Article

Lookup Tables Versus Stacked Rasch Analysis in Comparing Pre- and Postintervention Adult Strabismus-20 Data

David A. Leske[1], Sarah R. Hatt[1], Laura Liebermann[1], and Jonathan M. Holmes[1]

[1] Department of Ophthalmology, Mayo Clinic, Rochester, MN, USA

Correspondence: Jonathan M. Holmes, Ophthalmology W7, Mayo Clinic, Rochester, MN 55905, USA. e-mail: holmes.jonathan@mayo.edu

Purpose: We compare two methods of analysis for Rasch scoring pre- to postintervention data: Rasch lookup table versus de novo stacked Rasch analysis using the Adult Strabismus-20 (AS-20).

Methods: One hundred forty-seven subjects completed the AS-20 questionnaire prior to surgery and 6 weeks postoperatively. Subjects were classified 6 weeks postoperatively as "success," "partial success," or "failure" based on angle and diplopia status. Postoperative change in AS-20 scores was compared for all four AS-20 domains (self-perception, interactions, reading function, and general function) overall and by success status using two methods: (1) applying historical Rasch threshold measures from lookup tables and (2) performing a stacked de novo Rasch analysis. Change was assessed by analyzing effect size, improvement exceeding 95% limits of agreement (LOA), and score distributions.

Results: Effect sizes were similar for all AS-20 domains whether obtained from lookup tables or stacked analysis. Similar proportions exceeded 95% LOAs using lookup tables versus stacked analysis. Improvement in median score was observed for all AS-20 domains using lookup tables and stacked analysis ($P < 0.0001$ for all comparisons).

Conclusions: The Rasch-scored AS-20 is a responsive and valid instrument designed to measure strabismus-specific health-related quality of life. When analyzing pre- to postoperative change in AS-20 scores, Rasch lookup tables and de novo stacked Rasch analysis yield essentially the same results.

Translational Relevance: We describe a practical application of lookup tables, allowing the clinician or researcher to score the Rasch-calibrated AS-20 questionnaire without specialized software.

Introduction

Strabismus (ocular misalignment) is a condition that negatively impacts health-related quality of life (HRQOL).[1–8] The Adult Strabismus-20 questionnaire (AS-20)[8] is a strabismus-specific patient-derived HRQOL instrument (Table 1) that has been shown to be valid and responsive to the treatment of strabismus.[9–11] The AS-20 has been further refined using Rasch analysis in an effort to ensure unidimensionality in each domain and proper response orientation and to convert the original AS-20 score to a linear measure.[12] Identified as a rigorously developed instrument for assessing strabismus-related HRQOL,[13] the resulting Rasch AS-20 questionnaire has four domains: self-perception (five items), interactions (five items), reading function (four items), and general function (four items) (Table 1). Response

options in the general function domain were also reduced from five to four options (never/rarely, sometimes, often, and always).[12]

Two approaches are commonly used to analyze responsiveness data using Rasch analysis: (1) using available Rasch lookup tables[12,14,15] (ready-to-score spreadsheets that automatically calculate Rasch measures from raw responses) to compare pre- and postintervention scores[16] and (2) performing a de novo stacked Rasch analysis.[17,18] Because Rasch analysis is technically demanding and often not readily available to clinicians and researchers, the ability to Rasch-score questionnaire data using lookup tables is a convenient option, but, to our knowledge, the two methods of using Rasch lookup tables and de novo stacked Rasch analysis have not previously been compared. The purpose of the present study was to compare these two methods of analysis

Table 1. AS-20 Questionnaire Items

Self-Perception Domain
*1. I worry about what people will think about my eyes.
2. I feel that people are thinking about my eyes even when they don't say anything.
3. I feel uncomfortable when people are looking at me because of my eyes.
4. I wonder what people are thinking when they are looking at me because of my eyes.
6. I am self-conscious about my eyes.

Interaction Domain
5. People don't give me opportunities because of my eyes.
7. People avoid looking at me because of my eyes.
8. I feel inferior to others because of my eyes.
9. People react differently to me because of my eyes.
10. I find it hard to initiate contact with people I don't know because of my eyes.

Reading Function Domain
12. I avoid reading because of my eyes.
13. I stop doing things because my eyes make it difficult to concentrate.
16. I have problems reading because of my eye condition.
20. I need to take frequent breaks when reading because of my eyes.

General Function Domain
11. I cover or close one eye to see things better.
15. My eyes feel strained.
17. I feel stressed because of my eyes.
18. I worry about my eyes.

Original AS-20 Items Not Scored in the Rasch AS-20
14. I have problems with depth perception.
19. I can't enjoy my hobbies because of my eyes

The original non-Rasch-scored AS-20 had two domains: psychosocial (items 1–10) and function (items 11–20). The Rasch-scored AS-20[12] contains four unidimensional domains: self-perception (items 1–4 and 6), interactions (items 5 and 7–10), reading function (items 12, 13, 16, and 20), and general function (items 11, 15, 17, and 18). Items 14 and 19 are not scored in the Rasch AS-20.

* Original item numbers of the AS-20.

(Rasch lookup tables versus de novo stacked Rasch analysis) using pre- and postoperative AS-20 data in adults with strabismus. We hypothesized that results using each method would be the same.

Methods

Patients

Adult strabismus patients undergoing strabismus surgery between the years 2009 and 2012 by a single strabismus surgeon at the Mayo Clinic were prospectively recruited and completed the AS-20 questionnaire (available at no cost at www.pedig.net, accessed December 15, 2015) immediately prior to surgery and again at their 6-week postoperative examination (defined as a window between 3 weeks and 5 months). Patients could not have participated in the previous study in which Rasch lookup tables were created.[12] At the 6-week examination, surgical outcomes were classified as "success," "partial success," or "failure" based on previously reported postoperative angle and diplopia outcome criteria (Table 2).[19]

Scoring of the AS-20: Lookup Table

Rasch AS-20 logit measures in the present study were estimated using a lookup table based on item and structure calibrations from the previous Rasch analysis of the AS-20 performed in 348 adult strabismus patients[12] (from www.pedig.net). Rasch

Table 2. Criteria Used to Define Surgical Outcome Classification 6 Weeks Following Strabismus Surgery as Previously Described[19]

	Success (All Criteria Must Be Met)	Partial Success (All Criteria Must Be Met)	Failure (If Any One Criterion Met)
Motor criteria			
Angle of deviation by SPCT distance	<10 prism diopters	≤15 prism diopters	>15 prism diopters
Angle of deviation by SPCT near	<10 prism diopters	≤15 prism diopters	>15 prism diopters
Diplopia criteria			
Diplopia/visual confusion primary position distance	None or "rare"	None, "rare," or "sometimes"	"Always" or "often"
Diplopia/visual confusion reading	None or "rare"	None, "rare," or "sometimes"	"Always" or "often"
Prism	Not allowed	Allowed	Not applicable
Bangerter foil/occlusion	Not allowed	Not allowed	Allowed

SPCT, simultaneous prism and cover test.

scoring of the existing AS-20 is based on responses from 18 of the original 20 items (no. 14 and no. 19 were not scored) and combining of the "rarely" and "never" response options for items in the general function domain as previously described.[12]

Scoring of the AS-20: Stacked Rasch Analysis

Rasch AS-20 logit measures were also calculated by performing a de novo Rasch analysis on a stacked dataset (each subject had responses from both a pre- and a postoperative questionnaire in the same dataset). Rasch analysis was performed with Winsteps software (version 3.72.2, Winsteps Software Technologies, Seattle, WA; available at www.winsteps.com, accessed December 15, 2015) using the same methods as previously described.[12]

Data Analysis

Pre- to postoperative change in AS-20 domain scores were compared overall and with respect to surgical success classification (success, partial success, and failure), using three different approaches: (1) effect sizes, (2) proportion exceeding 95% limits of agreement (LOA), and (3) change in the distribution of scores.

Effect Sizes

For the first method of analyzing pre- to postoperative change in HRQOL scores, the effect size statistic was calculated by dividing the magnitude of the pre- to postoperative change by the standard deviation of the preoperative scores.[20] Effect sizes of 0.20 to 0.49 were considered small, 0.50 to 0.79 were

considered medium, and 0.80 and higher were considered large.[21]

Proportion Exceeding 95% LOA

The second method for analyzing pre- to postoperative change in HRQOL scores was calculating the proportion of patients showing postoperative change in HRQOL greater than the 95% LOA (test-retest variability derived from a previous test-retest study[22]), with respect to each domain individually and with respect to any of the four domains: that is, determining if any of the four domains change by a magnitude that exceeds the 95% LOA. Rescoring previous test-retest data[22] using Rasch-based values[12], 95% LOA were 2.99 logits for self-perception, 1.36 logits for interactions, 2.38 logits for reading function, and 1.83 logits for general function domains. Changes observed exceeding these thresholds are indicative of change that is greater than the amount of change expected by variability of the test itself.[23] Because AS-20 domain scores for some patients were high enough that improvement exceeding the 95% LOA was not possible, separate analyses were conducted comparing only subjects able to exceed the 95% LOA.

Change in Distributions

For the third method of analysis, distributions of pre- to postoperative changes in HRQOL score were compared using signed rank tests for each of the four Rasch AS-20 domains. Comparison of postoperative change in Rasch AS-20 domain scores between patient success classifications (success, partial success, and failure) was made using Kruskal-Wallis tests and individual Wilcoxon rank sum tests, with a Bonferro-

Table 3. Demographic and Clinical Characteristics of Original AS-20 Rasch Study and the Present Study

	Original AS-20 Rasch Study[12] (N = 348)	Present Study (N = 147)
Sex: Female	202 (58.0%)	87 (59.2%)
Race		
American Indian	3 (0.9%)	0
Asian	2 (0.6%)	2 (1.4%)
Black/African American	4 (1.2%)	1 (0.7%)
More than one race	3 (0.9%)	0
Unknown/not reported	6 (1.7%)	1 (0.7%)
White	330 (94.8%)	143 (97%)
Ethnicity		
Hispanic	3 (0.9%)	1 (0.7%)
Not Hispanic	335 (96.3%)	145 (98.6%)
Unknown/not reported	10 (2.9%)	1 (0.7%)
Age (years)		
18–30	61 (17.5%)	33 (22.4%)
31–40	46 (13.2%)	15 (10.2%)
41–50	57 (16.4%)	22 (15.0%)
51–60	77 (22.1%)	35 (23.8%)
61–70	53 (15.2%)	27 (18.4%)
71–80	42 (12.1%)	11 (7.5%)
81+	12 (3.5%)	4 (2.7%)
Previous surgery	153 (44.0%)	63 (42.9%)
Diplopic	246 (70.7%)	105 (71.4%)
Nondiplopic	102 (29.3%)	42 (28.6%)
Strabismus etiology		
Childhood/idiopathic	154 (44.3%)	64 (43.5%)
Neurologic	114 (32.8%)	47 (32.0%)
Sensory	21 (6.0%)	14 (9.5%)
Mechanical	59 (17.0%)	22 (15.0%)
Primary deviation		
Esotropia	116 (33.3%)	51 (34.7)
Orthotropia (post-op)	5 (1.4%)	2 (1.4%))
Torsion	7 (2.0%)	9 (6.1%)
Vertical	60 (17%)	27 (18.4)
Exotropia	160 (46.0%)	58 (39.5)

ni-corrected α of 0.0167 indicating significance (accounting for multiple comparisons).

The study was approved by the Institutional Review Board at the Mayo Clinic, and informed consent was obtained from all subjects. Data were collected and analyzed in a manner consistent with the Health Insurance Portability and Accountability Act guidelines and adhered to the Declaration of Helsinki.

Results

One hundred forty-seven adult strabismus patients were enrolled in the study (median age 52 years, range 18–87 years). Eighty-seven (59%) patients were female, 143 (97%) self-reported their race as white, 64 (44%) had childhood onset/idiopathic strabismus, 47 (32%) neurogenic strabismus, 22 (15%) mechanical strabismus, and 14 (10%) sensory strabismus. Ninety-seven (66%) of the patients had diplopia, 8 (5%) had symptoms of visual confusion, and 42 (29%) were nondiplopic. Table 3 reports demographics and clinical characteristics of subjects in both the original AS-20 Rasch study and the present study.[12] Data from 1 (1%) of the 147 have been previously reported in a study of responsiveness of the original AS-20.[10] Questionnaires were completed a median of 1 day prior to surgery (range 0–12 days) and 7 weeks following surgery (range 4 weeks–5 months). At the 6-week examination, 102 (69%) of 147 subjects were classified as a surgical success, 18 (12%) as a partial success, and 27 (18%) as a surgical failure. Although a separate population from the study population used for the initial Rasch analysis of the AS-20[12], Rasch analysis in the present study population led to the same scale and response option structure as previously described.

Effect Sizes

Effect sizes were very similar, both overall and by surgical outcome status, when comparing Rasch lookup table methods to de novo stacked Rasch analysis methods (Table 4). Effect sizes were generally large in patients classified as success using both lookup tables and stacked methods, whereas effect sizes were medium or small for partial success and small for failures using both methods.

Proportion Exceeding 95% LOA

The proportion of subjects exceeding 95% LOAs for each domain, both overall and based on surgical success status, were similar between Rasch lookup table methods and de novo stacked Rasch analysis (Table 5). The proportion of subjects exceeding 95% LOAs when limited to only those subjects able to improve beyond the LOA is also reported in Table 5

Table 4. Effect Sizes of HRQOL Scores by Surgical Success Status Using the Rasch-Scored AS-20 Questionnaire, Analyzed Using Lookup Tables and De Novo Stacked Rasch Analysis

	Lookup Table	Stacked Analysis
Overall		
Self-perception	0.58	0.59
Interactions	0.46	0.51
Reading function	0.66	0.72
General function	0.93	0.88
Success		
Self-perception	0.76	0.78
Interactions	0.63	0.72
Reading function	0.88	0.97
General function	1.26	1.17
Partial success		
Self-perception	0.28	0.30
Interactions	0.23	0.22
Reading function	0.57	0.49
General function	0.66	0.50
Failure		
Self-perception	0.10	0.09
Interactions	0.07	0.02
Reading function	−0.01	0.01
General function	0.20	0.17

and may give a more accurate estimate of improvement.

Change in Distributions

When comparing the distribution of AS-20 scores, median domain score improved across all AS-20 domains whether analyzed with lookup tables or with stacked Rasch methods ($P < 0.0001$ for all comparisons; Fig. 1). When comparing pre- and postoperative HRQOL scores within each surgical outcome classification, improvement was observed for surgical successes for each domain of the Rasch-scored AS-20 whether analyzed using Rasch lookup tables or a stacked Rasch analysis ($P < 0.0001$ for each comparison). The distribution of responses was somewhat greater when using the de novo Rasch analysis method. For partial surgical successes, improvement was much less, reaching statistical significance on the reading function domain with each method ($P < 0.007$) and the general function domain using lookup tables ($P = 0.003$). In contrast, no improvements were observed in patients classified

as failures for any domains by either method ($P \geq 0.2$ for each comparison). Comparing pre- to postoperative changes in scores between outcome categories (success, partial success, and failure), greater change in score was observed among successful outcomes compared with failures for self-perception ($P = 0.0008$), reading function ($P < 0.0001$), and general function ($P = 0.0002$) using the Rasch lookup tables and for all domains using the stacked Rasch analysis method ($P \leq 0.002$ for all comparisons) (Fig. 2). Greater change was observed for successful outcomes than for partially successful outcomes in the self-perception domain using a stacked Rasch analysis ($P = 0.01$). Numerically greater change, albeit nonsignificant when Bonferroni corrected ($P > 0.0167$), was observed for successful outcomes compared with partial success and partial success compared with failures for all remaining domains using either analysis method (Fig. 2).

Discussion

When using either Rasch lookup tables or a de novo stacked Rasch analysis for analyzing pre- to postoperative AS-20 data, we found essentially identical results, and subtle differences between methods did not change the interpretation of the data. Overall, the Rasch-scored AS-20 is responsive to changes 6 weeks following strabismus surgery measured using three different methods: effect size, proportion improving more than the 95% LOAs, and change in distribution of scores. The Rasch-scored AS-20 demonstrates construct validity, with greater change in HRQOL scores following successful strabismus surgery than following surgical failure. Practically, it may be more convenient to use Rasch lookup tables[12] than to perform a de novo stacked Rasch analysis, and it is reassuring that either method yields essentially identical results for AS-20 data. As noted, the distribution of responses for the de novo stacked Rasch analysis was somewhat greater than for the Rasch lookup tables. Nevertheless, the corresponding variability using the lookup tables was less, which is reflected in very similar effect sizes using either method.

We have previously reported a comparison of the original AS-20 to the National Eye Institute Visual Function Questionnaire-25 (VFQ-25) in response to strabismus surgery.[10] In that study, the strabismus-specific AS-20 was found to be more responsive to surgery than the VFQ-25, particularly for nondiplopic patients. The AS-20 has previously undergone Rasch

Table 5. Improvement in HRQOL Exceeding the 95% LOA by Surgical Success Status Using the Rasch-Scored AS-20 Questionnaire, Analyzed Using Lookup Tables and De Novo Stacked Rasch Analysis

	Lookup Table	Stacked Analysis	Lookup Table: Limited to Only Subjects Able to Exceed*
Overall (N = 147)			
Self-perception	43 (29%)	46 (31%)	43 of 87 (49%)
Interactions	44 (30%)	33 (22%)	44 of 77 (57%)
Reading function	57 (39%)	37 (25%)	57 of 103 (55%)
General function	58 (39%)	48 (33%)	58 of 107 (54%)
Any domain	94 (64%)	94 (64%)	94 of 137 (69%)
Success (N = 102)			
Self-perception	39 (38%)	41 (40%)	39 of 63 (62%)
Interactions	34 (33%)	29 (28%)	34 of 50 (68%)
Reading function	49 (48%)	33 (32%)	49 of 69 (71%)
General function	46 (45%)	39 (38%)	46 of 77 (60%)
Any domain	77 (75%)	77 (75%)	77 of 96 (80%)
Partial success (N = 18)			
Self-perception	2 (11%)	3 (17%)	2 of 9 (22%)
Interactions	4 (22%)	1 (6%)	4 of 10 (40%)
Reading function	6 (33%)	2 (11%)	6 of 12 (50%)
General function	8 (44%)	7 (39%)	8 of 11 (73%)
Any fomain	10 (56%)	10 (56%)	10 of 18 (56%)
Failure (N = 27)			
Self-perception	2 (7%)	2 (7%)	2 of 15 (13%)
Interactions	6 (22%)	3 (11%)	6 of 17 (35%)
Reading function	2 (7%)	2 (7%)	2 of 22 (9%)
General function	4 (15%)	2 (7%)	4 of 19 (21%)
Any domain	7 (26%)	7 (26%)	7 of 23 (30%)

* Number of subjects exceeding of those able to exceed. This analysis was not done for the stacked Rasch analysis due to the open-ended nature of the logit scale.

analysis[12] and now comprises four unidimensional domains, whereas the original AS-20 contained two domains. In the present study, we demonstrate responsiveness of each of the four Rasch AS-20 domains to successful strabismus surgery. The four new Rasch-derived domains (self-perception, interactions, reading function, and general function) likely provide even greater specificity than the original strabismus-specific AS-20 and the more generic VFQ-25. We speculate that the four Rasch-derived domains may be particularly useful across the spectrum of strabismus conditions because different types of strabismus may affect specific domains of strabismus-specific HRQOL differentially.

The results of the present study demonstrate the utility of the Rasch-scored AS-20 in cohort studies because we found marked improvement in average scores and larger effect sizes in surgical successes compared with failures. In addition to using the

Rasch-scored AS-20 to measure response to surgery, we suggest that the Rasch-scored AS-20 could be used to study many different modalities of strabismus treatment, for example, treatment with prism.[16]

Although average AS-20 scores are easily interpreted for studies comparing cohorts or a change in a cohort over time, interpreting change in an individual patient remains more challenging. The HRQOL measures are inherently variable, leading to the question of whether an observed change in scores reflects a true change or just test-retest variability. Using the 95% LOA (also known as the repeatability coefficient), as described by Bland and Altman,[23] provides a measure of the variability expected by readministration of a questionnaire in the absence of a change in the underlying condition. Thus, any change that exceeds the 95% LOA is likely to be a true change in the underlying condition rather than a result of the instrument's variability. In the present
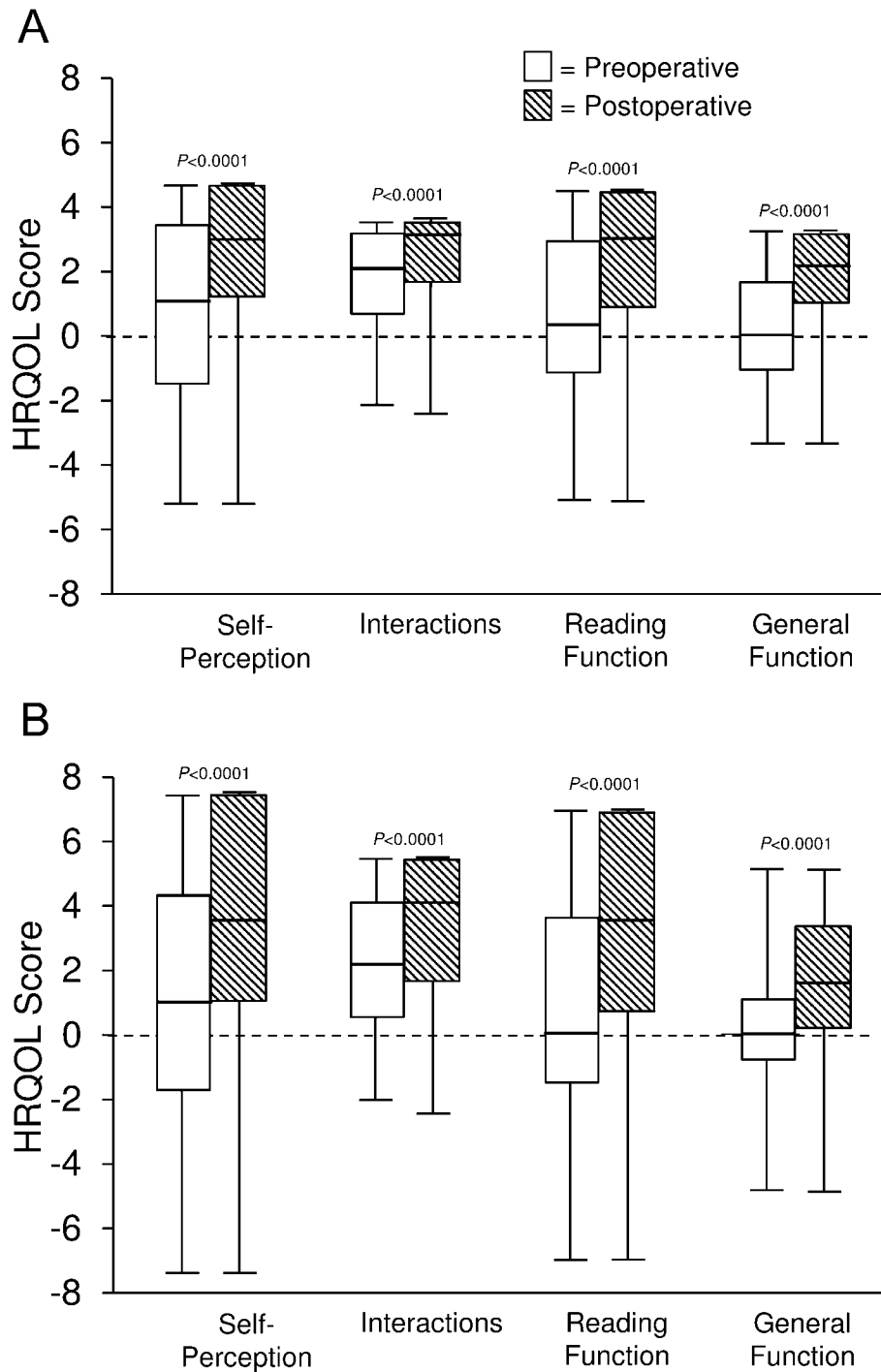
Figure 1. Pre- and postoperative HRQOL scores for the AS-20 domains, regardless of surgical outcome status, calculated using (A) Rasch lookup tables and (B) de novo stacked Rasch analysis. Whiskers represent extreme values. Signed rank *p*-values indicated for change in score.

study, despite significant changes in average scores across the cohort, not all patients showed an improvement that exceeded the 95% LOA for each domain. It is possible that patients may have concerns in only one domain. Nevertheless, one potential disadvantage of assessing improvement in any domain is that there is a somewhat higher probability of exceeding the 95% LOA by chance when assessing whether the 95% LOAs are exceeded on multiple domains versus on one domain. In the present study,
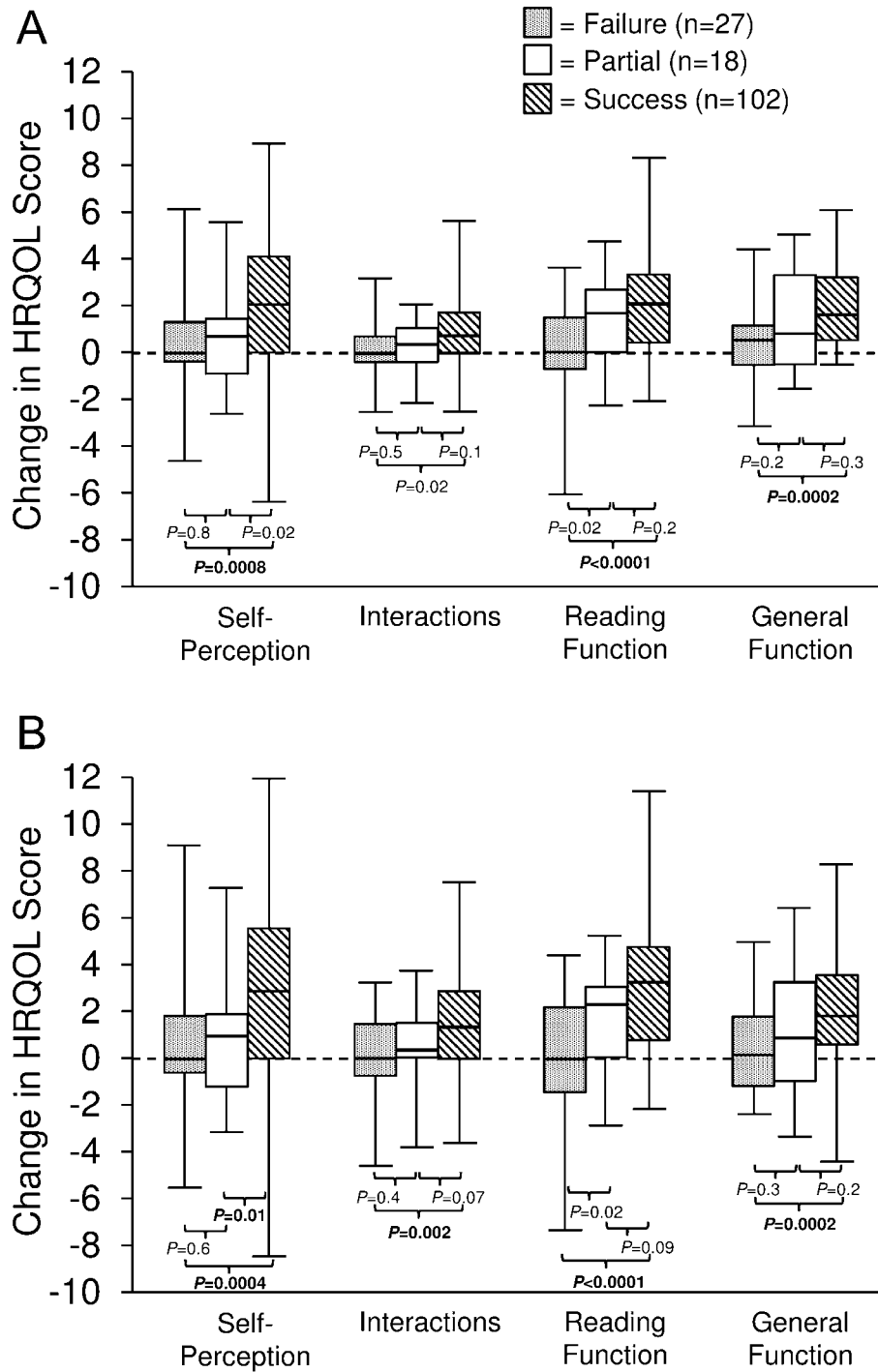
Figure 2. Change in HRQOL by surgical success classification for the AS-20 domains calculated using (A) Rasch lookup tables and (B) de novo stacked Rasch analysis. Wilcoxon rank sum comparisons between surgical success classifications, with *p*-values below 0.0167 (*in bold*) indicating statistical significance (adjusted for multiple comparisons).

26% of failures had improvement exceeding the 95% LOA in at least one of the four domains, although some of these patients may have exceeded the 95% LOA by chance. An alternative explanation for 26% of failures exceeding the 95% LOA on any of the four domains is that our definition of success may have been too strict. Some failures had measurable improvement based on clinical criteria (not meeting success criteria), and some also considered themselves subjectively improved; therefore, their change in

scores might have been expected to exceed the 95% LOAs.

The results of our study suggest that using a Rasch lookup table to convert raw responses to Rasch-calibrated values is a valid method of analysis for AS-20 data. The most evident advantage of using this approach is avoiding the need to conduct a separate Rasch analysis for each study, an analysis that requires specialized analysis software and expertise. Another advantage of using a Rasch lookup table is the ability to easily define whether an individual subject would be able to exceed the 95% LOA for one or more domains using previously described test-retest thresholds.[22] Thus, when determining whether or not a subject's score has changed more than would be expected due to test-retest variability alone, defining "ability to exceed" avoids erroneously concluding that a subject did not change following an intervention, when in reality the subject did not have room enough to change. In addition to logit values, the lookup table conveniently reports scores in a more familiar scale, such as from 0 to 100 (poor to good HRQOL).

Recently, Gothwal et al.[24] translated the AS-20 into Hindi and Telugu, administered the English, Hindi, or Telugu AS-20 (depending on primary language) to a cohort of 584 adult strabismus patients, and performed Rasch analysis on the response data. Wang et al.[25] translated the AS-20 into Chinese and then performed Rasch analysis on responses from a cohort of 247 adult patients with strabismus.[26] In both Rasch studies, the AS-20 was found to be unidimensional when analyzing the two original constructs (psychosocial and function) separately, although there were some slight differences in misfitting items and category response when compared to the Rasch analysis of the English version. Such differences are not unexpected given different cultural backgrounds and clinical characteristics (e.g., 70% of subjects with exotropia in the Chinese study[26]). Different lookup tables for non-English versions of the AS-20, such as those provided by Gothwal et al.[24], may be needed, particularly any time significant differences in performance are highlighted by Rasch analysis in different cultures.

Our study is not without limitations. Applying Rasch lookup tables requires making the assumption that subjects in a given study do not differ dramatically from the subjects used to create the lookup table itself. In the case of the AS-20, Rasch estimates were derived from a previous study of 348 adult strabismus patients, including both pre- and postoperative subjects with a wide spectrum of types and severities of strabismic conditions.[12] Despite efforts to be as representative as possible, it is still possible that the AS-20 may be less responsive to surgery for some types of strabismus or that the targeting may not be optimal for the present cohort, although demographics and clinical characteristics of the present study do not appear to differ greatly from those in the previous study (Table 3). Our study is also somewhat limited by racial and ethnic homogeneity, and further studies evaluating the lookup tables in a more diverse cohort may be needed. Finally, our data can only be generalized to the AS-20, and future studies comparing lookup tables to de novo stacked Rasch analysis for other instruments are warranted.

The Rasch-scored AS-20 is a responsive and valid instrument designed to measure strabismus-specific HRQOL. When analyzing pre- to postoperative change in AS-20 scores, Rasch lookup tables and de novo stacked Rasch analysis yield essentially the same results. Published lookup tables[12] (available free at www.pedig.net) are particularly convenient because no specialized software is needed.

Acknowledgments

Disclosure: **D.A. Leske**, None; **S.R. Hatt**, None; **L. Liebermann**, None; **J.M. Holmes**, None

References

1. Satterfield D, Keltner JL, Morrison TL. Psychosocial aspects of strabismus study. *Arch Ophthalmol*. 1993;111:1100–1105.
2. Burke JP, Leach CM, Davis H. Psychosocial implications of strabismus surgery in adults. *J Pediatr Ophthalmol Strabismus*. 1997;34:159–164.

3. Olitsky SE, Sudesh S, Graziano A, Hamblen J, Brooks SE, Shaha SH. The negative psychosocial impact of strabismus in adults. *J AAPOS*. 1999;3: 209–211.

4. Menon V, Saha J, Tandon R, Mehta M, Khokhar S. Study of the psychosocial aspects of strabismus. *J Pediatr Ophthalmol Strabismus*. 2002;39: 203–208.

5. Beauchamp GR, Black BC, Coats DK, et al. The management of strabismus in adults—III. The effects on disability. *J AAPOS*. 2005;9:455–459.

6. Jackson S, Harrad RA, Morris M, Rumsey N. The psychosocial benefits of corrective surgery for adults with strabismus. *Br J Ophthalmol*. 2006;90:883–888.

7. Hatt SR, Leske DA, Kirgis PA, Bradley EA, Holmes JM. The effects of strabismus on quality of life in adults. *Am J Ophthalmol*. 2007;144:643–647.

8. Hatt SR, Leske DA, Bradley EA, Cole SR, Holmes JM. Development of a quality-of-life questionnaire for adults with strabismus. *Ophthalmology*. 2009;116:139–144.

9. Hatt SR, Leske DA, Bradley EA, Cole SR, Holmes JM. Comparison of quality of life instruments in adults with strabismus. *Am J Ophthalmol* 2009;148:558–562.

10. Hatt SR, Leske DA, Holmes JM. Responsiveness of health-related quality of life questionnaires in adults undergoing strabismus surgery. *Ophthalmology*. 2010;117:2322–2328.

11. Hatt SR, Leske DA, Liebermann L, Holmes JM. Changes in health-related quality of life 1 year following strabismus surgery. *Am J Ophthalmol*. 2012;153:614–619.

12. Leske DA, Hatt SR, Liebermann L, Holmes JM. Evaluation of the Adult Strabismus-20 (AS-20) questionnaire using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2012;53:2630–2639.

13. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90: 720–744.

14. Pesudovs K, Gothwal VK, Wright T, Lamoureux EL. Remediating serious flaws in the National Eye Institute Visual Function questionnaire. *J Cataract Refract Surg*. 2010;36:718–732.

15. Leske DA, Holmes JM, Melia M. Pediatric Eye Disease Investigator Group. Evaluation of the Intermittent Exotropia Questionnaire (IXTQ) using Rasch analysis. *JAMA Ophthalmol*. 2015; 133:461–465.

16. Hatt SR, Leske DA, Liebermann L, Holmes JM. Successful treatment of diplopia with prism improves health-related quality of life. *Am J Ophthalmol*. 2014;157:1209–1213.

17. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care*. 2004;42:I25–I36.

18. Wright BD. Comparisons require stability. *Rasch Meas Trans*. 1996;10:506.

19. Hatt SR, Leske DA, Liebermann L, Holmes JM. Comparing outcome criteria performance in adult strabismus surgery. *Ophthalmology*. 2012;119: 1930–1936.

20. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991;12:142S–158S.

21. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1988:19–74.

22. Leske DA, Hatt SR, Holmes JM. Test-retest reliability of health-related quality of life questionnaires in adults with strabismus. *Am J Ophthalmol*. 2010;149:672–676.

23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.

24. Gothwal VK, Bharani S, Kekunnaya R, et al. Measuring health-related quality of life in strabismus: a modification of the Adult Strabismus-20 (AS-20) questionnaire using Rasch analysis. *PLoS One*. 2015;10:e0127064.

25. Wang ZH, Bian W, Ren H, Frey R, Tang LF, Wang XY. Development and application of the Chinese version of the Adult Strabismus quality of life questionnaire (AS-20): a cross-sectional study. *Health Qual Life Outcomes*. 2013;11:180.

26. Wang Z, Zhou J, Luo X, et al. Rasch analysis of the Adult Strabismus quality of life questionnaire (AS-20) among Chinese adult patients with strabismus. *PLoS One*. 2015;10:e0142188.