility analysis of glaucoma medication adherence. *Ophthalmology*. Published online 2019.

155. Hernández R, Burr J, Vale L, et al. Monitoring ocular hypertension, how much and how often? A cost-effectiveness perspective. *British journal of ophthalmology*. 2016;100(9):1263-1268.

156. Gazzard G, Konstantakopoulou E, Garway-Heath D, et al. Selective laser trabeculoplasty versus drops for newly diagnosed ocular hypertension and glaucoma: The LiGHT RCT. *Health technology assessment (Winchester, England)*. 2019;23(31):1.

157. Patel V, Ahmed I, Podbielski D, Falvey H, Murray J, Goeree R. Cost-effectiveness analysis of standalone trabecular micro-bypass stents in patients with mild-to-moderate open-angle glaucoma in canada. *Journal of medical economics*. 2019;22(4):390-401.

158. Stewart WC, Stewart JA, Nassar QJ, Mychaskiw MA. Cost-effectiveness of treating ocular hypertension. *Ophthalmology*. 2008;115(1):94-98.

159. Brown GC, Brown MM. Patient preference-based comparative effectiveness and cost-utility analysis of the prostamides for open-angle glaucoma. *Journal of Ocular Pharmacology and Therapeutics*. 2019;35(3):145-160.

160. Ordóñez JE, Ordóñez A, Osorio UM. Cost-effectiveness analysis of iStent trabecular micro-bypass stent for patients with open-angle glaucoma in colombia. *Current medical research and opinion*. 2019;35(2):329-340.

161. Teus MA, Belda JI, Lavin C, et al. Cost-effectiveness analysis of iStent inject® implantation during cataract surgery compared to cataract surgery alone for mild to moderate open-angle glaucoma patients in spain. *Expert Review of Ophthalmology*. 2021;16(4):319-328.

162. Nieland K, Labbé A, Schweitzer C, et al. A cost-effectiveness analysis of iStent inject combined with phacoemulsification cataract surgery in patients with mild-to-moderate open-angle glaucoma in france. *PLoS One*. 2021;16(6):e0252130.

163. Fea AM, Cattel F, Gandolfi S, Buseghin G, Furneri G, Costagliola C. Cost-utility analysis of trabecular micro-bypass stents (TBS) in patients with mild-to-moderate open-angle glaucoma in italy. *BMC health services research*. 2021;21(1):1-12.

164. King AJ, Fernie G, Hudson J, et al. Primary trabeculectomy versus primary glaucoma eye drops for newly diagnosed advanced glaucoma: TAGS RCT. *Health Technology Assessment*. 2021;25(72):1-158.

165. Choi JA, Song LD, Choi S, Park SM, Kwon JW, Jee D. The cost-effectiveness of medication, laser trabeculoplasty, and trabeculectomy for treatment of open-angle glaucoma in south korea. *Medicine*. 2019;98(2).

166. Ahmed IIK, Podbielski DW, Patel V, et al. A canadian cost-utility analysis of two trabecular micro-bypass stents at time of cataract surgery in patients with open-angle glaucoma. *Ophthalmology Glaucoma*. Published online 2019.

167. John D, Parikh R. Cost-effectiveness of community screening for glaucoma in rural india: A decision analytical model. *Public health*. 2018;155:142-151.

168. Azuara-Blanco A, Banister K, Boachie C, et al. Automated imaging technologies for the diagnosis of glaucoma: A comparative diagnostic study for the evaluation of the diagnostic accuracy, performance as triage tests and cost-effectiveness (GATE study). *Health Technology Assessment*. Published online 2016.

169. Thomas S, Hodge W, Malvankar-Mehta M. The cost-effectiveness analysis of teleglaucoma screening device. *PLoS One*. 2015;10(9).

170. Gestel A van, Schouten JS, Beckers HJ, Severens JL, Hendrikse F, Webers CA. The long term effectiveness and cost-effectiveness of initiating treatment for ocular hypertension. *Acta ophthalmologica*. 2014;92(6):513-523.

171. Wolfram C, Lorenz K, Breitscheidel L, Verboven Y, Pfeiffer N. Health-and vision-related quality of life in patients with ocular hypertension or primary open-angle glaucoma. *Ophthalmologica*. 2013;229(4):227-234.

172. Bartelt-Hofer J, Flessa S. Comparative efficacy and cost–utility of combined cataract and minimally invasive glaucoma surgery in primary open-angle glaucoma. *International Ophthalmology*. Published online 2020:1-11.

173. Li EY, Tham CC, Chi SC, Lam DS. Cost-effectiveness of treating normal tension glaucoma. *Investigative ophthalmology & visual science*. 2013;54(5):3394-3399.

174. Crane GJ, Kymes SM, Hiller JE, Casson R, Martin A, Karnon JD. Accounting for costs, QALYs, and capacity constraints: Using discrete-event simulation to evaluate alternative service delivery and organizational scenarios for hospital-based glaucoma services. *Medical Decision Making*. 2013;33(8):986-997.

175. Philippe Nordmann J, Lafuma A, Berdeaux G. Modelling the lifetime economic consequences of glaucoma in france. *Journal of medical economics*. 2009;12(1):9-16.

176. Vaahtoranta-Lehtonen H, Tuulonen A, Aronen P, et al. Cost effectiveness and cost utility of an organized screening programme for glaucoma. *Acta Ophthalmologica Scandinavica*. 2007;85(5):508-518.

177. Pen CL, Ligier M, Berdeaux G. Cost-effectiveness and cost-utility analysis of travoprost versus latanoprost and timolol in the treatment of advanced glaucoma in five european countries: Austria, france, germany, the netherlands and the united kingdom. *Journal of Medical Economics*. 2005;8(1-4):67-84.

178. Tang J, Liang Y, O'Neill C, Kee F, Jiang J, Congdon N. Cost-effectiveness and cost-utility of population-based glaucoma screening in china: A decision-analytic markov model. *The Lancet Global Health*. 2019;7(7):e968-e978.

179. Burr JM, Mowatt G, Hernández RA, et al. The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: A systematic review and economic evaluation. Published online 2007.

180. John D, Parikh R. Cost-effectiveness and cost utility of community screening for glaucoma in urban india. *Public Health*. 2017;148:37-48.

181. Stein JD, Kim DD, Peck WW, Giannetti SM, Hutton DW. Cost-effectiveness of medications compared with laser trabeculoplasty in patients with newly diagnosed open-angle glaucoma. *Archives of ophthalmology*. 2012;130(4):497-505.