THE REPRODUCIBILITY OF OPHTHALMIC UTILITY VALUES*

BY Gary C. Brown, MD, MBA, Melissa M. Brown, MD, MN, MBA (BY INVITATION), Sanjay Sharma, MD, MSc, MBA (BY INVITATION), George Beauchamp, MD, AND Hussein Hollands, MSc (BY INVITATION)

ABSTRACT

Purpose: Utility values have been used in the ophthalmic literature to measure the quality of life associated with a health state. By convention, a utility value of 1.0 is associated with perfect health, and a value of 0.0 is associated with death. Construct validity of utility values has been demonstrated, particularly in regard to decreasing utility values as the vision decreases in the better seeing eye, but long-term test-retest reliability has not been demonstrated. The purpose of this study was to demonstrate the test-retest reliability of ophthalmic utility values.

Methods: One hundred fifteen patients with ophthalmic diseases and stable visual acuity underwent time trade-off utility analysis with retesting at various intervals ranging from 1 month to 2 years. The results were analyzed using the Wilcoxon signed rank test. The study was designed to have an 80% power, using a two-sided alpha of 5%, to be able to detect a 10% difference between the test and retest groups.

Results: The mean time from testing to retesting was 0.87 years, with a median time of 1.0 year and range of 1 month to 2 years. The mean utility value in the test group was 0.766 (SD = .21; 95% CI, 0.730 - 0.802), while the mean utility value in the retest group was 0.763 (SD = .22; 95% CI, 0.724 - 0.802). The difference between the means of the test-retest groups was not significant (P = .99). The intraclass correlation between the initial and follow-up utility scores was .5246 (P < .00005).

Conclusions: Ophthalmic utility values appear to have good test-retest reliability over prolonged periods of time. This information is important because it gives researchers increased confidence in the validity of basic tools for ophthalmic cost-effective (cost-utility) analyses.

Tr Am Ophth Soc 2001;99:199-204

INTRODUCTION

Utility values are measures that quantify the quality of life associated with a health state.4 By convention, utility values range from 1.0, associated with perfect health, to 0.0, associated with death. The closer the value to 1.0, the better the quality of life associated with a health state, while the closer to 0.0, the poorer the quality of life. For example, mild angina has been associated with a utility value of 0.90, while severe angina has been associated with a value of 0.50.2

Utility value analysis has also been undertaken in patients with ophthalmic disease.3-7 It has been demonstrated that utility values in patients with ophthalmic disease most closely correlate with the visual acuity in the better-seeing eye. For example, a patient with bilateral ocular disease and a visual acuity of 20/20 in the better-seeing eye has been shown to have a utility value of 0.92, while a person with the same underlying ocular disease and acuity of 20/200 in the better-seeing eye has been shown to have a utility value of 0.66.11

Construct validity has been demonstrated for ocular utility values,3 as has test-retest reliability on a short-term (28-day) basis.11 To our knowledge, however, test-retest reliability has not been demonstrated on a longer-term basis. For this reason, we undertook a study to evaluate the test-retest reliability of ocular utility values in a patient population with known ocular disease.

PARTICIPANTS AND METHODS

Over a 2-year period, from December 1998 through November 2000, we collected time trade-off utility data on over 1,000 adult patients with various ocular diseases gathered from vitreoretinal (G.C.B.) and comprehensive ophthalmology (M.M.B.) practices. Data for the present
study were gathered from the group of these patients that was seen in follow-up examination. An attempt was made to consecutively include follow-up data on all patients who had initially undergone utility value evaluation. The utility study was approved by the Institutional Review Board of Wills Eye Hospital, Philadelphia.

Each person underwent a complete opthalmologic examination, including best-corrected Snellen visual acuity, slit-lamp biomicroscopy, and dilated fundus examination. Snellen visual acuity was selected for measurement, since it is the most commonly used system in clinical practice, and an objective of the study was to most closely simulate real-life practice. In those instances in which the visual acuity could be improved with pinhole above the best-corrected visual acuity, the pinhole vision was selected as the best vision. Since people often squint to improve vision, this again was thought to more accurately represent the actual visual potential in a real-life setting. The length of time of visual loss from the initial loss to the initial examination was recorded, as was the time from the initial examination to the follow-up examination.

Exclusion criteria included the absence of any ocular disease, unwillingness to answer the questions posed in the study, and obvious Alzheimer’s disease or some other form of dementia. A change in visual acuity between the initial and follow-up visits of 1 or more lines in the better seeing eye or 2 or more lines in the poorer seeing eyes was also an exclusion criterion.

**UTILITY VALUE ASSESSMENT**

The standardized and validated methodology of utility value assessment has been previously described. In essence, each patient was asked how many additional years of life he or she expected to live. The patient was then presented with the following theoretical scenario: “If the vision in each of your eyes could be returned to (or maintained as) normal (20/20) on a permanent basis by a 0.75, calculated as follows: 1.0 - (5/20) = 0.75.

At follow-up examination, the same time trade-off questions were asked in a double-blind fashion, with neither the patient nor the interviewer having access to previous utility values. The interviewers, however, did compare initial visual acuity with the acuity on follow-up examination at the time of data analysis, since those patients in whom there was a change in visual acuity (either worsening or improvement) between visits were ineligible for the study. Patients were not routinely asked to come in for follow-up for utility evaluation but were examined at their follow-up visits related to their ocular abnormalities.

It should be noted that in previous studies, about 5% to 10% of patients who initially agreed to participate in utility value assessments were unable to answer the questions once they were posed. In the present analysis, however, none of the patients who initially answered the ocular utility value questions were unwilling or unable to do so at the follow-up examination.

**STATISTICAL METHODS**

Means, medians, standard deviations, and 95% confidence intervals were calculated for the initial and follow-up groups using an Excel 97 statistical package (Microsoft, Tacoma, Wash). SPSS 10.1 software (SPSS, Inc, Chicago, Ill) was employed to perform the Kolmogorov Smirnov evaluation for normalcy, the Wilcoxon signed rank test for nonparametric data, the Spearman's rho correlation coefficient, and the intraclass correlation.

A power calculation using Epi-Stat (Centers for Disease Control, Atlanta, Ga) revealed that for a two-sided alpha of 5%, 115 patients in each of the initial and follow-up groups were required for a power of 80% to detect a 10% difference between the groups. Variances were obtained from a previous large utility value study on ophthalmic patients.

**RESULTS**

Repeat utility values were obtained on 175 patients, among whom 125 demonstrated stable visual acuity conforming with study entrance criteria during the follow-up period. Included among the 125 subjects in the study were 45 men and 80 women. The mean age was 65.1 years (SD = 12.1; 95% CI, 62.9 - 67.3). The median age was 68 years, and ages ranged from 37 to 82 years. There were 121 white and 4 black subjects. The mean number of years of education was 13.0 (SD = 2.7; 95% CI, 12.5 - 13.5), with a median of 12 years and a range of 4 to 20 years. A summary of the clinical characteristics of the sample is shown in Table I.

The predominant ocular diseases in the 125 subjects were as follows: diabetic retinopathy (54), macular degeneration (33), cataract (12), retinal detachment (9), retinal arterial or venous obstruction (6), amblyopia (3), glaucoma (2), corneal opacity (2), uveitis (2), macular hole (1), and macular pucker (1). These are shown in Table II.

The mean Snellen visual acuity in the better eye in the sample at initial entrance into the study was 20/40,
The Reproducibility of Ophthalmic Utility Values

with a range of 20/20 to light perception. The median visual acuity in the better seeing eye was also 20/40. The respective mean visual acuity in the poorer seeing eye was 20/100, with a range of 20/25 to no light perception. The median visual acuity in the poorer seeing eye was 20/300. Three subjects had 20/20 vision in each eye.

The mean time of visual loss to the level at the time of entrance into the study was 4.4 years, with a range of 1 month to 63 years and a median time of 2 years of visual loss. When the 3 subjects with visual loss occurring predominantly secondary to amblyopia were excluded, the mean time of visual loss was 3.1 years. As stated in the “Participants and Methods” section, only subjects in whom the vision remained the same in the better seeing eye and changed no more than 1 Snellen gradation better or worse in the poorer seeing eye were included in the study.

The mean utility value for the group at the time of initial examination was 0.766 (SD = .20; 95% CI, 0.724 - 0.802), with a median value of 0.80. Utility values in the follow-up group ranged from 0.17 to 1.0. The mean time of follow-up between the measurements was 0.87 year, with a median time of 1.0 year and a range of 1 month to 2 years.

The one-sample Kolmogorov-Smirnov evaluation for normality of distribution revealed a Z value of 2.196 for the initial utility values (P<.001) and a Z value of 2.214 (P<.001) for the follow-up utility values, indicating that neither of the utility value distributions was normal. Bar graphs of the utility value distributions at initial examination and follow-up examination are shown in Figs 1 and 2,

TABLE I: CLINICAL CHARACTERISTICS OF THE OPHTHALMIC SUBJECT COHORT (n=125)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Male: 45 (36%)</th>
<th>Female: 80 (64%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at initial examination</td>
<td>Mean: 65.6 yr</td>
<td>Median: 68 yr</td>
</tr>
<tr>
<td>Race</td>
<td>White: 121 (97%)</td>
<td>Black: 4 (3%)</td>
</tr>
</tbody>
</table>

| Education (years postkindergarten) | Mean: 13.0 | Median: 12.0 | Range: 4 - 20 |

TABLE II. PREDOMINANT OCULAR DISEASES IN THE COHORT OF 125 SUBJECTS

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>NO.</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macular degeneration</td>
<td>54</td>
<td>(43)</td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>33</td>
<td>(26)</td>
</tr>
<tr>
<td>Cataract</td>
<td>12</td>
<td>(10)</td>
</tr>
<tr>
<td>Retinal detachment</td>
<td>9</td>
<td>(7)</td>
</tr>
<tr>
<td>Retinal arterial or venous obstruction</td>
<td>6</td>
<td>(5)</td>
</tr>
<tr>
<td>Amblyopia</td>
<td>3</td>
<td>(2.5)</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>2</td>
<td>(1.5)</td>
</tr>
<tr>
<td>Corneal opacity</td>
<td>2</td>
<td>(1.5)</td>
</tr>
<tr>
<td>Uveitis</td>
<td>2</td>
<td>(1.5)</td>
</tr>
<tr>
<td>Macular hole</td>
<td>1</td>
<td>(1)</td>
</tr>
<tr>
<td>Macular pucker</td>
<td>1</td>
<td>(1)</td>
</tr>
</tbody>
</table>

FIGURE 1: Distribution of initial utility values (INITUTIL) in the 125-subject cohort.

FIGURE 2: Distribution of retest (follow-up) utility values (FUUTIL) in the 125-subject cohort.
respective. Because of the nonparametric characteristics of the utility values, the Wilcoxon signed rank test, the nonparametric variant of Student's t test, was used to statistically assess utility values and the utility calculation parameters of years to live and time of life traded.

The mean initial utility value for the cohort was .766 (SD = .21; 95% CI, 0.730 - 0.802), while the mean utility value in the retest group was .763 (SD = .22; 95% CI, 0.724 - 0.802). The difference between the means of the test and retest groups was not significant (P = .99).

When the mean amount of time each subject expected to live was examined, it was noted that the average number of remaining years at initial evaluation was 16.6 (SD = 9.8; 95% CI, 14.9 - 18.3), with a range of 5 to 40 years. At follow-up examination, the mean number of years remaining was 16.8 (SD = 10.5; 95% CI, 14.0 - 18.6), with a range of 2 to 45 years. The difference between the means of the initial and follow-up groups was not significant (P = .777).

The mean number of years traded at initial examination was 3.4 (SD = 3.7; 95% CI, 2.7 - 4.1), with a range of 0 to 20 years, while the mean number of years traded at follow-up was 3.9 (SD = 4.3; 95% CI, 3.1 - 4.7), with a range of 0 to 20 years. The difference between the means of the initial and follow-up groups was not significant (P = .474). A summary of these study results is shown in Table III.

An intraclass correlation was also performed between the initial and follow-up utility scores. The correlation was .5246 (95% CI, 0.3850 - 0.6408). This was highly significant (P < .0005).

Correlations were also performed between visual acuity and utility values using Spearman's rho correlation coefficient. The correlation between initial utility value and vision in the better seeing eyes was .524 (P < .001), while that between initial utility values and vision in the poorer seeing eye was .365 (P < .001).

**DISCUSSION**

Utility value analysis is a tool that allows the measure of the quality of life associated with a health state. According to Guyatt and associates,\(^{20,21}\) it is imperative for a good quality-of-life measurement tool to have construct validity, reliability, ease of administration, and the capacity to be readily interpretable. Overall construct validity was not addressed in the present analysis but requires comparing a parameter (in this case, utility values) with other patient characteristics and examining logical relationships that should exist. Construct validity has been previously shown for ophthalmic utility values, particularly in relation to the visual acuity in the better seeing eye.\(^{18}\) Data from the analysis presented here also indicate a substantially higher correlation between utility values and visual acuity in the better seeing eye than between utility values and visual acuity in the poorer seeing eye.

Our data suggest that ophthalmic utility values appear to be highly reproducible, even over prolonged periods. This information is critical to confirm the reliability of ocular utility values, an important aspect of construct validity. It has previously been demonstrated that ophthalmic test-retest utility values are reliable on a short-term (4-week) basis,\(^{19}\) with an intraclass correlation between initial and follow-up utility values of 0.7634. This indicates an excellent reliability for subsequent time trade-off utility responses, according to Rosner.\(^{22}\) In the present study, we found an intraclass correlation of .5246, considered to be good reproducibility.\(^{11}\) This gives us even greater confidence in the time trade-off methodology for ophthalmic utility value assessment, since the mean time between the initial and follow-up utility values in the present study was 0.87 years, with a range as high as 2 years between test and retest data.

The mean time of visual loss in subjects in this series was 3 years. Thus, it is uncertain whether this reproducibility found in subjects with longer-term visual loss holds for cases of acute visual loss (<1 month). It has been demonstrated in both patients with diabetic retinopathy\(^{10}\) and patients with visual loss occurring secondary to multiple causes\(^{11}\) that ophthalmic utility values are similar whether the visual loss has occurred for 1 year or less or for more than 1 year. In subjects with visual loss occurring secondary to age-related macular degeneration, however, the time trade-off utility value has been shown to be significantly lower in those who have had visual loss for 1 year or less than in those who have had visual loss for more than 1 year. Thus, at least for visual loss associated with age-related macular degeneration, there may be a compensatory mechanism that allows for some improvement of the quality of life over time.

In addition to reliability within the same population, it has been shown that ophthalmic utility values are comparable across international borders.\(^{10}\) They also appear to transcend differences in age, sex, race, and level of formal education.\(^{12}\) Other tools, including the VF-14,\(^{24}\) the

### TABLE III: CLINICAL DATA ON THE OPHTHALMIC SUBJECT COHORT AT INITIAL AND FOLLOW-UP EXAMINATIONS (N=125)

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>INITIAL EXAM</th>
<th>FOLLOW-UP EXAM</th>
<th>P VALUE*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean vision in better seeing eye</td>
<td>20/40</td>
<td>20/40</td>
<td>NA</td>
</tr>
<tr>
<td>Mean utility value</td>
<td>0.765</td>
<td>0.761</td>
<td>.991</td>
</tr>
<tr>
<td>Mean years of life expectancy</td>
<td>16.6</td>
<td>16.8</td>
<td>.777</td>
</tr>
<tr>
<td>Mean years traded</td>
<td>3.4</td>
<td>3.8</td>
<td>.474</td>
</tr>
</tbody>
</table>

NA, not applicable.

*Using the Wilcoxon signed rank test.
The Reproducibility of Ophthalmic Utility Values

51-item National Eye Institute Visual Function Questionnaire,23 and the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36).24 have also been used to measure quality of life in ophthalmic patients. These instruments measure primarily the functional ability associated with a health state and can provide valuable information. Utility values, however, are believed to be more all-encompassing quality-of-life measures,25,26 since they also take into account factors such as socioeconomic status, support systems, anxiety, psychologic overlay, and other entities that may not be addressed or emphasized by the previously mentioned instruments. Utility analysis furthermore allows a comparison of widely disparate medical specialties, an endeavor that can be difficult with certain of the other quality-of-life measures.

Perhaps the most important aspect of utility value analysis, compared with other quality-of-life tools, is that it is a critical component of cost-effective (cost-utility) analysis.27-31 With cost-utility analysis, the patient-perceived value conferred by a medical intervention can be amalgamated with the cost and compared across virtually all interventions in health care. Knowing that ophthalmic utility values are reproducible and reliable on a test-retest basis over a prolonged period of time, as found in the present study, gives researchers greater confidence in the validity of cost-effective analyses in the arena of ophthalmic interventions.

REFERENCES


DISCUSSION

Dr Malcolm L. Mazow. I appreciate the authors sending the paper in ample time for me to be able to review it and have a colleague assist with the statistics. There are a few issues that I would like to discuss.

Why did you choose to compute the confidence
Brown et al

interval instead of the predictive interval? By using the Bland Altman to look at the predictive interval, one could look at the individuals rather than groups. This would allow a better look at 1, 2 and 3 standard deviations. Test/retest should be for individuals and not groups as one would like to know how many fall out of the 68% range of 1 standard deviation. I am unclear about the comparison of good eye to bad eye, as visual perception occurs with (requires) use of both eyes.

This study looked at a very homogeneous group. I would like to see a more heterogeneous group tested and stratification of the various categories so that the individuals could be more accurately assessed. Ethnicity, age, occupation and diagnosis seem important. What about gender? Expected years to live and the number of years one is willing to give up is most certainly related to the patients’ age at the time of diagnosis. A youngster of 17 might be much more willing to give up several years at the end of their life, which is way down the road in their mind, than an adult of 70. Also, one’s education, vocation, and avocation, should impact the number of years they would be willing to give up for better vision. Another consideration might be the subjects overall health, and the condition of their other senses.

Perhaps, the practices involved did not lend themselves to such variations, but I think this should be looked at in future investigations.

As we are all aware, the name of the game in medical care is cost analysis. Therefore, it is incumbent on all of us to explore issues that can have an effect on the acceptance of various treatments that affect the quality of life of our patients, as this paper does.

I know you are all as concerned as I am about this issue and only hope that as we continue to prove how important quality of life can be for patients, we can impact the treatment available. In this regard, I’ve looked at the importance of the treatment of amblyopia in children, and now we are embarking on a study of adult strabismus and how improvement of this condition positively affects patients’ lives.

[Editor’s note] Dr Ivan R. Schwab asked how differing sociological values were factored into the equation, pointing out as an example that some individuals were willing to pay a significant amount for refractive surgery while others were happy that their vision was correctable with glasses or contact lenses. Dr Albert W. Biglan asked if and how adverse results from surgery were considered.

Dr Gary C. Brown. In answer to Dr Mazow’s excellent questions, visual utility values tend to be nonparametric at levels of 20/60 or better, but parametric, or normally distributed when the vision drops to 20/70 or worse. In the latter situation, confidence intervals can be employed.

In regard to the quality of life, it turns out that it is the good eye that is the most important and correlates with utility values. When we looked at variants of combining vision in both eyes, the vision in the better seeing eye correlates just as well as using a combination of 75% good eye, 25% bad eye and 50% good eye, 50% bad eye. The poorer seeing eye plays a secondary role; it is the better seeing eye that is more critical.

The vision in the second eye, however, is also valuable because people with 1 good eye constantly worry about their good eye. They worry about what is going to happen to that good eye, and they worry so much that it decreases their quality of life.

Concerning age and utility values, between the ages of 25 and 85 years the values for the same health state appear to be similar in younger and older groups. Utility values also transcend levels of society, since people with an 8th grade education have the same utility values as those with college education for the same health state. The values seem innate to human nature.

The comorbidity question is a good one. What about people with systemic comorbidities? For example, does someone who has diabetes derive less value from cataract surgery than somebody who doesn’t have diabetes? We thought this was a very important question. It turns out that when we do a multiple regression analysis, ocular utility values are not affected by systemic comorbidities. So, somebody who has heart disease or a bad hip receives just as much improvement in quality of life from cataract surgery as someone who otherwise has perfect health. This strongly suggests that we should not discriminate against the disabled when therapeutic interventions are being considered.

Dr Schwab talked about the willingness to pay method of utility value measurement. There are a number of ways to measure utility values and willingness to pay is one methodology. We have avoided it since we believe people with more money are willing to pay more for a therapeutic intervention, thus confounding quality of life issue. That’s why we elect to stay with the time trade-off methodology of utility value measurement.

The last question – Dr Biglin’s question. With decision analysis an integral part of cost utility analysis, one can take into account the adverse effects, or the disutilities, associated with any type of health state. Cost utility analysis incorporates utility values with decision analysis and takes into account improvement in length of life and quality of life, as well as adverse effects; it correlates the value obtained from an intervention with the associated costs. We believe cost-utility analysis will play a dominant role in how health care is delivered within a decade.