JAMA Ophthalmology | Original Investigation

Utility Assessment of Moderate to Severe Thyroid Eye Disease Health States

Terry J. Smith, MD; Kimberly Cockerham, MD; Gary Lelli, MD; Chitra Choudhary, MD; Stephanie Taylor, PhD; Naina Barretto, PhD; Ashley Enstone, BSc, MRes; Louisa Oliver, BSc; Judah Lynch, BSc, MRes; Robert J. Holt, PharmD, MBA

IMPORTANCE Thyroid eye disease (TED) results in varying degrees of proptosis and diplopia negatively affecting quality of life (QoL), producing possibly substantial visual changes, disfigurement, and disability.

OBJECTIVE To determine the association of varying TED severities with QoL in a non-TED population by assessing health state utility scores.

DESIGN, SETTING, AND PARTICIPANTS This qualitative study, conducted from April 20, 2020, to April 29, 2021, assessed health states for active, moderate-severe TED, and values were elicited using time trade-off methods. Six health states of varying severity were determined from 2 placebo-controlled clinical trials (171 patients with TED and clinical activity score ≥4, ±diplopia/proptosis) and refined using interviews with US patients with TED (n = 6). Each health state description was validated by interviews with additional TED patient advocates (n = 3) and physician experts (n = 3). Health state descriptions and a QOL questionnaire were piloted and administered to a general population. Visual analog scales (VASs) were also administered to detect concurrence of the findings.

MAIN OUTCOMES AND MEASURES TED health state utility scores and whether they differ from one another were assessed using Shapiro-Wilk, Kruskal-Wallis, pairwise Wilcoxon rank sum, and paired *t* tests.

RESULTS A total of 111 participants completed time trade-off interviews. The mean (SD) utility value was 0.44 (0.34). The lowest (worse) mean utility value was observed in the most severe disease state (constant diplopia/large proptosis) with 0.30 (95% CI, 0.24-0.36), followed by constant diplopia/small proptosis (0.34; 95% CI, 0.29-0.40), intermittent or inconstant diplopia/large proptosis (0.43; 95% CI, 0.36-0.49), no diplopia/large proptosis (0.46; 95% CI, 0.40-0.52), and intermittent or inconstant diplopia/small proptosis (0.52; 95% CI, 0.45-0.58). The highest (best) mean value, 0.60 (95% CI, 0.54-0.67), was observed for the least severe disease state (no diplopia/small proptosis).

CONCLUSIONS AND RELEVANCE These findings suggest that patients with active, moderate-severe TED may have substantial disutility, with increasing severity of proptosis/diplopia more likely to have detrimental associations with QoL. These health state scores may provide a baseline for determining QoL improvement in these TED health states (utility gains) treated with new therapies.

Invited Commentary page 166

Supplemental content

Author Affiliations: Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, Ann Arbor, Michigan (Smith); Division of Metabolism, Endocrine and Diabetes, Michigan Medicine and University of Michigan, Ann Arbor (Smith); Stanford Department of Ophthalmology, Palo Alto, California (Cockerham); Central Valley Eye Medical Group, Stockton, California (Cockerham); Senta Clinic, San Diego, California (Cockerham); Weill Cornell Medical Center, New York, New York (Lelli); Kansas University Medical Center, Kansas City (Choudhary); Horizon Therapeutics, Deerfield, Illinois (Taylor, Barretto, Holt); Adelphi Values PROVE, Manchester, United Kingdom (Enstone, Oliver, Lynch).

Corresponding Author: Robert J. Holt, PharmD, MBA, 1721 N Woods Way, Vernon Hills, IL 60061 (rholt@horizontherapeutics.com)

JAMA Ophthalmol. 2023;141(2):159-166. doi:10.1001/jamaophthalmol.2022.3225
Published online December 29, 2022. Corrected on February 16, 2023.

Thyroid eye disease (TED) is an inflammatory autoimmune condition most often presenting with Graves disease. In the US, TED prevalence is between 0.16% and 0.25%.[1,2] Those at risk for more severe disease include male individuals, those older than 65 years, tobacco users, and those with a high level of thyroid autoantibodies.[1,3,4]

TED begins with orbital and periocular inflammation typically followed by a less active but still symptomatic chronic disease course.[5-7] Active TED, defined by a clinical activity score (CAS), includes orbital/periocular pain, swelling, and redness with proptosis and diplopia present in more severe disease.[5-7]

Increasing levels of proptosis and diplopia are associated with worsening QoL.[8,9] These manifestations may prevent carrying out daily tasks such as driving and reading and have been associated with substantial psychosocial distress, most notably anxiety and depression from changes in appearance and visual function.[10,11]

Health state utilities are cardinal values that reflect an individual's preference for different health outcomes, such as trading off a shorter period of time in perfect health vs a longer time with a disease that reduces QoL. Utility values play an important role in cost-effectiveness analyses as well as health technology appraisals. They allow comparisons of QoL impact among differing diseases. To our knowledge, published utility values have not been defined for TED-induced health states with varying levels of proptosis and diplopia, against a backdrop (anchor health state) of heightened inflammation. Therefore, the primary objective of this study was to elicit utility values for health states associated with moderate-severe active TED of varying severities. These were assessed by magnitude of proptosis and diplopia (separately and combined), to demonstrate QoL burden associated with TED and allow for future assessment of QoL gains with treatments.

Key Points

Question What are utility values, or quality of life (QoL) burden measures, for the various thyroid eye disease (TED) health states?

Findings In this qualitative study of 111 participant interviews, the lowest mean value (representing the worst QoL outcome) was observed for the most severe health state (large proptosis and constant diplopia), and the highest mean value was observed for the least severe health state (low proptosis and no diplopia).

Meaning These findings suggest moderate-severe TED is associated with worse utility values, with increasing severities of proptosis and diplopia having greater association.

Methods

Overview of Study Design

This qualitative analysis, including the collection and evaluation of protected health information, was compliant with the Health Insurance Portability and Accountability Act and adhered to the ethical principles outlined in the Declaration of Helsinki as amended in 2013. This study was conducted from April 20, 2020, to April 29, 2021. Institutional review board approval was not required for this study as it is defined as market research and is exempt. Patients agreed to participate in the interview and were assured confidentiality. All patients, health care professionals, and members of the public voluntarily participated in this study following identification from a patient advisory council, an expert panel, and a general population US proprietary panel, respectively. Participants were identified using a predefined screening questionnaire. All respondents provided written consent before study recruitment. All respondents participating in interview development or validation remained anonymous according to regulations and practice guidelines of market research governing bodies. These included the Council of American Survey Research Organizations and followed the Standards for Reporting Qualitative Research (SRQR) reporting guidelines for qualitative studies. Respondents were remunerated for their time according to fair market values.

Health states were developed using results from 2 placebo-controlled clinical trials in patients with active, moderate-severe TED with differing severities of proptosis and diplopia and interviews with patients with TED.[12-14] The health state descriptions were validated through expert TED physician and patient advocate interviews. Utility scores were elicited using time trade-off (TTO) methods with a finite 10-year horizon.[15] TTO methodology is widely accepted for valuing preference-based health-related QoL. The Measurement and Valuation of Health TTO protocol was used, involving the respondent trading off years with a poor QoL for a shorter period of life in full health.[16] A utility score (decimalized) was then calculated using the interviewee's value from the TTO interview between perfect health (1) and death (0) based on these trade-offs. Lower utility score represents worse QoL. The TTO interviews were conducted by 2 trained TTO interviewers who were independent of the study authors.

Health State Development

Each health state description represents patients with active TED with varying severities of proptosis and diplopia. Six health states (**Box**) were defined through a development and validation process (eFigure in Supplement 1) according to baseline clinical features of patients participating in the 2 clinical trials (171 patients with TED, CAS ≥4, ±diplopia/proptosis).[12-14] Individual patient data were reviewed and categorized by proptosis degree (<3 mm or ≥3 mm [above upper limit of normal

Box. TED Health States in Order of Severity[a]

1. No diplopia, small proptosis (<3 mm)
2. No diplopia, large proptosis (≥3 mm)
3. Intermittent or inconstant diplopia, small proptosis (<3 mm)
4. Intermittent or inconstant diplopia, large proptosis (≥3 mm)
5. Constant diplopia, small proptosis (<3 mm)
6. Constant diplopia, large proptosis (≥3 mm)

[a] In order of increasing severity.

Abbreviation: TED, thyroid eye disease.

for race and sex]) and diplopia (none, intermittent, inconstant, or constant). These health state descriptions were further refined using telephone interviews (1 hour) with 6 US patients with TED. A semistructured interview guide was used subsequently to explore how the disease was associated with patient QoL (eAppendix 1 in Supplement 1).

Descriptive language development focused on how each health state was associated with physical, mental, and social QoL. All of these health state descriptions incorporated QoL outcomes, including limitations on daily activities, problems with self-image, reduced social activities and interactions with others, inability to work/reduced productivity, anxiety about the future, coping mechanism, and treatments. Photographs of patients with CAS scores of 4 or greater with varying proptosis levels and language (with imaging) describing the association with diplopia were also used to aid the respondents' understanding the outcome of each health state (eAppendix 2 in Supplement 1).

Health State Validation

To confirm the validity and ensure that each health state was representative of patients with TED and its association with QoL, interviews were conducted with 1 oculoplastic surgeon (G.L.), 1 comprehensive ophthalmologist (K.C.), and 1 endocrinologist (T.J.S.) with more than 4 years in practice, and 3 interviews were conducted with members of a patient advisory council, each with moderate-severe TED.

TTO Valuation Phase

Participants

Recruitment of a demographically representative sample comprising 100 interviewees (without TED) in the US was accomplished using a general population US proprietary panel identified by a third-party recruitment agency. Participants from the following race and ethnicity groups were included: Asian, Black or African American, Hispanic, and White. Race and ethnicity data were obtained from the screener questionnaire used for the TTO study. Based on existing guidance for health utility studies, a sample size of 100 was deemed adequate for achieving demographically representative sampling of the US general population.[17] General population interviews were used to minimize bias because this cohort does not have a vested interest in the treatment of particular health states, as opposed to patients with TED.[18] A 10% dropout rate was assumed before the interview stage. The selection criteria included (1) adult US residents 18 years or older willing to give explicit informed written consent to participate and have their answers audio-recorded, (2) no participation in market research within the past 6 months, and (3) no conflict of interest (defined as employment in health care or market research). In addition, recruitment quotas were applied to ensure a geographically and demographically representative sample.

Pilot Study

Pilot interviews with members of the public were conducted for linguistic validation, using the same methodology intended for formal interviews. These interviews validated the approach and were included in the final analysis.

Utility Interview Procedures and Scoring

The standard TTO method of eliciting utilities was used in this study and included a 10-year time horizon.[19] A time horizon beyond 10 years lacks comparability with other utility values previously reported using this methodology.[15,20]

The TTO assessment was chosen because the most common QoL measure from which utilities can be derived, the EuroQol 5 Dimension 5 Level (EQ-5D-5L) questionnaire, includes domains not adequately capturing all relevant dimensions of TED QoL. TED disease-specific measures, such as the Graves ophthalmopathy QoL (GO-QoL) questionnaire, have good validity, reliability, and responsiveness; however, that questionnaire has not been mapped previously or correlated with specific TED utility values.

To prevent bias during the utility study, each health state was titled with different geometric shapes. This allowed each health state to be referred to without any implication of progression and/or severity.

The visual analog scale (VAS) interviews were conducted first to familiarize participants with the health state descriptions and the concept of evaluating health states. The VAS exercise provided additional data to the TTO analysis for comparison and confirmation purposes, as well as valuing their own health at the time of analysis. The VAS method used a scale between 0 (worst health) and 100 (best health). Both TTO and VAS interviews were conducted using the health state descriptions in a randomized order in the general US population (without TED).[16]

Statistical Analysis

Data were analyzed in the statistical software environment CRAN R using the packages dplyr, tidyr, and ggpubr (Kurt Hornik, Friedrich Leisch).[21] Continuous variables (TTO scores and VAS scores by health state and the interviewee's own health) were analyzed descriptively, and the sample size, mean, median, SD, SE, and 95% CIs were presented for each variable. Histograms of continuous variables were developed to visually represent the distribution of each variable. A Shapiro-Wilk test was performed to assess distribution of each variable and determine the optimal test for significance (paired *t* test for normal data or Kruskal-Wallis test for nonnormal data).

The Kruskal-Wallis test was used to determine the presence of overall significant difference between utility values of multiple health states (2-sided *P* value; cutoff .05 for significance). A pairwise Wilcoxon rank sum test assessed utility values and VAS scores to determine whether individual health states differed (2-sided *P* value). *P* values were not adjusted to multiple comparisons. Exploratory subgroup analyses were conducted using a Kruskal-Wallis test if the subpopulation sample size was greater than 25, to determine the differences across health states or between multiple subgroups.

A weighted analysis was performed to test whether differing proportions of health states would affect the results and whether utility values in this study are generalizable to the US TED population. Because the prevalence of these health states within TED is unknown, we estimated the percentage of each health state at baseline from the 2 trials.[13,14]

Table 1. Utility Values for Each Health State[a]

Health state	Total No. of respondents	Mean (SD)	Median	95% CI	IQR
No diplopia and small proptosis	111	0.60 (0.34)	0.70	0.54-0.67	0.36-0.91
No diplopia and large proptosis	111	0.46 (0.32)	0.40	0.40-0.52	0.22-0.73
Intermittent or inconstant diplopia and small proptosis	111	0.52 (0.33)	0.50	0.45-0.58	0.28-0.81
Intermittent or inconstant diplopia and large proptosis	111	0.43 (0.33)	0.38	0.36-0.49	0.11-0.70
Constant diplopia and small proptosis	111	0.34 (0.31)	0.30	0.29-0.40	0.03-0.53
Constant diplopia and large proptosis	111	0.30 (0.31)	0.23	0.24-0.36	0.03-0.5
Average TED utility value	111	0.44 (0.34)	0.40	0.42-0.47	NA

Abbreviations: NA, not applicable; TED, thyroid eye disease.

[a] Scores rounded to 2 decimal places.

Figure. Distribution of Utility Values by Health State

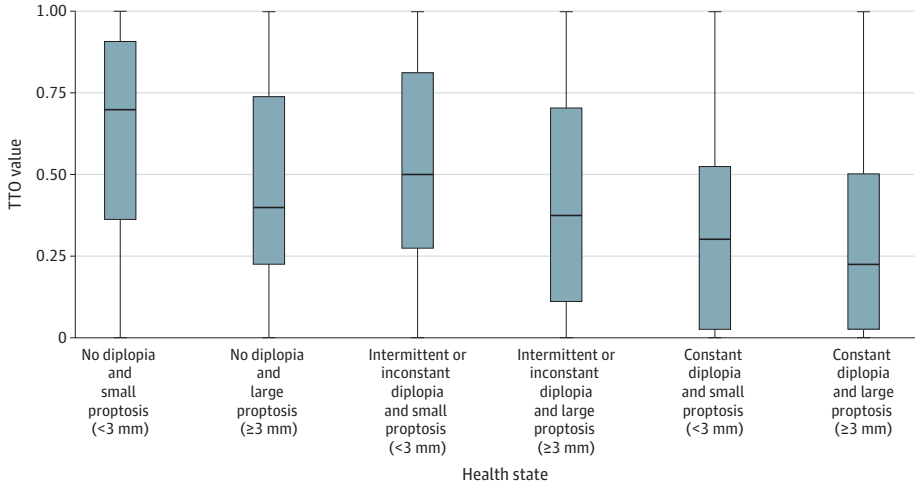

Figure displays median Q1, Q3. TTO indicates time trade-off.

Results

Participant Baseline Characteristics

A total of 111 participants (pilot, n = 10; main study, n = 101) completed the TTO interviews. The largest proportion of participants were 55 to 64 years old (47 [42.0%]), female individuals (75 [68.0%]), wore glasses (90 [81.0%]), and 46 (41%) self-identified as having received a visual impairment diagnosis, including near or farsightedness (19 [17.0%]). Participants from the following race and ethnicity groups were included: 4 Asian (3.6%), 11 Black or African American (9.9%), 4 Hispanic (3.6%), and 92 White (82.9%). Some characteristics differed from the aggregate US population, such as older age, higher percentage of female individuals, lower income level, and higher percentage of visual impairment; however, these differences more closely resembled those reported in the US TED population (eTable 1 in Supplement 1).[12-14]

Utility Values

Overall, mean utility values were found to decrease with increasing severity of the health states (Box). As would be expected, the highest mean utility values (least association with QoL) were reported for the least severe health state of no diplopia and small proptosis (0.60; 95% CI, 0.54-0.67) (**Table 1**). However, a smaller decrease in utility was detected when the severity of proptosis remained constant, and there was minimal change in those with lower diplopia states (ie, no diplopia or intermittent/inconstant diplopia) as highlighted by overlapping CIs in an analysis of variability and central tendency shown in the **Figure**. The lowest (worse) mean utility value was observed in the most severe disease state (constant diplopia/large proptosis) with 0.30 (95% CI, 0.24-0.36), followed by constant diplopia/small proptosis (0.34; 95% CI, 0.29-0.40), intermittent or inconstant diplopia/large proptosis (0.43; 95% CI, 0.36-0.49), no diplopia/large proptosis (0.46; 95% CI, 0.40-0.52), and intermittent or inconstant diplopia/small proptosis (0.52; 95% CI, 0.45-0.58).

Differences Between Health States

The pairwise Wilcoxon rank sum test indicated significant differences in utility values in 9 of 15 possible health state comparisons (**Table 2**). The other 6 comparisons were neighboring health states with small incremental differences in disease

Table 2. Overview of Significance Between Health States[a]

Health state	No diplopia and small proptosis	No diplopia and large proptosis ptosis	Intermittent or inconstant diplopia and small proptosis	Intermittent or inconstant diplopia and large proptosis	Constant diplopia and small proptosis
No diplopia and large proptosis	.001	NA	NA	NA	NA
Intermittent or inconstant diplopia and small proptosis	.053	.21	NA	NA	NA
Intermittent or inconstant diplopia and large proptosis	<.001	.40	.053	NA	NA
Constant diplopia and small proptosis	<.001	.001	<.001	.10	NA
Constant diplopia and large proptosis	<.001	<.001	<.001	<.001	.27

Abbreviation: NA, not applicable.

[a] Utility values from Table 1 were compared.

Table 3. Weighted Analysis of Utility Values by Health State

Health state	Total No. of respondents	% Sample[a]	Weighted utility score
No diplopia and small proptosis	22	12.9	0.080
No diplopia and large proptosis	24	14	0.064
Intermittent or inconstant diplopia and small proptosis	33	19.3	0.099
Intermittent or inconstant diplopia and large proptosis	57	33.3	0.142
Constant diplopia and small proptosis	14	8.2	0.028
Constant diplopia and large proptosis	21	12.3	0.037
Average weighted utility value	171	100	0.45

[a] Health state percentages taken from the 2 randomized trials.[7]

severity, with 1 comparison (intermittent/inconstant with small proptosis vs no diplopia/small proptosis) approaching significance. The weighted analysis of utility values produced a very similar utility value (0.45) to the mean (SD) utility value calculated in this study of 0.44 (0.34) (**Table 3**).

Analysis of Subpopulations

Most subgroups large enough to test (ie, n >25) demonstrated an overall significant difference between health states (Kruskal-Wallace test). Using the no diplopia/small proptosis health state, these included participants aged 55 to 64 years (mean [SD], 0.65 [0.34]; 95% CI, 0.03-1.00); female individuals (mean [SD], 0.61 [0.34], 95% CI, 0-1.00); male individuals (mean [SD], 0.59 [0.34], 95% CI, 0.03-1.00); White race (mean [SD], 0.64 [0.31]; 95% CI, 0.03-1.00); working full-time (mean [SD], 0.53 [0.34]; 95% CI, 0.03-1.00); eyeglass wearers (mean [SD], 0.61 [0.35]; 95% CI, 0-1.00); individuals with visual impairment (mean [SD], 0.59 [0.34]; 95% CI, 0-1.00); those without visual impairment (mean [SD], 0.61 [0.33], 95% CI, 0-1.00); those who previously smoked (mean [SD], 0.61 [0.32], 95% CI, 0-1.00); and those who never smoked (mean [SD], 0.60-0.35; 95% CI, 0-1.00) (eTables 2-10 in Supplement 1).

Results From the VAS Exercise

As with utility values, highest and lowest VAS scores were associated with no diplopia/small proptosis (51.71; 95% CI, 47.67-55.74) and constant diplopia/large proptosis (27.42; 95% CI, 24.41-30.44), respectively (**Table 4**).

Differences in VAS scores were found within the constant diplopia and large proptosis health state, when participants in the highest quartile (80.69) for their own health score were compared with those in the lowest quartile (74.59) (lowest quartile mean = 21.59 vs highest quartile mean = 33.61; $P = .01$).

There was no difference between own health VAS scores (mean [SD], 77.64 [80.00]) and VAS scores for the US general population (mean [SD], 80.4 [not reported]; 95% CI, 74.59-80.69; $P = .08$), supporting the hypothesis that the population interviewed was similar in general health to the overall population.[22] Differences existed between TED health state VAS scores and the mean VAS score for the US general public across all health states (eTable 11 in Supplement 1). This emphasizes the disutility associated with the TED health states. The mean (SD) VAS score (when converted to the utility scale) was 0.38 (.20) vs an average mean (SD) utility value of 0.44 (.34).

Discussion

To our knowledge, based on a thorough literature search (eAppendix 3 in Supplement 1), this was the first study validating health states of moderate-severe TED and assessing utility values from the US general population using the TTO approach. Rationale for using the general population to value health states in TED derives from an assumption that these individuals harbor no investment in the disease, unlike those with TED. Patient valuation could potentially bias study results because these individuals could adapt to their health state over time and therefore assign higher values to their own health state.[18] These higher values may be associated with patients developing coping mechanisms, allowing for improved function-

Table 4. Visual Analog Scale (VAS) Scores for Each Health State (n = 111)[a]

Health state	Mean (SD)	Median (95% CI)
No diplopia and small proptosis	51.71 (21.44)	50.00 (47.67-55.74)
No diplopia and large proptosis	38.95 (17.63)	40.00 (35.64-42.27)
Intermittent or inconstant diplopia and small proptosis	45.38 (17.74)	45.00 (42.04-48.72)
Intermittent or inconstant diplopia and large proptosis	36.52 (18.14)	35.00 (33.11-39.94)
Constant diplopia and small proptosis	30.56 (17.83)	25.00 (27.20-33.91)
Constant diplopia and large proptosis	27.42 (16.04)	25.00 (24.41-30.44)
Average VAS score	38.41 (19.94)	40.00 (36.90-39.94)
Own health	77.64 (80.00)	1.54 (74.59-80.69)

[a] Scores rounded to 2 decimal places; 0 = worst health; 100 = perfect health.

ing in daily activities.[17] Further, ethical concerns regarding patient valuation of health states may arise given the potential for emotional distress provoked by imagining hypothetical situations that they are unable to experience.[23]

Each TTO exercise followed the standard Measurement and Valuation of Health protocol, which is one of the most widely adopted techniques for TTO valuation and was used by EuroQol for the valuation of the EQ-5D-3L questionnaire.[15] It is acknowledged that EuroQoL has introduced another protocol, the EuroQoL Valuation Technology (EQ-VT), which includes a discrete choice component. Although utilities for quality-adjusted life-year calculations cannot be generated from discrete choice as a stand-alone technique, this represents an alternative approach to the complementary VAS scoring used in this study.[15] The EQ-VT was deemed to be more difficult to administer due to the time required to complete and keep interviewees engaged during the peak of the COVID-19 lockdown.

The mean utility value derived across these 6 health states of patients with active, moderate-severe TED was 0.44. This indicates a substantial level of disutility as compared with the value of 0.88 found in a study investigating QoL in hyperthyroid Graves disease and a chronic, mild TED subgroup.[24] These utility values highlight the substantial impact TED imposes when compared with the US general population (0.85).[22] Further, this mean utility value was nearly identical to that calculated with a weighted analysis using proportions of these health states found in 2 controlled trials. Results found here might thus be validly generalized to the population with active, moderate-severe TED.[12]

The calculated overall utility value in this study (0.44) allows contextualizing the results to other diseases. It is similar to or lower (worse) than values reported for several severely debilitating conditions, such as blindness with no light perception (0.26; CI, 0.19-0.33), noninfectious uveitis (median, 0.975; IQR, 0.8-1.0), and age-related macular degeneration (mean, 0.81; 95% CI, 0.76-0.86) (all comparisons used the TTO methodology).[25-27] This congruence highlights the severe burden moderate-severe active TED places on patients as perceived by an unbiased unaffected population.

Severity of proptosis and diplopia emerged as the key variables associated with utility values for the 6 health states studied here. These values decreased (worsened) as severity of diplopia increased; however, the decrement in those health states with large proptosis and milder diplopia was more modest. Both the presence of large proptosis and constant diplopia drove health state scores lower. Constant diplopia appears to produce a floor effect because scores clustered at their lowest when present in a health state, regardless of proptosis magnitude. This finding underscores the particular burden imposed by both large proptosis and constant diplopia.

No major differences emerged from 6 comparisons of the mostly neighboring middle health states. This may relate to a general similarity of association observed in these health states with daily life or our inability to discriminate differences. For instance, changes from intermittent/inconstant diplopia to no diplopia may be difficult to detect, thus resulting in only small utility value changes. As observed, moderate-severe TED health states were associated with worse utility values. This finding highlights the need for further identification of each health state characteristics having the greatest association with QoL. Furthermore, ongoing studies are attempting to bridge and validate these utility findings with the Graves Orbitopathy-Specific Quality of Life (GO-QOL) questionnaire, a specific TED QoL instrument.

Limitations

This study had some limitations. Despite best recruitment efforts, our cohort was older and included more women and more individuals with visual impairment than the US general population. However, the differences in our sample more closely approximated the known demographic characteristics of the US TED population.[28] These differences serendipitously increased our confidence in the findings emanating from the interviewed population.

Although TED persists throughout life, the fixed time interval of 10 years used for health state assessment in this TTO may not represent a large subset of patients with active disease whose inflammatory symptoms are shorter lived. Primary progressive symptoms driving disutility identified here (proptosis and diplopia) can persist for extended time periods with variable severity over a lifetime, as was assessed here. Further, the 10-year time horizon is recommended by Health Technology Assessment as facilitating comparisons across treatments and diseases, despite very different patterns of disease manifestations.[15,20] The TED health state descriptions studied here were based on patients with highly active, moderate-severe disease and thus may not represent individuals with milder or less active disease.

Conclusions

Results of this qualitative study suggest that patients with moderate-severe, active TED may have substantial disutility. The association of TED with negative QoL has been well documented; however, using the unbiased TTO method in a general US population to assess this association (health score util-

ity scoring) is novel, to the best of our knowledge. The results suggest that utility values decrease with increasing severity of TED symptoms as is reflected by the presence of diplopia and/or proptosis in 6 different health states. Substantial differences were found between nonneighboring health states and in certain subpopulations. These very low utility values suggest the substantial QoL burden potentially confronting patients with TED and highlight the need for additional studies in this area. These data might now serve as baseline for prospectively assessing improvements (utility gains) with therapies.

ARTICLE INFORMATION

Accepted for Publication: October 27, 2022.

Published Online: December 29, 2022. doi:10.1001/jamaophthalmol.2022.3225

Correction: This article was corrected on February 16, 2023, to update the status of the article so that it is now Open Access.

Author Contributions: Drs Smith and Holt had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Smith, Cockerham, Lelli, Taylor, Barretto, Enstone, Oliver, Holt.
Acquisition, analysis, or interpretation of data: Cockerham, Lelli, Choudhary, Taylor, Barretto, Enstone, Oliver, Lynch, Holt.
Drafting of the manuscript: Cockerham, Choudhary, Taylor, Barretto, Oliver, Lynch, Holt.
Critical revision of the manuscript for important intellectual content: Smith, Cockerham, Lelli, Choudhary, Taylor, Barretto, Enstone, Holt.
Statistical analysis: Oliver, Lynch.
Obtained funding: Taylor, Holt.
Administrative, technical, or material support: Cockerham, Lelli, Taylor, Barretto, Enstone, Oliver, Holt.
Supervision: Lelli, Choudhary, Taylor, Barretto, Enstone, Holt.

Conflict of Interest Disclosures: Dr Smith reported being issued US patents covering the use of IGF-1 receptor inhibitors in thyroid eye disease, which are held by UCLA and the Lundquist Institute and receiving consultant fees from Horizon Therapeutics, Veridian, and Lundbeck outside the submitted work. Dr Cockerham reported receiving consultant fees from Horizon Pharmaceuticals and Viridian during the conduct of the study. Drs Lelli and Choudhary reported receiving consultant fees from Horizon Therapeutics outside the submitted work. Dr Taylor reported being an employee of Horizon Therapeutics during the study and completion of the article. Dr Barretto reported being an employee of Horizon Therapeutics and owning stock in the company outside the submitted work. Dr Enstone reported receiving grants from Horizon Therapeutics and being an employee of Adelphi Values PROVE during the conduct of the study. Drs Enstone, Oliver, and Lynch reported being employees of Adelphi Values PROVE (Adelphi Values received budgetary compensation from Horizon Therapeutics for conduct of the study). Dr Holt reported being an employee of Horizon Therapeutics and owning stock in the company outside the submitted work. No other disclosures were reported.

Funding/Support: This study was funded by Horizon Therapeutics.

Role of the Funder/Sponsor: Author employees of the funder did have a role in the design and conduct of the study; analysis and interpretation of the data; preparation and approval of the manuscript; and the decision to submit for publication but not the interviews for or collection of data.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Lazarus JH. Epidemiology of Graves' orbitopathy (GO) and relationship with thyroid disease. *Best Pract Res Clin Endocrinol Metab*. 2012;26(3):273-279. doi:10.1016/j.beem.2011.10.005

2. Hyman L, Penne R, Fujino D, et al. Prevalence and associated factors for thyroid eye disease (TED) in the AAO IRIS registry. *Invest Ophthalmol Vis Sci*. 2020;61(7):5425.

3. Sikder S, Weinberg RS. Thyroid eye disease: pathogenesis and treatment. *Ophthalmologica*. 2010;224(4):199-203. doi:10.1159/000260224

4. Manji N, Carr-Smith JD, Boelaert K, et al. Influences of age, gender, smoking, and family history on autoimmune thyroid disease phenotype. *J Clin Endocrinol Metab*. 2006;91(12):4873-4880. doi:10.1210/jc.2006-1402

5. Bahn RS. Graves' ophthalmopathy. *N Engl J Med*. 2010;362(8):726-738. doi:10.1056/NEJMra0905750

6. Wang Y, Padnick-Silver L, Francis-Sedlak M, Holt RJ, Foley C, Douglas RS. Inflammatory and noninflammatory thyroid eye disease: comparison of disease signs, symptoms, and quality of life in US patients. *Endocr Pract*. 2022;28(9):842-846. doi:10.1016/j.eprac.2022.06.003

7. Cook LD. Oxford handbook of ophthalmology. *J Neuroophthalmol*. 2009;29(1):81-82. doi:10.1097/01.wno.0000347033.35456.80

8. Egle UT, Kahaly GJ, Petrak F, et al. The relevance of physical and psychosocial factors for the quality of life in patients with thyroid-associated orbitopathy (TAO). *Exp Clin Endocrinol Diabetes*. 1999;107(suppl 5):S168-S171. doi:10.1055/s-0029-1212177

9. Ponto KA, Pitz S, Pfeiffer N, Hommel G, Weber MM, Kahaly GJ. Quality of life and occupational disability in endocrine orbitopathy. *Dtsch Arztebl Int*. 2009;106(17):283-289. doi:10.3238/arztebl.2009.0283

10. Farid M, Roch-Levecq A-C, Levi L, Brody BL, Granet DB, Kikkawa DO. Psychological disturbance in graves ophthalmopathy. *Arch Ophthalmol*. 2005;123(4):491-496. doi:10.1001/archopht.123.4.491

11. Kahaly GJ, Hardt J, Petrak F, Egle UT. Psychosocial factors in subjects with thyroid-associated ophthalmopathy. *Thyroid*. 2002;12(3):237-239. doi:10.1089/105072502753600205

12. Kahaly GJ, Douglas RS, Holt RJ, Sile S, Smith TJ. Teprotumumab for patients with active thyroid eye disease: a pooled data analysis, subgroup analyses, and off-treatment follow-up results from 2 randomised, double-masked, placebo-controlled, multicentre trials. *Lancet Diabetes Endocrinol*. 2021;9(6):360-372. doi:10.1016/S2213-8587(21)00056-5

13. Douglas RS, Kahaly GJ, Patel A, et al. Teprotumumab for the treatment of active thyroid eye disease. *N Engl J Med*. 2020;382(4):341-352. doi:10.1056/NEJMoa1910434

14. Smith TJ, Kahaly GJ, Ezra DG, et al. Teprotumumab for thyroid-associated ophthalmopathy. *N Engl J Med*. 2017;376(18):1748-1761. doi:10.1056/NEJMoa1614949

15. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*. 2016;34(10):993-1004. doi:10.1007/s40273-016-0404-1

16. Gudex C. *Time Trade-off User Manual: Props and Self-Completion Methods*. Centre for Health Economics University of York; 1994.

17. Tolley K. What are health utilities? Accessed June 1, 2021. http://www.bandolier.org.uk/painres/download/What%20is%202009/What_are_health_util.pdf

18. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull*. 2010;96(1):5-21. doi:10.1093/bmb/ldq033

19. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095-1108. doi:10.1097/00005650-199711000-00002

20. Brazier J, Rowen D. NICE decision support unit technical support documents. In: *NICE DSU Technical Support Document 11: Alternatives to EQ-5D for Generating Health State Utility Values*. National Institute for Health and Care Excellence (NICE); 2011.

21. R Core Team. The R project for statistical computing. Accessed June 1, 2021. http://www.R-project.org/

22. Jiang R, Janssen MFB, Pickard AS. US population norms for the EQ-5D-5L and comparison of norms from face-to-face and online samples. *Qual Life Res*. 2021;30(3):803-816. doi:10.1007/s11136-020-02650-y

23. Brazier J, Rowen D, Karimi M, Peasgood T, Tsuchiya A, Ratcliffe J. Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *Eur J Health Econ*. 2018;19(6):881-891. doi:10.1007/s10198-017-0931-5

24. Mladenovic M, Buchberger M, Qerimi V, Puntscher S, Siebert U, Rochau U. Health state utilities in individuals with goiter, hypothyroidism, hyperthyroidism and Graves' disease as an example for thyroid disorders: a systematic review. *Value Health*. 2017;20(9):A483. doi:10.1016/j.jval.2017.08.479

25. Brown GC, Sharma S, Brown MM, Kistler J. Utility values and age-related macular degeneration. *Arch Ophthalmol*. 2000;118(1):47-51. doi:10.1001/archopht.118.1.47

26. Brown MM, Brown GC, Sharma S, Kistler J, Brown H. Utility values associated with blindness in an adult population. *Br J Ophthalmol*. 2001;85(3):327-331. doi:10.1136/bjo.85.3.327

27. Niemeyer KM, Gonzales JA, Doan T, Browne EN, Rao MM, Acharya NR. Time trade-off utility values in noninfectious uveitis. *Am J Ophthalmol*. 2019;208:47-55. doi:10.1016/j.ajo.2019.06.005

28. Wang Y, Sharma A, Padnick-Silver L, et al. Trends in treatment of active, moderate-to-severe thyroid eye disease in the US. *J Endocr Soc*. 2020;4(12):bvaa140. doi:10.1210/jendso/bvaa140

Invited Commentary

Insight Into Quality of Life in Patients With Moderate to Severe Thyroid Eye Disease

Marissa J. Carter, PhD, MA

Quality of life is a key concept in assessing the consequences of having a medical condition. Although physical examination and testing can provide objective measures of the disease, they do not inform us how the patient feels about the condition or the effect on everyday life. To address this issue, both general and disease-centric instruments are becoming increasingly used as patient-reported outcome measures (PROMs) with at least 160 disease-focused instruments regularly used in the fields of ophthalmology and optometry.[1] As useful as PROMS are, they lack a singular value that tells us where a patient lies on a universal scale. In order to do that, the idea of utility has been adopted in which the scale is anchored to values of 1 representing perfect health and 0 death. Determining, therefore, utility values for uncommon eye diseases, which have a prevalence of something like 0.1% in a general population, is a worthy goal. In this issue of *JAMA Ophthalmology*, Smith et al[2] report on their endeavors to measure utility in thyroid eye disease (TED), a condition that involves chronic mild to severe symptoms of diplopia and proptosis.

Related article page 159

The authors considered 6 health states to be assessed with various combinations of the 2 principal symptoms, varying from no diplopia and small proptosis (mild) to constant diplopia and large proptosis (severe) using the time-trade-off (TTO) approach and a cohort drawn from the US general population as the body of individuals tasked with answering questions to obtain the individual utility values. The most striking results are the mean utility value of the 6 states—0.44—and a range of 0.6 to 0.3 in terms of increasing severity of the disease, which equates to very considerable perceived debilitation. For comparison, the utility of the degree of blindness in which individuals can perceive light or count fingers is 0.47 while no light perception is estimated at 0.26.[3] The most important question we can ask, therefore, is how credible are the results?

The population from which participants are drawn is an important factor. The main choices are a general population, health care professionals, especially those familiar with the relevant health field, caregivers, or patients who have the condition. In ophthalmic disease in which visual acuity is affected, it is not uncommon to find that physicians provide higher estimates of utility compared to patients or even general populations as was shown in a systematic review of neovascular age-related macular degeneration.[4] Research has also shown that utility values derived from general vs patient populations can be quite different with several proposed explanations: it can be challenging for respondents in poor health states to imagine full health; poor descriptions of health states that are augmented or interpreted by a respondent's experience; the response shift, in which individuals can change their internal standards for evaluating their own health in response to changes in their health (eg, older persons rate expectations based on their age group); adaptation by patients to the disease over time; and health respondents considering transition to an imagined health state rather than the longer-term consequences with the result that the general population provides lower values compared with patient self-reported values for chronic states of health.[5] That said, many agencies and health care decision-making bodies have endorsed the use of general populations provided measures are taken to prevent the most egregious forms of hidden bias. In this respect, the authors went to considerable lengths to ensure health state descriptions were accurate and useful and participant recruitment mirrored the US population. Results showed that the actual population interviewed differed slightly from the reference population with a skew toward a US TED population, which probably was helpful.

Another highly discussed point in the literature is the method by which utilities are estimated. The TTO method involves asking participants a series of questions designed to elicit their preferences in trading remaining life expectancy to avoid remaining in subperfect health states and is popular in ophthalmology health economics. However, its use does pose several problems. First, it can be cognitively challenging and sometimes requires guidance from interviewers leading to interviewer interference effects.[6] Second, the many different methods that TTO can use does affect the final obtained values (ie, the devil is in the details).[7] Other commonly cited issues are time preferences and the respondents' life expectancies (eg, 40 years vs 5 years), and lack of theoretical measurement properties such as unidimensionality and the invariance principle.[6] But perhaps the biggest objection to TTO is that on a theoretical basis it produce values and not utilities because uncertainty and risk are not involved in the measurement process. These are not fundamental reasons for immediately discontinuing TTO in favor of another method as all methods have their shortcomings, but rather a recognition that we need to develop better methods and recognize that the numbers we produce from some TTO studies may not be utili-