

**THE LOCALIZED SCLERODERMA QUALITY OF LIFE
INSTRUMENT (LOSQI): A DISEASE-SPECIFIC SURVEY
USING ANCHORING VIGNETTES**

by

Christina Kelsey Zigler

Bachelor of Science, University of Miami, 2007

Master of Science in Education, University of Miami, 2010

Submitted to the Graduate Faculty of
the School of Education in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Christina Kelsey Zigler

It was defended on

July, 26

and approved by

Committee Co-Chair: Suzanne Lane, Professor, Psychology in Education

Committee Co-Chair: Clement Stone, Professor, Psychology in Education

Kathryn Torok, Assistant Professor, School of Medicine

Lan Yu, Associate Professor, Medicine and Psychiatry

Copyright © by Christina Zigler

2017

**THE LOCALIZED SCLERODERMA QUALITY OF LIFE INSTRUMENT: A
DISEASE-SPECIFIC SURVEY USING ANCHORING VIGNETTES**

Christina Zigler, PhD

University of Pittsburgh, 2017

The main goal of this project was to develop and provide validity evidence for a disease-specific quality of life survey to be used with pediatric localized scleroderma (LS) patients. This new survey, called the Localized Scleroderma Quality of Life Instrument (LoSQI), incorporated unique features associated with the disease, not captured by current surveys. As a secondary goal, the feasibility and usefulness of anchoring vignettes with pediatric patients were examined. The project included three phases; content domain development and item generation, a pilot study, and a field test. Validity evidence was gathered from multiple sources including test content, internal structure, and in relation to other variables. Overall, there was initial support for use of the LoSQI with pediatric LS patients. Patients indicated general understanding and readability of the items, and there was qualitative evidence for content validity. Exploratory factor analysis suggested the utility of reporting a total score along with two subscale scores, (1) Pain and Physical Functioning and (2) Body Image and Social Support. Reliability of both the subscale and total scores was acceptable. There was less evidence for use of anchoring vignettes in this context, as there was a high frequency of ties in rankings, which limited the utility of

statistical models. Despite limitations from a small sample size and skewed response distributions, the pilot study and the field test provided promising initial evidence that the LoSQI can be used to capture HRQoL in LS patients ages 10-20 years. Future studies should examine responsiveness of the scores to change and optimal capture of HRQoL in patients <10 years of age.

TABLE OF CONTENTS

PREFACE.....	XX
1.0 INTRODUCTION.....	1
1.1 STATEMENT OF PROBLEM	1
1.1.1 Health related quality of life	2
1.1.2 Anchoring vignettes	2
1.1.3 Localized scleroderma as target population	3
1.2 PURPOSE	4
1.3 RESEARCH QUESTIONS.....	4
1.4 STUDY SIGNIFICANCE.....	5
2.0 LITERATURE REVIEW	6
2.1 PATIENT REPORTED OUTCOMES.....	6
2.1.1 Current standards for PRO'S.....	7
2.2 LOCALIZED SCLERODERMA: ETIOLOGY AND PROGRESSION.....	9
2.2.1 Health related quality of life in localized scleroderma: What do we know?..	13
2.2.1.1 In children	13
2.2.1.2 In adults.....	14
2.2.1.3 Correlates with HRQoL in LS.....	15

2.2.1.4	HRQoL and physician outcomes: Disagreement in children	15
2.2.1.5	Parental proxy.....	16
2.2.2	Issues and concerns with HRQoL measurement in LS	16
2.2.2.1	Potential HRQoL impact of LS: Further gaps in the literature.....	22
2.2.2.2	Skin sensations	24
2.2.2.3	Physical functioning.....	24
2.2.2.4	Body image	25
2.2.2.5	Peer relationships.....	26
2.2.2.6	Treatment burden.....	27
2.2.3	Summary: LS.....	28
2.3	METHODOLOGICAL CONSIDERATIONS FOR PEDIATRIC SURVEY DESIGN.....	29
2.3.1	Validity as an argument.....	29
2.3.1.1	Validity evidence based on test content	29
2.3.1.2	Validity evidence based on internal structure (Factor Analysis) ...	30
2.3.1.3	Reliability.....	33
2.3.1.4	Validity evidence in relation to other variables	33
2.3.1.5	Validity evidence based on response process	34
2.3.1.6	Validity evidence based on testing consequences and fairness.....	34
2.3.2	Anchoring vignettes	35
2.3.2.1	Design and structure of anchoring vignettes.....	36
2.3.2.2	The C-Scale and B-Scale	38
2.3.2.3	Other statistical models	41

2.3.2.4	Use of anchoring vignettes in the literature	48
2.3.2.5	Considerations for use of anchoring vignettes with children	50
2.3.2.6	Sample size considerations.....	52
2.3.2.7	Summary: Anchoring vignettes.....	52
2.4	SUMMARY OF LITERATURE.....	53
3.0	METHODS.....	55
3.1	DEVELOPMENT OF ORIGINAL SURVEY	55
3.1.1	Proposed use and interpretations	56
3.1.2	Content domain & development.....	56
3.1.2.1	Theoretical Domain 1: Skin sensations.....	57
3.1.2.2	Theoretical Domain 2: Physical functioning and musculoskeletal sequelae.....	58
3.1.2.3	Theoretical Domain 3: Body image and social support	58
3.1.2.4	Theoretical Domain 4: Medication attitudes and important side effects....	59
3.1.2.5	Theoretical factor structure.....	60
3.1.3	Timeframe.....	61
3.1.4	Response scale	61
3.1.5	Age range	62
3.1.6	Scoring of theoretical domains.....	63
3.1.6.1	Total score	63
3.1.6.2	Scoring of optional medication subscale.....	63
3.1.7	Anchoring vignettes	63

3.1.7.1	AV Set 1: Skin sensations.....	65
3.1.7.2	AV Set 2: Physical functioning.....	66
3.1.7.3	AV Set 3: Body image and social support	67
3.1.7.4	AV Set 4: Medication attitudes and side effects.....	67
3.1.7.5	Instructions for anchoring vignettes	68
3.1.8	Order of items and vignettes	68
3.1.9	Background items	69
3.2	PILOT STUDY	70
3.2.1	Purpose.....	70
3.2.2	Sample	70
3.2.3	Methods.....	71
3.2.3.1	Administration	72
3.2.4	Analysis	72
3.2.4.1	Item revision.....	73
3.3	FIELD TEST.....	75
3.3.1	Purpose.....	75
3.3.2	Sample	75
3.3.3	Methods.....	76
3.3.3.1	Administration procedures	76
3.3.3.2	Additional assessment instruments	77
3.3.4	Analysis	79
3.3.4.1	Sample characteristics.....	80
3.3.4.2	Missing item analysis.....	80

3.3.4.3	Reading level	80
3.3.4.4	Item level descriptive statistics	80
3.3.4.5	Internal structure.....	81
3.3.4.6	Descriptive Statistics for subscales and total scores	81
3.3.4.7	Reliability: Stability and Consistency.....	81
3.3.4.8	Evidence in relation to other variables	82
3.3.4.9	Anchoring vignette analysis.....	83
3.3.4.10	Software for analysis	84
3.4	SUMMARY OF METHODS.....	84
4.0	RESULTS	87
4.1	PILOT STUDY	87
4.1.1	Sample and administration	87
4.1.2	Pilot study results: 15-18 years old.....	91
4.1.2.1	LoSQI (15-18 yo).....	91
4.1.2.2	Optional medication subscale (15-18 yo)	95
4.1.2.3	AVs (15-18 yo).....	95
4.1.3	Pilot study results: 11-14 years old.....	96
4.1.3.1	LoSQI (11-14 yo).....	96
4.1.3.2	Optional medication subscale (11-14 yo)	98
4.1.3.3	AVs (11-14 yo).....	98
4.1.4	Pilot study results: 8-10 years old.....	99
4.1.4.1	LoSQI (8-10 yo).....	100
4.1.4.2	Optional Medication Subscale (8-10 yo).....	102

4.1.4.3	AVs (8-10 yo)	103
4.1.5	Anchoring vignettes rank order and assumptions	105
4.1.6	Revisions based on pilot data	107
4.1.6.1	LoSQI revisions.....	108
4.1.6.2	Anchoring vignette revisions	109
4.1.7	Summary	111
4.2	FIELD TEST	112
4.2.1	Field test sample	112
4.2.2	Missing item analysis	113
4.2.3	Reading level.....	114
4.2.4	Item level descriptives.....	114
4.2.5	Exploratory factor analysis (EFA)	118
4.2.5.1	Justification for score reporting	124
4.2.6	Subscale and total score descriptives	125
4.2.6.1	Potential subgroups of interest	129
4.2.7	Reliability	134
4.2.7.1	Stability	134
4.2.7.2	Consistency	135
4.2.7.3	Standard error of measurement.....	136
4.2.7.4	Item-total score correlations	137
4.2.8	Relationships with other variables	140
4.2.9	Anchoring vignette results.....	145
4.2.9.1	Descriptives and missing values over entire sample.....	145

4.2.9.2	Rank ordering	148
4.2.9.3	Vignette-adjusted LoSQI scores	160
5.0	DISCUSSION	162
5.1	DISCUSSION OF PRIMARY RESEARCH QUESTIONS	162
5.1.1	RQ 1: What was the validity evidence to support the definition of the content domain for the LoSQI?	162
5.1.2	RQ 2: What was the evidence to support the reliability of the LoSQI scores in this sample?	164
5.1.3	RQ 3: What is the internal and external validity evidence to support the LoSQI score interpretations?	167
5.1.3.1	Internal validity evidence	167
5.1.3.2	External validity evidence	168
5.1.4	RQ 4: How did the LoSQI perform compared to the Children’s Dermatology Life Quality Index (CDLQI)?	171
5.2	DISCUSSION OF SECONDARY RESEARCH QUESTIONS	172
5.2.1	RQ 5: What validity evidence is there to support using the LoSQI anchoring vignettes with pediatric patients?	172
5.2.2	RQ 6: Did patients rank the LoSQI anchoring vignettes the same way or were there differences?	173
5.2.3	RQ 7: Did adjusting for patients’ response styles using anchoring vignettes change the LoSQI scores significantly? If not, did the anchoring vignette results impart any new information concerning the content domain?. 174	
5.3	LIMITATIONS	175

5.4 FUTURE DIRECTIONS	178
5.5 RECOMMENDATIONS FOR PRACTITIONERS	180
5.6 CONCLUSION	181
APPENDIX A: QUALITATIVE INTERVIEW SCRIPT FOR PILOT STUDY	183
APPENDIX B: PILOT VERSION OF THE LOCALIZED SCLERODERMA QUALITY OF LIFE INSTRUMENT.....	187
APPENDIX C: PILOT VERSION OF ANCHORING VIGNETTES.....	191
APPENDIX D: REVISED VERSION LOSQI.....	196
APPENDIX E: INTER-ITEM CORRELATIONS.....	203
BIBLIOGRAPHY	205

LIST OF TABLES

Table 1. Current classification of LS based on skin involvement.	12
Table 2. Health-related quality of life surveys used with <i>pediatric</i> localized scleroderma patients.	18
Table 3. Health-related quality of life questionnaires used with <i>adult</i> scleroderma patients (both localized and systemic disease).	19
Table 4. Other HRQoL questionnaires that show potential for use in this population.	20
Table 5. Identified domains of health related quality of life that are impacted by pediatric localized scleroderma (LS) and corroborating quotations from LS patients and their parents. ...	23
Table 6. All potential values of the non-parametric C-scale (King & Want, 2007) from one self- assessment, y , adjusted by two anchoring vignette responses, z_1 and z_2	40
Table 7. Male and female names chosen for anchoring vignettes.	65
Table 8. Spiral technique for item coverage for pilot study.	71
Table 9. Common reasons for changing an item during PRO Development from FDA Standards (2009; page 9).	74
Table 10. Sources of data examined during the pilot study and field test.	86
Table 11. Frequency of LS classifications in the sample of patients enrolled in the pilot study..	88
Table 12. Years from diagnosis of localized scleroderma to study visit in the pilot sample.	88
Table 13. Average time (in minutes) for patients to complete the quality of life instruments.	89

Table 14. Domain and total LoSQI scores for 17 patients enrolled in the pilot study.....	90
Table 15. Frequencies for responses to anchoring vignettes in the pilot study.	106
Table 16. Anchoring vignettes used in the pilot study and their revised versions.....	110
Table 17. Rates of disease subtype in the study sample versus a large international sample of 750 children with LS.....	113
Table 18. Item level frequencies for responses on the LoSQI (n = 74).....	115
Table 19. Descriptives for additional question (n = 61).	116
Table 20. Descriptive statistics for Domain 4: Medication side effects (n = 42).	117
Table 21. Descriptives for additional questions related to Domain 4: medication side effects..	118
Table 22. Eigenvalues from the exploratory factor analysis of 21 LoSQI items (n = 74).....	119
Table 23. Factor loadings for 1- and 2-factor solutions with oblimin rotations for 21 items of the LoSQI.....	121
Table 24. Factor loadings from final solution of 20 LoSQI items (n = 74).....	123
Table 25. Fit statistics for the initial 1- and 2-factor solution of the 20-item LoSQI.	124
Table 26. Descriptive information for LoSQI total score and subscale scores (n = 74).....	127
Table 27. Results of independent-samples Mann Whitney U tests comparing score distributions between patients with and without linear LS of the trunk/limb.....	132
Table 28. Results of independent-samples Mann Whitney U tests comparing score distributions between LS patients with and without facial involvement.	133
Table 29. Test-re-test reliability of the LoSQI (n = 30).....	135
Table 30. Internal consistency coefficients for LoSQI domains and total score (n = 74).	135
Table 31. Cronbach’s alpha for full response scales and dichotomized item responses.	136

Table 32. Standard errors of measurement for LoSQI subscale and total scores using test-retest coefficients and Cronbach’s alpha.	137
Table 33. Item-to-total score Spearman’s correlations for the three LoSQI domains and the total LoSQI score.	138
Table 34. Item-to-total score correlations for LoSQI domain 4 (medication side effects).	139
Table 35. Relationships between the LoSQI total score, the CDLQI, and other patient, parent, and physician reported outcomes for patients with complete data ($n = 24$; Spearman’s rho). ...	143
Table 36. Relationships between the LoSQI total score, the CDLQI, and other patient, parent, and physician reported outcomes for all patients with the outcome (Spearman’s rho).	144
Table 37. Correlations between domain scores on the LoSQI and subscale scores on the Peds QL Rheumatology Module ($n = 55$).	145
Table 38. Descriptive statistics for the four vignette sets ($n = 73$).	146
Table 39. Most common ranking patterns for vignettes in set 1 (skin sensations).	149
Table 40. Proportion of cases that a vignette was ranked less than another vignette for set 1 (skin sensations).	151
Table 41. Most common ranking patterns for vignettes in set 2 (physical functioning).	152
Table 42. Proportion of cases that a vignette was ranked less than another vignette for set 2 (physical functioning).	154
Table 43. Most common ranking patterns for vignettes in set 3 (body image and social support)	155
Table 44. Proportion of cases that a vignette was ranked less than another vignette for set 3 (body image and social support).	157
Table 45. Most common ranking patterns for vignettes in set 4 (medication side effects).	158

Table 46. Proportion of cases that a vignette was ranked less than another vignette for set 4 (medication side effects).....	160
Table 47. Comparison between the raw medians and adjusted medians for items with associated anchoring vignettes using the B-Scale (Wand, 2013).....	161

LIST OF FIGURES

Figure 1. Erythematous and waxy plaque lesion in a pediatric localized scleroderma patient. ...	10
Figure 2. Joint contracture and change in pigmentation of right hand corresponding to linear lesion on right arm of a pediatric localized scleroderma patient.	11
Figure 3. Theoretical factor structure of the Localized Scleroderma Quality of Life Instrument (LoSQI).	60
Figure 4. Summary of study procedures.	85
Figure 5. Distribution of domain scores for the LoSQI of 17 patients enrolled in the pilot study.	91
Figure 6. Scree plot for exploratory factor analysis of 21 items on the LoSQI (n = 74).	119
Figure 7. Scatterplots examining the relationship between the LoSQI total score and the subscale scores.	125
Figure 8. Boxplot showing distribution of scores on two LoSQI subscales (n = 74).	127
Figure 9. Distribution of total LoSQI scores (n = 74)	128
Figure 10. Distribution of scores on the medication subscale (n = 41).	128
Figure 11. Scatterplot of relationship between age and LoSQI total score.	129
Figure 12. Scatterplot between disease duration in months and LoSQI scores.	130

Figure 13. Distribution of LoSQI total scores for patients with and without linear involvement of the trunk/limb.....	131
Figure 14. Distribution of the Body Image and Social Support Subscale of the LoSQI for patients with and without facial involvement.....	134
Figure 15. Scatterplot of total LoSQI scores with DLQI/CDLQI scores.	141
Figure 16. Distribution of scores on the DLQI/CDLQI and LoSQI for sample of pediatric LS patients.	142
Figure 17. Ranking patterns for vignettes in set 1 (skin sensations).	150
Figure 18. Ranking patterns for vignettes in set 2 (physical functioning).....	153
Figure 19. Ranking patterns for vignettes in set 3 (body image and social support).....	156
Figure 20. Ranking patterns for vignettes in set 4 (medication side effects).....	159

PREFACE

I would like to acknowledge my committee members and their invaluable support through the entire doctoral process. Thank you Dr. Suzanne Lane for your detailed mentorship and guidance, Dr. Clement Stone for your critical thinking about this project, Dr. Kathryn Torok for your vast clinical and research experience with this population, and Dr. Lan Yu for your strong measurement expertise.

I would also like to thank Kaila Schollaert-Fitch for enrollment, data entry, and overall study coordination, Dr. Heidi Jacobe and Elaine Kunzler for enrollment at their site, the late Dr. Kevin Kim and Dr. Feifei Ye for being my research advisers during my time in the program, and Dr. Kaveh Ardalan for his expertise regarding the focus group data and thoughtful suggestions for my discussion section. Cassie, a special thank you for being my mentor and friend, and allowing me access to your brain and your patients. I also would like to acknowledge the support of the pediatric localized scleroderma patients and their families, whose experiences drive this work.

Thank you to my family: Mom, Dad, Eddie, Tricia, Matt, and Margaret. I love you all.

This work is dedicated to Mike and Theresa: everything I do is for you.

1.0 INTRODUCTION

1.1 STATEMENT OF PROBLEM

Historically, localized scleroderma (LS) was understudied due to its rarity (Laxer & Zulian, 2006), and the demand for research was likely not as strong as for other more severe rheumatic conditions due its non-fatal nature. Only in the past 10-15 years were there enough specialized clinicians to gather a sufficiently large sample of patients for accurate research, and thus, the recent understanding of LS has changed dramatically. For example, it was originally believed LS patients did not experience pain (Uziel, Laxer, Krafchik, Yeung, & Feldman, 2000), but more recent research shows uncomfortable skin sensations are one of the most common symptoms of LS (Das, Berstein, & Jacobe, 2014). Even though knowledge of this condition has grown considerably, especially in pediatric patients, large gaps continue to exist in the literature in regards to the impact of LS on quality of life. Attempts at research into this area have been limited by the lack of valid patient-reported outcomes. The main goal of this project was to develop and provide validity evidence for a disease-specific quality of life survey to be used with pediatric LS patients, incorporating a promising measurement technique, anchoring vignettes. Anchoring vignettes (AVs) are used to increase measurement precision by modeling interpersonal differences in ranking Likert scales. A search for the use of AV's in the literature showed no studies with children or rare chronic diseases.

1.1.1 Health related quality of life

Health related quality of life (HRQoL) is defined as the perceived impact of health or disease on the physical, emotional, and social aspects of a person's life (Calver, Blazeby, Altman, Revicki, Moher, & Brundage, 2013). There is a current trend towards integrating HRQoL into outcomes for clinical trials, complementary to physician-reported data (Chang & Reeve, 2005). HRQoL is typically measured through survey items and can be general or disease-specific. General HRQoL surveys are advantageous due to their ability to compare target patients with similar diseases/disorders or even healthy populations, while disease-specific HRQoL surveys capture the unique challenges associated with a disorder. Disease-specific HRQoL surveys are also useful in tracking changes within a patient over time.

1.1.2 Anchoring vignettes

HRQoL measurement involves collecting self-reported answers on a number of targeted items in order to determine how individuals and groups are impacted by their health. Although this approach is common, problems arise when individuals interpret the 'same' item differently (King, Murray, Salomon, & Tandon, 2004). This can be a major issue; resulting in incomparable data or incorrect conclusions, and is especially likely to occur when subjective response scales are used. Interpersonal incomparability due to response scale definitions is of particular concern when researchers seek to compare results among very dissimilar groups or attempt to capture abstract concepts (Chevalier & Fielding, 2011). AVs are an appealing solution to this problem as they are easy to develop, inexpensive to administer, and should not lengthen the survey significantly (King et al., 2004). Statistical methods exist to assist researchers in adjusting

responses based upon vignette rankings; however, they have not been studied in children or in rare disease populations.

1.1.3 Localized scleroderma as target population

The target population for this research was pediatric localized scleroderma patients. Current standardized LS outcomes rely on physician-report of disease activity and damage (Arkachaisri, Vilaiyuk, Li, et al., 2009; Arkachaisri, Vilaiyuk, Torok, & Medsger, 2010), and there are no developed surveys to specifically capture self-report of pediatric LS patients and no studies examining the psychometric qualities of currently used surveys. Physicians performing LS research tend to implement existing HRQoL surveys developed for other conditions; like the Children's Dermatology Life Quality Index (CDLQI), which was designed for children with more common dermatologic conditions like psoriasis (Salek, Jung, Brincat-Ruffini, MacFarlane, Lewis-Jones, Basra, Finlay, 2013), or the Peds QL Rheumatology Module, which was designed for rheumatic conditions in general (Varni, Seid, Knight, Burwinkle, Brown & Szer, 2002). Despite this gap in the literature, LS can negatively impact HRQoL in many unique ways including uncomfortable skin sensations (Das et al., 2014; (Kroft, de Jong, & Evers, 2008), musculoskeletal effects (Saxton-Daniels & Jacobe, 2010), changes in appearance which could result in negative body image (Palmero, Uziel, Laxer, Forrest, & Pope, 201), potential impact on social support and peer relationships (Ardalan, Switzer, Zigler, Hershey, & Torok, in submission), as well as harsh side effects from systemic medications (Li, Torok, Pope, Dedeoglu, Hong, et al., 2012). The clinical features of LS are also distinct from other comparable diseases, which suggest generic HRQoL surveys might not capture the entire content domain and currently used surveys might underestimate HRQoL in LS patients (Ardalan et al., in submission).

Advances in clinical care and further research in LS, including much needed clinical trials, will continue to be incomplete until a disease-specific HRQoL survey is developed.

1.2 PURPOSE

The purpose of this study was to develop and provide validity evidence for a disease-specific quality of life survey for pediatric LS patients. This new survey, called the Localized Scleroderma Life Quality Instrument (LoSQI), incorporates unique features associated with the disease not currently captured by other surveys. As a secondary goal, feasibility and usefulness of anchoring vignettes with pediatric patients were examined.

1.3 RESEARCH QUESTIONS

The primary research questions this study addressed were as follows:

1. What is the validity evidence to support the definition of the content domain for the LoSQI?
2. What is the evidence to support the reliability of the LoSQI scores in this sample?
3. What is the internal and external validity evidence to support the LoSQI score interpretations?
4. How did the LoSQI perform compared to the Children's Dermatology Life Quality Index (CDLQI)?

The secondary research questions were:

5. What validity evidence is there to support using the LoSQI anchoring vignettes with pediatric patients?
6. Did patients rank the LoSQI anchoring vignettes the same way or were there differences?
7. Did adjusting for patients' response styles using anchoring vignettes change the LoSQI scores significantly? If not, did the anchoring vignette results impart any new information concerning the content domain?

1.4 STUDY SIGNIFICANCE

There are no HRQoL surveys that have published validity evidence for LS patients, and the disease's unique features make generic surveys less accurate. A new survey that provides valid score interpretations would allow researchers to incorporate HRQoL as an outcome measure in future LS research, including clinical trials. This is particularly salient as the current treatments are evaluated; the medications used for treatment of LS are highly toxic and can have serious side effects (Li et al., 2012). Also, if anchoring vignettes (AVs) are found to be feasible and useful in this population, they are likely to be efficacious in measuring response shift (i.e. adjustment to a disease and its affects over time; Schwartz, 2010), comparing outcomes for diverse groups (King, Murray, Salomon, & Tandon, 2004), and adjusting for comparisons between parent and child report.

2.0 LITERATURE REVIEW

2.1 PATIENT REPORTED OUTCOMES

Patient reported outcomes (PRO's) are defined as a “report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else” (FDA, 2009). PRO's are significant as they capture information that can only be obtained from patients. For example, clinicians cannot objectively record unobservable symptoms, nor can they detect a patient’s feelings and emotions or problems not happening in clinic visits like sleep disturbances (Deshpande, Rajan, Sudeepthi, & Abdul Nazir, 2011). Furthermore, only patients can impart information concerning the severity of their symptoms, the impact of their disease, and their feelings towards the disease or treatment (Deshpande et al., 2011). Obtaining this information directly from patients, as opposed to through the lens of physicians, is most appropriate, as physicians tend to underrate physical and emotional effects of disease in their patients (Nelson, Conger, Douglass, Gephart, Kirk, et al., 1983).

Uses for PRO's are varied and include clinical trials, diagnosis, monitoring progress, and even performance assessment or quality improvement (Nelson, Eftimovska, Lind, Hager, & Wasson, 2015). From a patient advocacy standpoint, PRO's are an important bridge between physicians and their patients and can help to improve outcomes by increasing understanding of

disease, treatment, and impacts (Nelson et al., 2015). Currently, momentum is building to include patient reported outcomes (PRO) alongside physician measures in research and clinical trials among all medical specialties (Chang & Reeve, 2005). The cancer-research field was at the forefront of this movement, as combining PRO's with clinical outcomes allowed the development of new standards of care and the reconsideration of some treatments (Bottomley, Flechtner, Efficace, Vanvoorden, Coens, Therasse, et al., 2005).

PRO's can measure a number of concepts, and health-related quality of life (HRQoL) is becoming one of the most widely used. HRQoL is defined as "how a disease and its treatment affect patients' well-being" (Calver, Blazeby, Altman, Revicki, Moher, & Brundage, 2013) and can be either general to a patient's overall life experience or specific to one disease and its ramifications. General HRQoL surveys are typically used for making between group comparisons with other similar diseases or 'healthy' populations, while disease-specific surveys are usually more sensitive to changes within a patient over time (Patrick & Deyo, 1989). Regardless if the construct is general or disease-specific, HRQoL is usually hypothesized as multidimensional, consisting of a large general factor (i.e. 'quality of life') and several smaller factors like mental health, physical functioning, and cognitive symptoms (Chen, Sousa, and West, 2006). This multidimensionality is one of the main challenges to capturing PRO's (Gibbons, Immekus & Bock, 2007).

2.1.1 Current standards for PRO'S

A multitude of guidelines and standards exist to help researchers develop PRO's. Different approaches to the development of these measurement tools depend on the organization's broader goals. For example, the Patient Reported Outcomes Measurement Information System

(PROMIS), funded by the National Institute of Health, created a set of standards to guide the development of PRO's and the evaluation of general item banks and instruments (PROMIS, 2013). Item response theory (IRT) is fundamental to their development standards as their goal is to make general quality of life surveys useful for physicians and thus, short-forms and computer adaptive tests are a main focus. In contrast, the Food and Drug Administration's (FDA) guidance is limited to PRO's used in clinical trials to make 'treatment claims'. Their suggestions are similar to the standards provided by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement and Evaluation (NCME) and include information regarding the content domain, generating items, and obtaining validity and reliability evidence.

The International Society for Quality of Life Research (ISOQOL) recommends survey developers take into account traditional concepts like reliability, validity, responsiveness, interpretability of scores, and burden on the patient (Ahmed, Berzon, Revicki, Lenderking, Moinpour, et al., 2012). They also offer advice for researchers and clinicians who want to incorporate PRO's into their clinical practice (ISOQOL.org, retrieved 2016). Their standards are based on the traditional definition of validity.

AERA, APA, and NCME established standards for designing tests to measure achievement or psychological constructs (2014, which will be henceforth referred to as *The Standards*). *The Standards* are based on an all-encompassing view of validity (AERA, APA, NCME, 2014). Under this definition, validity is directly related to the proposed interpretation or use of the instrument and not a property of the test itself (Cronbach, 1971). *The Standards* state building a 'validity argument' requires information from multiple sources, both supporting and

undermining the specified uses/interpretations of the instrument (Kane, 2001). The validity-argument approach is the main framework for the methodological design of this project.

2.2 LOCALIZED SCLERODERMA: ETIOLOGY AND PROGRESSION

Localized scleroderma (LS), otherwise known as morphea, is an autoimmune condition resulting in lesions on the skin and in the subcutaneous tissue due to a lymphocytic infiltrate and excess production of collagen (Kreuter, 2012). The underlying cause of LS is unknown, although there is evidence of elevated pro-inflammatory Th-associated cytokine profiles in these patients compared with healthy controls (Torok, Kurzinski, Kelsey, Yabes, Magee, et al., 2015). Although LS shares some pathogenic features with systemic sclerosis (SSc), it seems to be a distinct condition, as having both or developing SSc after a diagnosis of LS is rare (Laxer & Zulian, 2006; Fett, 2012).

LS occurs most frequently in self-identified White patients (Fett & Werth, 2011) and females (Murray & Laxer, 2002), and when presenting in childhood, typical onset is around school age (Zulian, Artheya, Laxer, Nelson, et al., 2006). The overall incidence of LS is rare with 0.4-2.7 of 10,000 people (Murray & Laxer, 2002), which, in the past, limited doctors' ability to gather adequate samples for research. Typically, a delay in diagnosis is frequently seen (Zulian et al., 2006), as it is common to mistake the initial lesion symptoms for a bruise or injury. Confusing parents and patients even more, roughly 16% of patients self-report they developed their morphea lesion at the site of a trauma (Grabell, Hsieh, Andrew, Martires, Kim, et al., 2014).

LS has two main disease phases; an initial 'active' phase, where lesions are erythematous and expanding (see Figure 1), and a 'damage' phase which is characterized by fibrosis, sclerosis,

and atrophy in the skin, connective tissues, and even the underlying fat and bone (see Figure 2; Torok et al., 2015). The skin at the lesion site can also be discolored and if on a limb, cause joint contractures like those seen in the hand of the patient depicted in Figure 2. Disease damage can be especially distressing on a cosmetic location of the body (Palmer, Uziel, Laxer, Forrest, & Pope, 2010). The length of the active disease phase can vary, but children diagnosed with LS continue to have problems well into adulthood (Saxton-Daniels & Jacobe, 2010). Total disease duration also varies; in one sample, almost half of patients had no improvement or deteriorating symptoms after a year of follow-up (Herrick et al., 2011). Even more concerning, recurrence rates after a period of ‘remission’ are high for both pediatric- and adult-onset patients (Mertens, Seyger, Kievit, Hoppenreijs, Jansen, et al., 2014), indicating LS typically continues to be a concern throughout a patient’s life.



Figure 1. Erythematous and waxy plaque lesion in a pediatric localized scleroderma patient.



Figure 2. Joint contracture and change in pigmentation of right hand corresponding to linear lesion on right arm of a pediatric localized scleroderma patient.

Subtypes of LS are usually characterized by extent of skin involvement (Table 1, Laxer & Zulian, 2006). Extracutaneous manifestations (ECMs) are not uncommon among all subtypes, with between 15%-39% of patients indicating problems beyond their skin (Herrick, Ennis, Bhushan, Silman, & Baildam, 2011; Zulian, Vallongo, Woo, Russo, Ruperto, et al., 2005; Christen-Zaech, Hakim, Afsar, & Paller, 2008). The most common ECM is arthralgia (joint pain) but patients can also experience fatigue, Raynaud's phenomenon, headaches, seizures, and gastrointestinal symptoms (Christen-Zaech et al., 2008). ECMs tend to be more severe in the linear subtype with high frequencies of joint contractures (Falanga, Medsger, Morris Reichlin, & Rodnan, 1987) and possible neurologic manifestations (Leitenberger et al., 2009).

Table 1. Current classification of LS based on skin involvement.

Main group	Subtype	Description (Laxer & Zulian, 2006)
(1) Circumscribed morphea	(a) Superficial	“Oval or round circumscribed areas of induration limited to epidermis and dermis, often with altered pigmentation and violaceous, erythematous halo (lilac ring), they can be single or multiple.”
	(b) Deep	“Oval or round circumscribed deep induration of the skin involving subcutaneous tissue extending to fascia and may involve underlying muscle. The lesions can be single or multiple. Sometimes the primary site of involvement is in the subcutaneous tissue without involvement of the skin.”
(2) Linear scleroderma	(a) Trunk/limbs	“Linear induration involving dermis, subcutaneous tissue and, sometimes, muscle and underlying bone and affecting the limbs and trunk.”
	(b) Head	“ <i>En coup de sabre</i> (ECDS). Linear induration that affects the face and the scalp and sometimes involves muscle and underlying bone.”
(3) Generalized morphea		“Induration of the skin starting as individual plaques (four or more and larger than 3cm), that become confluent and involve at least two out of seven anatomic sites (head-neck, right upper extremity, left upper extremity, right lower extremity, left lower extremity, anterior trunk, posterior trunk)”
(4) Pansclerotic morphea		“Circumferential involvement of limbs affecting the skin, subcutaneous tissue, muscle and bone. The lesion may also involve other areas of the body without internal organ involvement.”
(5) Mixed morphea		“Combination of two or more of the previous subtypes. The order of the concomitant subtypes, specifies in brackets, will follow their predominant representation in the individual patient [i.e. mixed morphea (linear-circumscribed)]”

As the course and duration of the disease can be variable, treatment is ultimately aimed at decreasing the amount of cutaneous damage incurred by LS lesions. Although attempts have been made at consensus treatment plans, differences in treatment recommendations still exist between rheumatologists and dermatologists (Li, Torok, Pope, Dedeoglu, Hong, et al., 2012). For rheumatologists, systemic medications like methotrexate (MTX) and corticosteroids are the most commonly used, while dermatologists tend to prescribe topical creams and phototherapy (Li et

al., 2012). Both systemic medications (Weibel, Sampaio, Visentin, Howell, Woo & Harper, 2006) and ultra-violet light have been shown to impart clinical improvement (Kreuter, Hyun, Stucker, Sommer, Altmeyer, & Gambichler, 2006). Surgery is also an option for patients with facial involvement. In one study, 9 of 10 patients who were treated with facial surgery would recommend it to other patients (Palmero, Uziel, Laxer, Forrest, & Pope, 201).

2.2.1 Health related quality of life in localized scleroderma: What do we know?

2.2.1.1 In children

HRQoL in pediatric LS has only become a focus of research relatively recently, and the results from most currently published studies are mixed. An older study stated LS patients had normal scores on self-perception and appearance (Uziel, Laxer, Krafchik, Yeung, & Feldman, 2000). However, the same study also stated, “patients with morphea feel no pain”, which is now known as grossly incorrect (Das et al., 2014). A more recent article from 2009 found pediatric LS patients had only slight impairment on a dermatology-specific measure of HRQoL compared with healthy controls (Orezechowski, Davis, Mason, Crowson, & Reed, 2009). In another more comprehensive study, 28% of children with LS were found to have moderate or higher HRQoL impact, as per interpretation of scores, and 78% of patients (adults and children) interviewed had at least one negative symptom (Das, Bernstein, & Jacobe, 2014). An additional study found half of the parents and children interviewed believed their scleroderma had serious consequences on their everyday life (Ennis, Herrick, Baildam, & Richards, 2012). Baildam and colleagues’ 2011 study found, on average, LS had a moderate effect on HRQoL and physical functioning, although a smaller group reported much greater impairment than the sample average (Baildam, Ennis, Foster, Shaw, Chieng, et al., 2011). It is important to note, most of the studies listed have

major methodological limitations like small sample sizes (Orezechowski et al., 2009; Baildam et al., 2011), conflating the results from both LS and systemic sclerosis (Ennis et al., 2012; Baildam et al., 2011), and uncertainty in whether patients or their parents actually responded to the survey (Ennis et al., 2012).

2.2.1.2 In adults

More research exists on HRQoL in adult LS patients; although it is unclear if results can be generalized to children since, on average, adults with LS self-report more HRQoL impact (Das et al., 2014). A study of adults in 2008 found 62% of patients experience pain, fatigue, or itch (Kroft, de Jong, Evers, 2008) and a study done by the same group found 38% of adult LS patients could be considered ‘at risk’ psychologically (2009). In another study, 47% of adult patients experienced moderate to strong negative HRQoL impact from their LS (Klimas, Shedd, Bernstein, & Jacobe, 2014), which is higher than rates in children.

Although results regarding the general levels of HRQoL impact may or may not be generalizable to children, research with adults suffering from LS point to the long-term implications of this disease. One small study found moderate to large HRQoL impact for adults whose disease began in childhood (Saxton-Daniels & Jacobe, 2010). In another study, 30% of adults with pediatric onset had functional impairment, even though they were less likely to admit HRQoL involvement than patients with adult onset (Condie, Grabell, & Jacobe, 2014). The later result might mean patients with long disease trajectories have time to adjust to their ‘new normal’, despite continuing to experience negative effects.

2.2.1.3 Correlates with HRQoL in LS

Studies examining correlates with HRQoL in LS have found variables associated with both risk of and protection from poor outcomes. Skin sensations (i.e. itch and pain), functional impairment (i.e. limited range of motion), extracutaneous manifestations (ECMs), and identifying as female were all found to negatively impact HRQoL (Condie et al., 2014; Saxon-Daniels et al., 2010; Klimas et al., 2014, and Ardalan, Kelsey, & Torok, 2016) – while optimism and time from initial clinic visit were found to positively impact HRQoL (Szramka-Pawlak, 2013; Ardalan et al., *in submission*). Not surprisingly, skin sensations like itch and pain correlate with decreased HRQoL in both children and adults (Das et al., 2014). Itch in particular was also associated with disease activity (meaning the patient exhibited erythematous and expanding lesions) and poor HRQoL (Klimas et al., 2014). Regarding functional limitations, Saxton-Daniels and colleagues found the number of lesions on a patient’s body and limited range of motion in lesion-covered joints were associated with more HRQoL impact in adults (2010). Patients also report more pain if their lesions were on their trunk or limb, as opposed to their face (Baildam et al., 2011).

2.2.1.4 HRQoL and physician outcomes: Disagreement in children

In children, self-reported HRQoL does not correlate with physician-reported outcomes that capture cutaneous involvement, although there seems to be more agreement between adults and their doctors. Arkachaisri and colleagues (2010) found no correlation between the modified-Localized Scleroderma Skin Severity Index (mLoSSI), a LS disease activity measure, and pediatric HRQoL, a result that was replicated in another study (Kelsey & Torok, 2013). However, in adults, disease activity was associated with general HRQoL in two separate studies (using the SF-36; Klimas et al., 2014; Szramka-Pawlak et al., 2013). This disagreement between physicians and pediatric patients might be unique to LS, as moderate agreement exists between

child-patients and physicians in other rheumatic diseases, like lupus (Moorthy, Peterson, Baratelli, Harrison, Onel, Chalom, Haines, Haskhes, & Lehman, 2007). It was hypothesized that the low correlation between disease activity and HRQoL in children is due to a stronger relationship between HRQoL and disease *damage* (Kelsey & Torok, 2013). However, poor correlations have also been seen between HRQoL and cutaneous disease damage measures (Arkachaisri, Vilaiyuk, Torok, & Medsger, 2010).

2.2.1.5 Parental proxy

Parents also tend to report more HRQoL impact for their children than the children's self-report. In one study, parents reported higher total HRQoL scores and higher subscale scores than their children, and this difference increased with the child's age (Baildam et al., 2011). The same study also found emotional impact on parents, family activity, and family cohesion were all rated as more serious for LS than systemic sclerosis (SSc), which was surprising as SSc is considered more severe clinically (Baildam et al., 2011).

2.2.2 Issues and concerns with HRQoL measurement in LS

The conflicting results in the aforementioned research could be related to the lack of standardized HRQoL outcomes, which leaves the decision of which survey to use up to the individual researcher. Unsurprisingly, there is a lot of variation in that choice. Since 2000, at least nine different HRQoL questionnaires were used with children and adult LS patients (Tables 2 & 3), and some researchers even chose to create original surveys (Worret & Jessberger, 2004; Palmero, Uziel, Laxer, Forrest, & Pope, 2010). The Dermatology Life Quality Index (DLQI) and corresponding children's version, the Children's Dermatology Life Quality Index (CDLQI), are

by far the most popular with clinicians treating LS patients (see Table 2). The DLQI/CDLQI were designed to capture the unique concerns of common dermatological conditions like psoriasis (Salek, Jung, Brincat-Ruffini, MacFarlane, Lewis-Jones, Basra, & Finlay, 2013), which share some of the same cutaneous features as LS.

Table 2. Health-related quality of life surveys used with *pediatric* localized scleroderma patients.

Survey	General or Disease-Specific	Respondent	Age Range	No. domains	No. items	Domains	Author(s)	Score Range	Relevant Articles (* includes LS patients)
CDLQI	Derm	Self	3-16 *adult form available	6	10	Symptoms/ feelings, Leisure, School/holidays, Personal relationships, Sleep, Treatment	Lewis-Jones, Finlay	0-30 high = bad	Ardalan, 2010* Condie, 2014* Grabell, 2014* Das, 2014* Kelsey, 2014* Baildam, 2011* Saxton-Daniels, 2010* Orezechowski, 2009* Arkachaisri, 2010*
CHAQ	Gen – physical function (JIA)	Parent proxy	1-19	3	30	Disability, discomfort, pain	Athreya, Fries, Goldsmith	0-3 high = bad	Ennis, 2012* Baildam, 2011*
CHQ	Gen	Self *parent proxy available	10-18	11	87	Physical functioning, bodily pain, role/social-physical, general health perceptions, role/social-emotional behavior, mental health, general behavior, self-esteem, parental emotional impact, parental time impact, family impact.	HealthACT CHQ	0-100 high = good	Baildam, 2011*
CQOL	Gen	Self *parent proxy available	9-15	15	15	Activities, appearance, communication, continence, depression, discomfort, eating, family, friends, mobility, school, sight, self-care, sleep, worry	Graham, 1997	1-105 high = bad?	Baildam, 2011*
KINDL	Gen	Self *parent proxy available	4-16	6	24	Physical well-being, emotional well-being, self-esteem, family, friends school	Ravens-Sieberer & Bullinger 1998	0-100 high = good	Orezechowski, 2009*

CDLQI = Children's Dermatology Life Quality Index; CHAQ = Children's Health Assessment Questionnaire, CHQ = Childhood Health Questionnaire; CQOL = Child Quality of Life Questionnaire; KINDL = German generic quality of life tool for children.

Table 3. Health-related quality of life questionnaires used with *adult* scleroderma patients (both localized and systemic disease).

Survey	General or Disease-Specific	Respondent	Age Range	No. domains	No. items	Domains	Author(s)	Score Range	Relevant Articles (* includes LS patients)
SF-36*	Gen	self	>18	10	36	Physical functioning, Role limitations due to physical functioning, Bodily pain, General health perceptions, Vitality, Social functioning, Role-limitations due to emotional problems, Mental health	RAND	0-100 high = good	Condie, 2014* NG, 2012 Valentini, 2007
Skindex-29	Derm		> 18		29	Emotions, symptoms, functioning	Chren, MM 1997	0-100 high = bad	Condie, 2014* Szramka-Pawlak, 2013*
ISDL	Derm	Self	>16	5	42	Physical functioning, psychological functioning, stressors, illness cognitions, and social support.	Evers, 2007	For each domain	Kroft, 2009* Kroft, 2008*

SF-36 = Short Form Health Survey, ISDL = Impact of Chronic Skin Disease on Daily Life.

Other frequently used surveys with adult and pediatric scleroderma patients were created to capture general HRQoL impact and encompass domains like physical, emotional, and psychological functioning (see Tables 2 & 3). Three surveys designed for general rheumatic diseases have potential for use in LS patients (see Table 4).

Table 4. Other HRQoL questionnaires that show potential for use in this population.

Survey	General or Disease-Specific	Respondent	Age Range	No. domains	No. items	Domains	Author(s)	Score Range	Articles (* LS)
Peds QL	Gen	Self	5-18 *parent proxy available	4	23	Physical, emotional, social, school-functioning	Varni, 1998	0-100 high = good	Mulligan, 2015
Peds QL Rheum	Rheum Gen	Self	8-18	6	22	Pain, hurt, daily activities, treatment, worry, communication	Varni, 1998	0 -100 high = good	Mulligan, 2015
Peds QL-Family Impact	Family Impact	Parent	0-18	6	36	Physical, emotional, social, cognitive functioning, communication, worry	Varni, 1998	0-100 high = good	Varni, 2004

Most saliently, no published studies examining the psychometric properties and validity evidence of any HRQoL questionnaire exist with LS patients in either children or adults. To remedy this need, there is currently a paper being submitted for publication examining the psychometric properties of the CDLQI in pediatric LS patients (Ardalan, Switzer, Zigler, Hershey, & Torok, in submission). Based upon those initial results and the validity standards of *The Standards* (AERA, APA, & NCME, 2014), there is insufficient information to support the CDLQI use in pediatric LS patients at this time.

The CDLQI is a 10-item survey published by Lewis-Jones and Finlay in 1995. It is likely attractive to rheumatologists due to its short length and relevance to the cutaneous features of LS. There is also extensive psychometric research into its properties when used with skin diseases like atopic dermatitis (Salek et al., 2013). However, as designed, the interpretations of its score were never meant for LS patients and a validity argument has yet to be made for the new use. In a sample of 94 pediatric LS patients, the CDLQI had moderate reliability (Cronbach's alpha = .73) and the results indicated a 3-factor solution, eliminating 4 items, fit the data best (Ardalan et

al., in submission). Additionally, only two items saliently loaded on each factor, which is not ideal. These initial results indicate while the CDLQI likely addresses some important domains of LS HRQoL, it does not optimally cover the entire content domain.

As for the initial exploratory factor analysis (EFA) results, the first factor was comprised of the two items dealing with sports and playing (Ardalan et al., in submission), both of which assess functional limitations in daily activities. The second factor consisted of items related to embarrassment and teasing; which both suggest important emotional and social domains. The third factor was comprised of a skin sensation item, which assessed pain, itch, and other skin-related discomfort, and a treatment item, which assessed treatment burden and side effects. The six included items indicate important parts of the content domain that will be expanded upon in the LoSQI.

To provide examples of how the aforementioned items will be further improved, it is important to discuss in supplementary detail two items captured by the CDLQI related to skin sensations and embarrassment. The CLDQI only includes one item regarding skin symptoms and reads, *“how itchy, “scratchy”, sore or painful has your skin been?”* Although this item is attempting to capture the extent of problematic sensations experienced by patients, it is not ideal as it assesses upon more than one negative sensation but only allows for one answer. Thus, the new survey will expand this item into multiple items each capturing different sensations common to these patients, allowing the ranking of each symptom separately. For the second example, more than half of patients answered the CDLQI item *“how upset or embarrassed, self-conscious, or sad have you been because of your skin?”* positively, indicating negative impact (Ardalan et al., in submission). This item again attempts to capture multiple issues relating to general

negative emotions (i.e. ‘upset’/’sad’), self-perceptions (‘embarrassment’), and body image (‘self-conscious’).

2.2.2.1 Potential HRQoL impact of LS: Further gaps in the literature

The total impact of LS on HRQoL is not fully understood due to the lack of comprehensive and valid PRO’s. In 2014, focus groups were conducted with pediatric LS patients and their families to gain a comprehensive view of their experiences. Although the results of this study are not yet published, the transcripts are coded and preliminary themes emerged echoing the results of the factor analysis of the CDLQI. The topics listed in Table 5 encompass the gaps in the literature, and are discussed in more detail in the subsequent sections (2.2.2.2-2.2.2.6).

Table 5. Identified domains of health related quality of life that are impacted by pediatric localized scleroderma (LS) and corroborating quotations from LS patients and their parents.

Domain	Examples from focus group transcripts
Uncomfortable skin sensations like itch, pain, and tightness.	“I keep [my skin] soft because if I don't I can feel it pull and it kind of hurts.”
	“Mine doesn't usually hurt, but it sometimes itches.”
	Parent: “[Redacted] doesn't have pain, but he does feel it.”
Nuanced effects on fine motor functioning after prolonged repeated movements	“I've been taking standardized tests lately and my hand will cramp...I'm given 25 minutes to write an essay and...[I spend] 5 minutes just trying to wrestle my hand back to normal.”
	Parent: “She had to wear braces on her fingers...she had to use putty. She had all these little mechanisms of things she had to squeeze.”
Limitations or worry about elective activities	“They ask...‘You've been dancing for so long why can't you do these basic things?’ And [I] tell them that I'm physically unable to.”
	“I love playing soccer...when I first noticed it, I would always be careful. I would always be watching everywhere just to make sure nothing would hit me.”
	Parent: “She wanted to play volleyball, but I said, you'll hurt your fingers. You know, you hit with the palm of your hand [where your scleroderma is]”
Joint/muscle pain or cramping	“When I'm doing school work, [my hand] will just cramp up. Like what [redacted] said, you know, it's painful.”
	“I play the viola in orchestra...and sometimes I have to position my hand a different way so my hand doesn't cramp up.”
	Parent: “At one time her hand was hurting so bad they thought she was going to have to go to using a computer in school”
Body image/appearance	“You wake up and you just look in the mirror and you're just like what's wrong? Why is that here?”
	“The staring makes me feel different and I used to wear a bunch of long-sleeved shirts and never show my arms.”
	Parent: “He wished his face would change and knowing he had to go to school that day and be seen by his friends and the girls in his classes.”
Peer relationship	“The bad days are when people are commenting and...you just don't want to leave your room. You just want to stay inside because you are afraid of what

Table 5 (continued).	people will say.”
	Parent: “They laugh at her because [of] her skin condition and even her cousins are really mean. They laugh at her. They make fun of her and its terrible.”
Systemic medication side effects	“In the middle of the week I'd get sick and start throwing up and... then I'd just throw up and then go to school and try to fight through it.”
	“With the methotrexate it would just feel weird, it'd make me throw up, I'd feel dizzy.”
	Parent: “She went from being a skinny little kid...to having big puffy cheeks and gaining weight from the steroids. She hated it...and I think that affected me horribly.”

2.2.2.2 Skin sensations

Recent research has shed new light on skin symptoms LS patients experience, including itch, pain, and tightness (Das et al., 2014; Kroft et al, 2008). The majority of the children who participated in the focus groups mentioned some kind of negative, upsetting, or bothersome skin sensation (see Table 5). Itch seemed particularly distressing and often occurred at nighttime, interfering with sleep.

2.2.2.3 Physical functioning

The focus group participants reported nuanced effects on fine motor function that become worse after prolonged repeated movements, e.g. writing in class. They also mentioned significant limitations in elective activities, such as being unable to perform handstands in dance class (see Table 5). Participants also revealed their limitations caused them worry and distress, and it was another way they felt ‘different’ from their peers. Musculoskeletal ECMs like joint contractures, myalgias, and arthralgias are common among LS patients (Christen-Zach et al., 2008) and can affect both their functioning and their emotional responses to situations that might expose their limitations.

The Childhood Health Assessment Questionnaire is traditionally used to measure physical functioning in rheumatic disease (CHAQ; Klepper, 2003). The CHAQ is a 30-item survey including three main domains; disability, discomfort, and pain (Singh, Arthreya, Fries, & Goldsmith, 1994). It was developed specifically for use with juvenile rheumatoid arthritis (RA) patients in 1994, when the availability of treatments resulted in much poorer outcomes than those seen today. The CHAQ was mainly designed to measure gross functioning (for example, the ability to dress independently or walk with/without the assistance of wheelchairs or walkers). Since treatments have improved, researchers have found significant floor effects in the total CHAQ score when used in rheumatic patients (Bode, Klein-Gitelman, Miller, Lechman, & Pachman, 2003; Pouchot, Ruperto, Lemelle, Sommelet, Grouteau, et al., 2001). The problems and concerns of LS patients would not be captured well by the CHAQ, or even by scales created to measure fine motor functioning (like those for Parkinson's disease or other tremor disorders).

2.2.2.4 Body image

In the focus groups, LS patients revealed the relationship between disease and appearance is full of nuances. For some, their feelings towards their skin changed as they grew, while others reported negative feelings towards their skin or mental or visual avoidance. New situations were reported as the most distressing for patients, especially when they were encountering people who did not already know about their condition (see Table 5) and probing questions regarding their skin lesions were asked.

There is little currently published research on body image in LS patients. One study enrolling patients with facial involvement found bullying, low self-confidence, and insecurity in appearance were reported reasons for seeking cosmetic surgery (Palmero et al., 2010). However, in another study, having lesions in a cosmetically sensitive place did not correlate with HRQoL

(using the CDLQI; Das et al., 2014). Studying how LS patients are affected by their appearance and the corresponding psychological and emotional affects is of great importance to the outcomes of this patient population, especially since living with a visible disfigurement can affect body image and self-confidence into adulthood (Rumsey & Harcourt, 2004).

Surveys concerning body image and appearance are used commonly in the burn literature as part of HRQoL (Laitakari, Kolionene, Pyorala, Rintala, et al., 2015; Kyannli, Finlay, Edgar, Wu, & Wood, 2011; Murphy, Holzer, Richardson, Epperson, Ojeda, et al., 2015). LS lesions can sometimes appear to be severe burns, so it follows LS patients might deal with some of the same body image issues as burn patients. Furthermore, LS patients can develop ECMs that can also affect their appearance, including limb circumference discrepancies, hemifacial atrophy, and dental problems. Studies regarding long term effects of children who experience severe burns have mixed results, which are similar to those found in the LS literature. For example, one study found HRQoL in pediatric burn patients was relatively similar to health controls (Laitakari et al., 2015), while another study found burn survivors reported lower HRQoL (Maskell et al., 2013).

2.2.2.5 Peer relationships

A surprising finding in the focus groups was LS patients reported their close friends provided support in stressful situations. Specifically, it was mentioned friends help mitigate the stress many LS patients have with new interactions or when people ask probing questions regarding their lesions. This interesting interaction between protective peer relationships, potentially negative communication with new people, and individual body image may help explain why some patients are more at risk for low HRQoL than others. There is no detailed literature on peer relationships in LS patients, although items regarding social effects are usually included as part of general HRQoL surveys. The CLDQI, as indicated above, does include one item on

embarrassment/teasing, and another that asks how their skin has ‘affected their friendships’, although it is not clear how respondents interpret this item. Participants in the focus groups did mention bullying and teasing as being especially distressing (see Table 5), although not all children experienced these negative interactions.

2.2.2.6 Treatment burden

LS patients are also affected psychologically, physically, and potentially emotionally by the medications they are prescribed. As mentioned above, systemic medications are standard care (Li et al., 2012) and are reported as mostly tolerated in LS patients (Zulian, Vallongo, Patrizi, Belloni-Fortina et al., 2012). As part of their treatment, many patients self-administer MTX, or have their parents administer it, subcutaneously on a weekly basis at home (Li et al, 2012). MTX in particular can have very severe side effects (Li et al., 2012), although studies in LS are infrequent. Anticipatory nausea occurs in roughly 20% of pediatric cancer patients who take the drug (Vol, Flank, Lavoratore, Nathan, Taylor, Zelnka, Maloney, & Lee, 2016), and more than half of arthritis patients (Mulligan, Kassoumeri, Etheridge, Moncrieffe, Wedderburn, & Newman, 2013). MTX gastrointestinal side effects can also result in conditioned behavioral responses that have additional negative impact (van der Meer, Wulffraat, Prakken, Gijssbers, Rademaker, & Sinnema, 2007). MTX is also often prescribed in combination with low dose steroids to increase efficacy of treatment (Kroft, Creemers, van der Hoogen, Boezeman, de Jong, 2009), which can have a number of negative side effects including weight gain and aggression (Foster, van Sonderren, Lee, Sanderman, Dijkstra, Postma, & van der Molen, 2006). Patients and their parents who participated in the focus groups discussed medication side effects at length and mentioned that the effects of prednisone and methotrexate were particularly distressing (Table 5).

The literature on side effects in arthritis patients could be generalizable to LS patients, as they are usually prescribed similar doses of the medication. MTX intolerance rates, meaning the medication was stopped due to side effects, are typically high in these patients (Bulatovic, Heijstek, Verkaaik, van Dijkhuizen, et al., 2011). One study found pediatric arthritis patients who were prescribed MTX had poor HRQoL (Muligan, Wedderburn, & Newman, 2015). Specifically, patients receiving subcutaneous administration indicated they believed MTX to be less helpful and had more difficulty taking it than patients using oral methods of administration (Mulligan et al., 2015). Anxiety regarding the subcutaneous administration and/or feeling sick or nauseous after taking MTX was also significantly related to poor HRQoL (Mulligan et al., 2013). There is validity evidence for a survey designed to quantify MTX intolerance in pediatric idiopathic arthritic patients (The Methotrexate Intolerance Severity Score; MISS, Mulligan et al., 2013), however no validity evidence exists for its use with LS patients. Examining the effect of treatment burden on HRQoL in LS patients is an important consideration and should not be ignored.

2.2.3 Summary: LS

The unique aspects of LS suggest a disease-specific HRQoL survey is necessary to capture a complete picture of the impact on pediatric patients. In summary, LS patients suffer from uncomfortable skin sensations, subtle musculoskeletal effects, changes in appearance which could result in negative body image, impact on social support and peer relationships, as well as harsh side effects from systemic medications. None of the currently used HRQoL surveys or any existing surveys capture all of these domains, further justifying the creation of a new LS-specific HRQoL survey.

2.3 METHODOLOGICAL CONSIDERATIONS FOR PEDIATRIC SURVEY DESIGN

2.3.1 Validity as an argument

Validity is the most important aspect of any test or survey, and building a validity argument drives the design of a study, data collection, and final conclusions. The ‘validity as an argument’ approach is multifaceted, comprehensive, and tied explicitly to the proposed interpretations and uses of an instrument (AERA, APA, & NCME, 2014). A systematic validity argument for a new survey requires consideration of multiple sources of validity evidence including; test content, internal structure, traditional reliability, relationships to other variables, response processes, and an evaluation of fairness (AERA, APA, & NCME, 2014). When designing a pediatric survey, the sources of evidence should be tailored to the cognitive level of the child.

2.3.1.1 Validity evidence based on test content

Test content is extremely important when evaluating the validity of test score interpretations and uses (AERA, APA, NCME, 2014), whether it is designed for adults or children. In fact, all four of the PRO standards included explicit considerations of test content (AERA, APA, NCME, 2014; FDA, 2009; PROMIS, 2013; Ahmed, Berzon, Revicki, Lenderking, Moinpour, et al., 2012). The content framework should be conceived using multiple sources of information including the published literature, the target population, and both content and measurement experts (PROMIS, 2013). The consideration of additional issues related to test content are imperative when designing PRO’s for children such as age-related vocabulary, language comprehension, and duration of recall (FDA, 2009). The FDA standards recommend pediatric

surveys with narrow age groupings to account for development, and they discourage reliance on parental proxy-report (FDA, 2009). During development, qualitative methods, including interviews with respondents, are advised to determine suitable wording and age appropriate language (PROMIS, 2013). Negatively worded items should be avoided because children tend to interpret them differently than positively worded items (Borgers, Hox, & Sikkel, 2004).

When examining psychological traits like HRQoL, survey directions must indicate an appropriate time frame of recall to respondents (FDA, 2009). Different time frames, i.e. ‘in the past week’ versus ‘in the past month’, can result in different answers (PROMIS, 2013). When determining the appropriate time frame to reference, the FDA’s standards recommend researchers take into account “the instrument’s purpose and intended use, the variability, duration, frequency, and intensity of the concept measures; the disease or condition’s characteristics; and the tested treatment” (FDA, 2009). In children, it is even more important to choose a short enough time frame to provide accurate recall (FDA, 2009). The administration process is also an important consideration across all standards (PROMIS, 2013; AERA, APA, NCME, 2014; FDA, 2009; ISOQOL, 2016) in order to limit bias. The time it takes to complete the survey should not be unduly burdensome on respondents (PROMIS, 2013), which might mean limiting the number of included items.

2.3.1.2 Validity evidence based on internal structure (Factor Analysis)

For test interpretations to be valid, the relationships among test items and conceptual components must conform to the construct on which the proposed test score interpretations are based (AERA, APA, NCME, 2014). Theoretical models originating in the field of education are now being applied to clinical settings and multidimensional constructs like HRQoL, (Gibbons, Clark, Cavanaguh, & Davis, 1985) and are appropriate when one overall strong factor and several

weaker factors/domains are present. There are two ways of conceptualizing this structure, (1) the second-order factor model and (2) the bi-factor model, which are described in the context of HRQoL below. Both models are beneficial when trying to assess for dimensionality and determine factor structure of a new questionnaire.

In a second-order model, there is a higher-order factor that represents HRQOL and several lower-order factors (like cognition, physical functioning, or mental health) that directly predict the survey items. This model is appropriate when the lower-order factors are correlated and the higher-order factor is thought to account for those correlations (Chen et al., 2006). The model can be written in an SEM framework and includes a structural component:

$$\eta = \Gamma\xi + \zeta \quad [1]$$

where η is a vector of the lower-ordered factors, Γ is a matrix of factor loadings of the lower-ordered factors on the higher-ordered factor, ξ vector represents the higher-order factor, and ζ represents the disturbances of the lower-ordered factors (Chen et al., 2006). The second-order model also includes a measurement component:

$$Y = \Lambda_y\eta + \epsilon \quad [2]$$

where Y is a vector of observed variables, Λ_y is the factor loadings of the measured variables on the first ordered factors, η is a vector of the lower-ordered factors, and ϵ is the residuals. This model is most commonly used in psychological research areas like personality and depression (Gibbons, Rush, & Immekus, 2009; Carroll, 1997).

The bi-factor model is also known as a ‘general-specific’ or ‘nested’ model. It is similar to the second-order factor model, except now the general factor accounts for the commonality of the items instead of the domains (which were called ‘lower-order factors’ in the previously

described model; Chen et al., 2006). The bi-factor model is very useful when a researcher wants to determine the effect of the domains beyond the general factor, or has reason to believe that each domain offers unique variance to its own set of items (Chen et al., 2006). It can be represented by the following equation:

$$Y = \Lambda_y \eta + \epsilon \quad [3]$$

where Y is a vector of observed variables, matrix Λ_y represents the factor loadings of the general and domain specific factors, vector η represents the factors themselves, and ϵ represents the residual variances for each variable (Chen et al., 2006). As mentioned before, the bi-factor model is commonly used in education research to represent a primary ability dimension and secondary constructs (Watkins & Beaujean, 2013; Gustafsson & Balke, 1993) and more recently has been adopted by clinicians who are performing HRQoL research (Chen et al, 2006).

If multidimensionality is present, researchers have suggested that both the bi-factor and second-order models be directly compared using a chi-squared test to determine best fit (Chen, Sousa & West, 2006). Direct comparison between the two models is made possible because the bi-factor and the second-order models are nested (Yung, Thissen, & McLeod, 1999; Chen et al., 2006). Both models are similar in structure as they include one general factor, and the lower-order factors in the second-order model correspond directly to the domains in the bi-factor model. Although there were initial arguments to the contrary, the models are not mathematically equivalent unless proportionality constraints are imposed using the Schmid-Leiman transformation method (Schmid & Leiman, 1957; Yung, Thissen, & McLeod, 1999). The two models also have similar interpretations although there are important differences. Specifically, the bi-factor model provides an estimate of the general factor after *controlling for the other domains* (DeMars, 2013; Gustafsson & Balke, 1993). This can be particularly helpful when using

testlets or items that are grouped by scenario, as either example would result in non-meaningful domain factors (DeMars, 2013; Rijmen, 2010). However, if the domain scores are of interest, using the bi-factor model can be confusing as the trait scores no longer include the effects of the general factor (DeMars, 2013).

2.3.1.3 Reliability

Reliability of scores must be considered when designing surveys and tests. The traditional definition of reliability comes from classical test theory and encompasses both ‘stability’ and ‘consistency’ of the scores. The term is also commonly defined in relation to measurement error (i.e. a reliable score minimizes measurement error; ISOQOL, 2016). Both test-retest correlation coefficients and Cronbach’s alpha, a quantitative measure of internal consistency, are recommended to quantify reliability in most standards (FDA, 2009; PROMIS 2013; ISOQOL, 2016). *The Standards* offers a more precise approach and suggests reliability be examined separately for each relevant subgroup (AERA, APA, NCME, 2014). Basic internal consistency should also be examined using item-to-total score correlations, adjusted by removing the item in question from the total score calculations.

2.3.1.4 Validity evidence in relation to other variables

The expected relationship between a new test score and other variables is tied intrinsically to the proposed interpretation of the scores (AERA, APA, NCME, 2014). *A priori* hypotheses set by the researcher determine the choice of the variables, and relationships are usually detected using correlations (FDA, 2009), which can either be positive or negative, convergent or divergent. Scores on HRQoL should positively correlate with other subdomains affecting quality of life (e.g. pain, psychosocial symptoms, and disease severity) while not correlating with unrelated

domains (e.g. cognitive ability). Whenever possible, a newly developed score should be tested against a “gold standard” measure of the same construct (FDA, 2009; PROMIS, 2013), although this is not usually possible when measuring HRQoL. PRO survey results also should be triangulated by examining relationships with other patient scored measures, physician-scored disease activity and damage measures, and clinical indicators of disease (ISOQOL, 2016).

2.3.1.5 Validity evidence based on response process

Likert scales are the most widely used type of response scale to measure psychological constructs (‘Likert Scale’, 2008), like HRQoL. Juvenile respondents prefer Likert scales to visual or numeric analogue scales (van Laerhoven, van der Zaag-Loonen, & Derkx, 2004), and unlike in adults, a 4-option response format seems optimal (Borgers, Hox, & Sikkels, 2004). Since Likert scales are subjective, many potential errors can obscure the true value of scores, including interpersonal bias (‘Likert scale’, 2008). A measurement technique called anchoring vignettes (AVs) can be used to limit this bias (see section 2.3.2 of this document for more information).

2.3.1.6 Validity evidence based on testing consequences and fairness

Typically high stakes are not associated with the interpretation of HRQoL scores, and thus some considerations in *The Standards* will not be applicable to this project. In particular, there will be no raters to train and items will be scored in a standardized way. There also will be relatively low-risk of adverse consequences due to testing; although it is possible children could become distressed while reflecting on their experiences. Although unlikely, the IRB submission incorporated a protocol to ensure children who become distressed are debriefed appropriately. Throughout the examination of the test constructs, subgroup analysis was performed to determine any adverse or unintended issues.

2.3.2 Anchoring vignettes

Although the use of surveys is common in the social and medical sciences, problems arise when individuals interpret the ‘same’ item differently (King, Murray, Salomon, & Tandon, 2004). This can be a major issue; resulting in incomparable data or incorrect conclusions, and is likely to occur when subjective response scales are used (e.g. ‘strongly disagree’, ‘disagree’, ‘neutral’, ‘agree’). Interpersonal incomparability due to response scale definitions is of concern especially when researchers seek to compare results between very dissimilar groups or attempt to capture abstract concepts, like HRQoL (Chevalier & Fielding, 2011). Anchoring vignettes (AVs) are an appealing solution to this problem as they can be tailored directly to the concept being measured, are inexpensive to administer, and should not lengthen the survey significantly (King et al., 2004). There also are a number of statistical methods available to assist researchers in adjusting responses based upon the vignette rankings.

AVs describe hypothetical individuals with different levels of a targeted construct (e.g. depression, HRQoL, job satisfaction, etc). The actual level of the construct is constant for each vignette, and thus any differences in ranking is due to interpersonal or cultural factors (both considered sources of error) which can be statistically adjusted for (King et al., 2004). Vignettes are used primarily in large-scale surveys for inter-cultural or -country comparisons on constructs like health inequities (Dowd, 2011), depressed mood (Mojtabai, 2015), and job satisfaction (Kristensen, 2005), although potential applications for use with smaller samples exist. After taking into account the information provided by the vignettes, study conclusions can change significantly. For example, when conducting a cross-cultural comparison of depression levels, researchers found Americans tended to rate both themselves and vignettes as more depressed than their European counterparts (Mojtabai, 2015). However, when the difference in rankings of

vignettes was taken into account, Americans emerged as being less depressed than most of the European countries (Mojtabai, 2015). AVs can also “discover, clarify, and define additional dimensions of complicated concepts” (page 204, King et al., 2004) and at their basic level, AVs can be used to measure the implicit assumptions that a survey item is unambiguously written and clearly conceptualized (King et al., 2004).

2.3.2.1 Design and structure of anchoring vignettes

An example vignette from the survey of depressed Americans and Europeans is shown below (Mojtabai, 2015):

Part 1:¹ Karen/Paul enjoys her/his work and social activities and is generally satisfied with her/his life. She/he gets depressed every 3 weeks for a day or two and loses interest in what she/he usually enjoys but is able to carry on with her/his day-to-day activities.

Part 2: Overall in the last 30 days, how much of a problem did Karen/Paul have with feeling sad, low, or depressed?

None Mild Moderate Severe Extreme

As shown above, AVs typically have two parts. Part 1 describes the state of the hypothetical individual using examples from the content area. It is usually tailored to the gender of the targeted respondent by changing the name and pronouns used (King et al., 2004), and it is recommended the age of the character in each vignette be clearly stated to avoid age-related response inconsistency (Grol-Prokopczyk, 2014). The second part of the vignette is the stem (Part 2), which asks the respondent to provide their subjective response of the described

¹ The wording ‘Part 1:’ and ‘Part 2:’ was not part of the original vignette text but was added by the author for ease in describing the vignette.

individual's experience (which in the above example is depression). When designing the vignettes, the response scale should be the same as the self-report items (King et al., 2004). Oftentimes the stem is identical to the actual survey item except reworded to indicate the subject should report his or her own situation. For example:

Overall in the last 30 days, how much of a problem did you have with feeling sad, low, or depressed?

None Mild Moderate Severe Extreme

More than one vignette can be used for each content area, and typically enough are included to cover the entire scale range (Chevalier & Fielding, 2011). For the depression example above, five vignettes would cover none, mild, moderate, severe, and extreme depression levels. The optimal number of vignettes depends on the information the investigator is targeting. There is also a trade-off between gaining information and survey-cost due to vignettes increasing survey length and respondent fatigue (King et al., 2004).

Although traditionally provided after the survey (King et al., 2004), some researchers recommend the AVs precede the survey items to promote consistent definitions of the response scale (Hopkins & King, 2010). It is also advised that the directions explicitly instruct respondents to imagine themselves in the state of the hypothetical individuals. Au & Lorgelly (2013) found respondents who reported visualizing themselves as experiencing the state depicted in the vignette were more likely to rate their own state and the vignettes in the same way. When using multiple vignettes, the order should be random and varied, as some researchers have confirmed the presence of order effects (Buckley, 2008).

2.3.2.2 The C-Scale and B-Scale

The main statistical model used with AVs to capture differences in response scales is a non-parametric approach called the C-scale. It is the simplest mathematically and was devised by King and colleagues (2004). Their model recodes an individual's self-assessment relative to the set of vignettes and is shown in equation 6:

$$C_i = \begin{cases} 1 & \text{if } y_i < z_{i1}, \\ 2 & \text{if } y_i = z_{i1}, \\ 3 & \text{if } z_{i1} < y_{is} < z_{i2}, \\ \vdots & \vdots \\ 2J + 1 & \text{if } y_i > z_{iJ} \end{cases} \quad [6]$$

Where y_i is the categorical survey self-assessment for respondent i , z_{ij} is the categorical survey response for respondent i on vignette j ($j = 1, \dots, J$), and C_i is the new vignette-corrected variable *for respondents with identical ordinal rankings on all vignettes*. The new scale is expanded when compared to the original responses. For example, for 5 response options, 5 vignettes will be associated with the item ($J = 5$), and thus, C_i will now range from 1 to 11 (or $2J + 1$). The new adjusted score for each individual, C_i , is interpersonally comparable between individuals and can be interpreted and analyzed just like any ordinal variable (King et al., 2004).

The non-parametric method is limited in the following ways; all respondents must respond to all vignettes, all respondents must rank the vignettes in the same order, and there must be no ties in vignette rankings (King et al., 2004). If ties and inconsistencies do occur, C_i becomes limited due to a reduction in information (King et al., 2004). In response to the latter two limitations, the C-scale was updated to a more general equation in 2007 (King & Wand) to allow for more realistic data. Using the new formula, C_i could be either one value, like the original equation, or multiple values if ties or incorrect ordering exist (Table 6 provides an example of all possible C values for a survey item with two corresponding vignettes; King &

Wand, 2007; page 6). A ‘parametric supplement’ based on the censored ordered probit model is then used to determine the proportion of the sample falling into each category of C , which can then be used to create frequency distributions (King & Wand, 2007).

Table 6. All potential values of the non-parametric C-scale (King & Want, 2007) from one self-assessment, y , adjusted for two anchoring vignette responses, z_1 and z_2

Example	Survey responses						C
		1	2	3	4	5	
		$y < z_1$	$y = z_1$	$z_1 < y < z_2$	$y = z_2$	$y > z_2$	
1	$y < z_1 < z_2$	1	0	0	0	0	{1}
2	$y = z_1 < z_2$	0	1	0	0	0	{2}
3	$z_1 < y < z_2$	0	0	1	0	0	{3}
4	$z_1 < y = z_2$	0	0	0	1	0	{4}
5	$z_1 < z_2 < y$	0	0	0	0	1	{5}
6	$y < z_1 = z_2$	1	0	0	0	0	{1}
7	$y = z_1 = z_2$	0	1	0	1	0	{2, 3, 4}
8	$z_1 = z_2 < y$	0	0	0	0	1	{5}
9	$y < z_2 < z_1$	1	0	0	0	0	{1}
10	$y = z_2 < z_1$	1	0	0	1	0	{1, 2, 3, 4}
11	$z_2 < y < z_1$	1	0	0	0	1	{1, 2, 3, 4, 5}
12	$z_2 < y = z_1$	0	1	0	0	1	{2, 3, 4, 5}
13	$z_2 < z_1 < y$	0	0	0	0	1	{5}

Both the original C-Scale (King et al., 2004) and the updated version (King & Wand, 2007) require at least two assumptions, response consistency and vignette equivalence. King et al. (2004) define response consistency as “the assumption that each individual uses the response categories for a particular survey question in the same way when providing a self-assessment as when assessing each of the hypothetical people in the vignettes”, and vignette equivalence as “the assumption that the level of the variable represented in any one vignette is perceived by all respondents in the same way and on the same unidimensional scale, apart from random measurement error”. Although seemingly straightforward, in practice, there can be issues meeting these assumptions (d’Uva, Lindeboom, O’Donnell, van Doorslaer, 2011), and it is possible the C-scale can provide incorrect conclusions even when these two assumptions are met (Wand, 2013). It is also assumed the vignettes, as written, are unidimensional. If this assumption

is violated, individuals might be more likely to rank vignettes in different orders (King & Wand, 2007).

In his 2013 paper, Wand reports two more assumptions are needed for correct use of the C-Scale, interval equivalence and moderate vignette. He defines interval equivalence as “all individuals must agree on where the cut-points may be placed, even if they disagree on which cut-point to use at a particular location” and moderate vignette as “an anchoring vignette cannot be rated as being either the top or the bottom of the ordinal scale by any individual” (Wand, 2013). Wand (2013) proposes the use of the B-Scale, which relaxes the assumptions of vignette equivalence and interval equivalence, but requires response consistency. The B-Scale is statistically equivalent to the Aldrich and McKelvey scaling method, used by political scientists since 1977 (Hare, Armstrong, Bakker, Carroll, & Poole, 2015). The B-scale and C-scale often agree in ranking subjects, however, the B-scale claims less information when a tie occurs in vignette rankings (Wand, 2013).

2.3.2.3 Other statistical models

Alternative statistical models used with AVs are an extension of the ordered probit model, termed the compound hierarchical ordered probit model (CHOPIT; also referred to as the HOPIT) and a latent variable model. Like the non-parametric method, the first model allows researchers to correct for differences on the response scales by using vignettes; however, this approach performs the correction by estimating interpersonal thresholds (King et al., 2004). It improves on one of the main weaknesses of the non-parametric model; not all individuals need to respond to all vignettes (King et al., 2004). It also allows for researchers to include covariates and estimate their effects on the thresholds (King et al., 2004). The model equations are allowed

to vary over items and persons and include parameters for the *actual* level of the variable being measured, the *perceived* level, and the *reported* level (King et al., 2004).

The *perceived* level of the variable is now theorized as random and includes a normal error term (model notation from King et al., 2004):

$$Y_{is}^* \sim N(\mu_i, 1) \quad [7]$$

$$\mu_i = X_i\beta + \eta_i \quad [8]$$

where Y_{is}^* is the unobserved *perceived* level for each self-assessment item s for respondent i , μ_i is the *actual* level of respondent i , X_i represents observed covariates, and η_i is the independent normal random error (normally distributed with a mean of 0 and a variance of ω^2). The respondent theoretically turns the continuous perceived level into the *reported* level (y_{is}) for ordinal response category k depending on the threshold parameters (τ_{is}):

$$y_{is} = k \quad \text{if} \quad \tau_{is}^{k-1} < Y_{is}^* < \tau_{is}^k \quad [9]$$

The vignette component is set up much the same:

$$Z_{lj}^* \sim N(\theta_j, \sigma^2) \quad [10]$$

where Z_{lj}^* is the *perceived* level of person l for vignette j , θ_j is the *actual* level described in vignette j . The respondent turns the continuous perceived level into the *reported* categorical answer by:

$$z_{lj} = k \quad \text{if} \quad \tau_{l1}^{k-1} < Z_{lj}^* < \tau_{l1}^k \quad [11]$$

where k again represents the ordinal response category. All the thresholds for both the regular survey items (τ_{is}) and vignettes (τ_{l1}) are also functions of a vector of covariates, V_i and V_l (which

can overlap with the covariates in X_i), and unknown parameter vectors (γ_s and γ_l) as follows for survey items;

$$\tau_{is}^1 = \gamma_s^1 V_i, \quad [12]$$

$$\tau_{is}^k = \tau_{is}^{k-1} + e^{\gamma_s^k V_i} \quad (k = 2, \dots, K_s-1)$$

and vignettes;

$$\tau_{l1}^1 = \gamma_l^1 V_l, \quad [13]$$

$$\tau_{l1}^k = \tau_{l1}^{k-1} + e^{\gamma_l^k V_l} \quad (k = 2, \dots, K_l-1)$$

The vignettes and corresponding survey items should use the same response scale, so K_s will be equal to K_l . Interpersonal differences in responses are represented as variation in the thresholds over respondents (King et al., 2004).

In the initial publication, the HOPIT model was found to be unbiased using simulated data (King et al., 2004). It also performed better than the basic ordered probit model at detecting interpersonal differences between groups when the differences occurred in the same direction for the actual level and the threshold value (King et al., 2004). In an empirical example comparing political efficacy between Mexico and China, the vignette model significantly changed the conclusions drawn from the survey (King et al., 2004). For all response categories, the estimated thresholds were higher for Mexico; so, even though citizens of Mexico responded with lower raw levels of political efficacy than citizens of China, they had higher standards for what comprised political efficacy and higher adjusted mean values (as estimated by the model; King et al., 2004).

In addition to the initial empirical and Monte Carlo evidence provided by King et al. (2004), two articles provided further evidence regarding HOPIT's validity by comparing an objective measure to the uncorrected and HOPIT corrected results. Although the approach used in these two examples is useful in determining the validity of vignettes, in applied research, it is unlikely or impossible when measuring more abstract constructs like HRQoL. In the first example, half of respondents were randomly chosen to receive an eye test along with a visual acuity survey that included one item and 8 vignettes (King et al., 2004). According to the eye test, citizens from China had worse eyesight than citizens of Slovakia. The HOPIT model identified an effect in the same direction as the eye test, whereas the basic ordered probit model showed no difference in subjective visual acuity between the two groups (King et al., 2004). In another real dataset, researchers collected subjective information on drinking behavior in an Irish sample (i.e. whether they viewed their own behavior as a 'problem') along with the more objective count of drinks per week (Van Soest, Delaney, Harmon, Kapteyn, & Smith, 2011). When they corrected survey responses using vignette data and compared them to the objective drink count, they found the model corrected for interpersonal differences in responses well (Van Soest et al., 2011).

This model is slightly disadvantaged as it requires the assumptions of response consistency and vignette equivalence (King et al., 2004), it assumes individuals with the same covariate values interpret the scale in the same way (Wand, 2013), and relies on a linear relationship between covariates and the threshold parameters (Paccagnella, 2013). As for the two assumptions, response consistency can be examined in a number of ways. d'Uva, Lindeboom, O'Donnell, & van Dooslaer (2011) suggested response consistency be tested by checking if the vector of unknown parameters is the same for the survey items and vignettes as follows:

$$\gamma_s^k = \gamma_t^k \quad (k = 2, \dots, K-1) \quad [14]$$

However, this test still requires vignette equivalence (d’Uva et al., 2011). Similarly, the test they propose for vignette equivalence relies on response consistency, leaving both tests less than ideal (d’Uva et al., 2011). Van Soest et al. (2011) suggested response consistency be tested using the following equation:

$$\gamma_s^k - \gamma_s^{k-1} = \gamma_t^k - \gamma_t^{k-1} \quad (k = 2, \dots, K-1) \quad [15]$$

d’Uva et al., (2011) report the assumptions of this test are demanding as it requires a proxy for the content of interest, which is often unavailable.

Qualitative methods for examining response consistency are often used (Au & Lorgelly, 2013), as well as an approach that provides individuals new vignettes that were past descriptions of their own health (Kapteyn et al., 2011). Grol-Prokopczyk (2014) researched ways to improve response consistency and suggested the age of the hypothetical vignette individual be highlighted and potential gender differences examined closely. Vignette equivalence relies heavily on individual background and is one of the strongest criticisms to the use of AVs (Paccagnella, 2013). The wording of the vignette, especially when translating to compare levels in different cultural groups, is especially important and also depends on the targeted content (Paccagnella, 2013). When used on real data, neither the response consistency or vignette equivalence assumptions are commonly met, although the results can be mixed for different domains of the same construct (d’Uva et al., 2011; Hirve, Gomez-Olive, Oti, Debpuur, Juvekar, Tollman, et al., 2013; Hirve, Verdes, Lele, Juvekar, Bolmstedt, Tollman, et al., 2014; Jurges & Winter, 2013; Kapteyn et al., 2011). However, there is insufficient information regarding the performance of

the HOPIT model when the assumptions are not met, although it would be important knowledge to have since it is a common issue in real data.

Despite the disadvantages with HOPIT's assumptions, in recent years, researchers have started extending the HOPIT model to take into account other data characteristics and situations, increasing its practical use. Kapteyn, Smith, & van Soest (2007) extended the model so thresholds can vary both with individual characteristics and an unobserved heterogeneity term for individuals, Paccagnella (2011) modified it to account for sample selection bias, Angelini et al. (2011) extended it to be used with longitudinal data, and Peracchi & Rossetti (2012) allow for the simultaneous capture of multiple domains under a larger conceptual area (in their case, they used pain, mobility, sleeping problems, etc to capture overall 'health'). In 2013, Paccagnella compared the results of four HOPIT models mentioned above (not including Peracchi & Rossetti's) and found each to be useful. He recommended AVs and their models be used in new contexts, like adjusting for heterogeneity in larger populations, as opposed to solely comparing self-report of diverse cultures or countries.

The most recent statistical model was proposed by Bolt, Lu, & Kim (2014) and consists of a latent variable model, using AVs, controlling for multiple styles of response to Likert scales. Some examples of typical response styles are: individuals who tend to pick the most extreme categories, individuals who tend to choose the neutral/middle option, and/or individuals who tend to respond on average in a positive or negative manner (Bolt et al., 2014). The proposed model is an extension of the 2009 model by Bolt and Johnson, which only allowed for correction for one response style, and it is based under a multidimensional Item Response Theory framework (IRT).

The model states the probability that respondent r selects category k on item i is:

$$P(U_{ri} = k | \theta_r, s_r) = \frac{\exp(\alpha_{ik} \theta_r + s_{rk} + c_{ik})}{\sum_{h=1}^K \exp(\alpha_{ih} \theta_r + s_{rh} + c_{ih})} \quad [16]$$

where h indicates all the score categories and ranges from $h = 1, \dots, K$ (K being the highest value), θ_r is the actual level of the construct being measured for each respondent, s_r and c_i are vectors over the number of response categories with s_r indicating the specific response style for each respondent r accounting for differences in selecting each category, and c_i representing item difficulties (as related to category selection; Bolt et al., 2014). The α_i term is set to be 0 for all AVs so only the actual survey item directly estimate θ . Bolt et al. (2014) demonstrated their model using a real dataset, comparing country-level mean conscientiousness, and found it to work adequately.

An advantage of this model is that a baseline model can be estimated and compared to other models in order to test the presence of response styles (Bolt et al., 2014). The baseline model is a special case, and defined when the vector of s_r is set to be 0 (example $s_r = [0, 0, 0, 0, 0]$, for items with 5 response categories). In addition, this modeling method can accommodate any form of response style, be expanded to account for multiple response styles at one time, and accommodate multi-group comparisons (Bolt et al., 2014). This is particularly of interest when researchers use vignettes to account for response differences across cultural groups and is modeled by assuming the distributions of the construct being measured and response styles vary at both the group and respondent levels (Bolt et al., 2014). In addition to the mean estimates of the target construct for each group, response styles for each individual can also be interpreted through specific profiles estimated by the model and can be useful in understanding how response styles contribute to bias in the construct (Bolt et al., 2014). Furthermore, the researchers found the profiles in their example data often indicated styles not accounted for by the usual

literature-based response style types, providing further support that flexibility in estimating response styles is necessary in real data (Bolt et al., 2014).

A disadvantage of this model is it is highly complex, and thus, a Bayesian approach must be used to estimate the model and priors must be specified for each of the parameters (Bolt et al., 2014). The initial authors have some suggestions on which priors to use (Bolt et al., 2014), although the effect of misspecifying priors has not been specifically examined. Another disadvantage is the assumptions of response consistency and vignette equivalence still need to be met. However, Bolt and colleagues (2014) provide suggestions for how to statistically test these assumptions using a generalized form of their equation (which is an advantage over the HOPIT model). This model is also relatively new; it has not been tested extensively or used in applied research with diverse datasets.

2.3.2.4 Use of anchoring vignettes in the literature

AVs have started to be used more frequently in applied research to increase measurement precision by adjusting for differences in self-ratings between groups. Studies using AVs compare clusters in many diverse applied topic areas like disability severity (Heiland & Yin, 2015), extent of work disability (Kapteyn et al. 2007), job satisfaction (Kristensen & Johansson, 2005), and health system responsiveness (Rice et al., 2010). This section will briefly discuss two examples of how incorporating AVs into surveys have significantly changed study conclusions, although each set of authors use vignettes in a slightly different way. First, Salomon, Tandon, and Murray (2004) used the non-parametric ranking approach to examine pilot data from a World Health Organization survey on mobility (i.e. difficulties in moving around, walking, etc). Their goal was to discover if expectations of mobility differed between Asian countries (China, Sri Lanka, Pakistan, Turkey, Myanmar, & United Arab Emirates) or across demographic variables.

Although only a pilot study, they found evidence that expectations of mobility change with age (i.e. get worse as individuals get older) and might also vary by country (Salomon et al., 2004).

A second study done by Dowd & Todd (2011) used the HOPIT model to determine differences on self-reported health (6 domains: pain, sleep, mobility, memory, shortness of breath, and depression) by demographic variables such as education, race/ethnicity, and gender. They found that adjusting the self-reported data resulted in mixed differences across the domains for each covariate. Their main focus was exploratory and to examine if the estimated coefficients could determine which groups had lower cut-points and, in turn, which groups tended to rate the domains as having higher severity. To report just one of many interesting findings, the researchers found for shortness of breath, sleep, and mobility, Black and Hispanic individuals were significantly disadvantaged compared to Whites. This was in direct contrast to the unadjusted results, which indicated these groups had advantages or no differences. The researchers overall concluded constructs like health, which contain multiple heterogeneous domains, should be reported individually and not aggregated as they found very diverse results between domains and demographic variables (Dowd & Todd, 2011).

AVs are now being applied to other interesting issues in medical survey research, like identifying response shift and determining minimal clinically important differences. Response shift is defined as a change in meaning of a construct over time and is often referenced in quality of life research (Schwartz, 2010). Korfage, de Koning, & Essink-Bot (2007) used AVs to examine reprioritization, a component of response shift that indicates a change in the importance attributed to quality of life domains. They had adult men rank vignettes describing urinary, bowel, or erectile dysfunction two months prior and one month after a diagnosis of prostate cancer. The researchers found the vignettes describing the dysfunction were rated more

positively *after* diagnosis, and thus, vignettes can be useful in describing response shift (Korfage et al., 2007). Thissen and colleagues (2015) created anchoring vignette comparisons for hypothetical patients, and had real patients judge them for ‘change’. They used IRT methods to determine the minimal clinically important difference (MID), which was defined as the place where 50% of judges identified ‘a change’. Their study was successful in identifying the MID and their results were consistent with other MID identification methods (Thissen et al., 2015).

2.3.2.5 Considerations for use of anchoring vignettes with children

To be able to respond to AVs accurately, children will need to (1) be able to think abstractly, (2) imagine themselves in the place of the described vignette individual, (3) remember their own experiences in relation to the domains included in the vignettes, and (4) make a judgment whether the vignette example is ‘worse’ or ‘better’ than their experiences. The cognitive processes AVs require seem at the surface to be rather complex, however, most of these skills are required for answering basic self-report surveys, like the CDLQI. Children as young as 5 years old are capable of self-reporting HRQoL (Varni et al., 2007), and by 8-11 years old their responses become even more reliable (Riley, 2004).

Abstract thinking skills usually start to develop around kindergarten (ages 5-7 years old), when children begin to understand symbolism (Dumontheil, 2014). This ability is instrumental to reading and writing, but also in retrieving past information, thoughts, and memories (Dumontheil, 2014). In Piaget’s theory of cognitive development, a child typically begins thinking abstractly during the pre-operational stage (ages 2-7) although they continue to have trouble seeing the viewpoint of others until the next stage, concrete operational (Huitt & Hummel, 2003). Abstract thinking continues to improve into adulthood, and is related to prefrontal cortex processing and development (Dumontheil, 2014). In the context of Piaget’s

theory, the formal operational stage (ages 11-20 years) allows the development of advanced abstract thinking, deductive reasoning, and meta-cognition (Huitt & Hummel, 2003). This suggests that the use of AVs with children 11 years and older is reasonable.

Empathy is defined as “an affective response more appropriate to someone else’s situation than to one’s own” (Hoffman, 1994). Empathy is a naturally occurring human state that seems to have evolutionary benefits, and can be seen throughout the lifespan. Basic empathy is observed even in infants, who often cry upon hearing the distress of other children (Hoffman, 1994). Empathy for another’s feelings starts around 2-3 years old, when children recognize other people have different feelings than they do (Hoffman, 1994). The highest level of empathy (according to Hoffman, 1994) is ‘empathy for another’s life condition’, which develops during late childhood (Reiffe, Ketelaar, & Wiefferink, 2010). This would likely be the level of empathy needed to complete AVs, as respondents would need to be able to imagine how the described person felt on a weekly basis given their explained symptoms.

In general, researchers believe episodic memory, or the ability to recall specific events, commences around age 4 (Raj & Bell, 2010). However, development of this ability continues through adolescence and into adulthood (Ghetti, DeMaster, Yonelinas, & Bunge, 2010). Research into the self-report of psychological and physical constructs in children are mixed, with some studies finding self-report and more objective behavioral measures of the same construct do not correlate (Beyer, McGrath, & Berde, 1990). However, it is likely certain domains are harder to recall than others. Subjective experiences like pain (measured in the Beyer et al. study, 1990) are likely harder to recall than more objective experiences, and emotional memories and behaviors are more easily remembered than non-emotional experiences (Davidson, Luo, & Burden, 2010). Thus, it is likely children will be better able to recall HRQoL experiences that

have stronger emotional ties. Different children have varying levels of empathy (Reiffe et al., 2010) and cognitive abilities (Dumontheil, 2014), so it is likely some children will be better than others at responding to AVs, although general ability should increase with age.

2.3.2.6 Sample size considerations

There are no current recommendations regarding sample size requirements for the C-scale/B-scale, the HOPIT, or the latent variable approach. However, in the literature, sample sizes are large. For the two introductory articles on anchoring vignettes, authors used a sample of greater than 2,000 people for both the C-scale/B-scale and HOPIT (King et al, 2004 and King & Wand, 2007). For the latent-variable approach, Bolt's article used a sample of 2,965 (Bolt et al. 2012). In rare disease populations like pediatric localized scleroderma, obtainable sample sizes are likely small, (i.e. less than 100 people). Thus, the non-parametric C-scale and B-scale are likely the best options for these cohorts.

2.3.2.7 Summary: Anchoring vignettes

AVs are useful in adjusting self-reported data for differences in response styles resulting from individual or cultural disparities. Although AVs are a promising approach for a multitude of reasons, they are not being extensively used in practice, especially in quality of life research (Szkulicka-Debek, Drozd, Bem, Kierpurska, & Mazure, 2015). Szkulicka-Debek and colleagues (2015) found only seven quality of life studies from 2011-2014 that incorporated AVs into their methodology. The CHOPIT/HOPIT model seems to be the most widely implemented statistical method and also has the strongest research base, although its assumptions can be problematic. The new latent variable approach (Bolt, Lu, & Kim, 2014) seems promising, although more methodological and applied research is needed to determine how it functions

under a wide variety of data characteristics. AVs have not been studied in small sample sizes, in rare diseases (including LS), or pediatric patients. Achievable sample size in LS will likely limit the utility of statistical models in this population.

2.4 SUMMARY OF LITERATURE

It is important for HRQoL surveys be included in treatment studies for LS. Current knowledge of the HRQoL impact of LS is limited by the significant methodological concerns of most published research, lack of consensus by physicians over which survey to implement, and the lack of valid standardized outcomes from which to choose from. No published research studies examine the psychometric properties of currently used HRQoL questionnaires in LS patients. The most popular survey, the DLQI/CDLQI, is not ideal based upon initial examinations of its psychometric properties, although it does attempt to capture some important aspects of LS HRQoL. The unique aspects of LS lend itself well to a disease-specific HRQoL survey and the new survey should include items that assess uncomfortable skin sensations, physical functioning, body image, peer relationships, and treatment side effects.

When developing new measurement surveys, the ‘validity as an argument’ approach requires evidence supporting the intended interpretations from multiple sources. Small sample size is a limitation when developing surveys for use with rare patient populations and appropriate analysis methods need to be considered.

Anchoring vignettes (AVs) have many interesting and practical applications to measuring HRQoL in LS but are not yet studied in children or in rare disease populations. Three main statistical methods are used with AVs, and the practicality of each needs to be examined with the

specific target population. Different children have varying levels of empathy (Reiffe et al., 2010) and cognitive abilities (Dumontheil, 2014), and it is likely that the ability to accurately answer AVs will increase with age.

3.0 METHODS

The study consisted of three stages (1) survey development, (2) a small pilot study, and (3) a larger field test. The survey was developed using information from the literature, focus groups with patients, and input from physicians and measurement experts. The pilot study consisted of qualitative interviews with patients regarding the developed items and responses, and results were used to revise the survey appropriately. The field test included larger data collection from two diverse, specialty clinics in North America. All three stages of this research project were designed to contribute different types of validity evidence for the survey and its proposed interpretations and uses. The use and feasibility of anchoring vignettes were also examined at each stage of the project. Appropriate IRB approval was procured and consent obtained from each subject and their parent prior to initiating study procedures.

3.1 DEVELOPMENT OF ORIGINAL SURVEY

The Localized Scleroderma Quality of Life Instrument (LoSQI) was designed to be a disease-specific HRQoL survey to be used with pediatric LS patients. The pilot version of the survey included 18-items in the main body of the survey assessing three main domains, as well as an optional 11-item medication subscale for patients who were actively being treated for LS (Appendix B). The LoSQI focused on the unique aspects of localized scleroderma not currently

captured by generic HRQoL tools and expanded on the domains included in the CDLQI found to be important in prior studies. Common LS ECMs and their associated general symptoms were also considered during survey development.

3.1.1 Proposed use and interpretations

The proposed use of the LoSQI was to accurately measure disease-specific quality of life in pediatric localized scleroderma patients. The LoSQI was intended to be part of a larger battery of PRO's to be used in LS research and clinical trials, thus, general HRQoL and family impact scales should be collected in tandem.

3.1.2 Content domain & development

The content domain was developed over a period of four months. Initially, the literature was used to build a list of common HRQoL domains from currently used questionnaires. Expert opinion was sought from two pediatric rheumatologists with experience treating LS patients and the theoretical domains revised and restructured. The survey developer reviewed the psychometric analysis of the CDLQI (Ardalan et al., in submission) so that items of interest could be expanded and explored further in the new survey. Focus group data from pediatric LS patients were also evaluated for inclusion of additional items based upon common themes coded from the transcripts. Finally, additional expert opinion was requested to finalize the theoretical domains and develop items.

All items went through multiple rounds of expert evaluation and revision, in accordance with the guidelines of the PROMIS network (DeWalt, Rothrock, Yount, & Stone, 2007), to

ensure the wording was simple, slang was avoided, and none were double-barreled (i.e. items that ask for one response to two different questions). In accordance with the PROMIS guidelines, items were developed so that they would have similar formatting and structure (DeWalt et al., 2007). All items had a stem added that indicated each described symptom should be considered in terms of where their localized scleroderma is. For example, to differentiate itchy skin due to psoriasis or eczema from itchy skin due to a scleroderma lesion, the item was written as “itchy skin *where my scleroderma is*”.

The LoSQI consisted of four theoretical domains: (1) skin sensations, (2) physical functioning and musculoskeletal sequelae, (3) body image and social support, and (4) medication attitudes and side effects. As mentioned above, support for the included domains was gathered via three main sources (1) the current LS published literature, (2) past focus groups with pediatric LS patients, and (3) opinions from content experts (i.e. board-certified pediatric rheumatologists who specialize in treating LS patients) and measurement/survey-design experts.

3.1.2.1 Theoretical Domain 1: Skin sensations

Domain 1 quantified the patient’s experience of uncomfortable skin sensations. As mentioned in section 2.2.2.2, recent research into LS HRQoL documented that uncomfortable skin sensations are much more bothersome to patients than previously thought (Das et al., 2014; Kroft et al., 2008). Items under this domain specifically asked how frequently the patients are bothered by pain, itch, tightness, and soreness in their skin. Consequently, the three items generated from this domain ask about (1) *painful skin*, (2) *itchy skin*, (3) *uncomfortably tight skin*.

3.1.2.2 Theoretical Domain 2: Physical functioning and musculoskeletal sequelae

Domain 2 quantified the effect of LS on fine motor function during everyday activities (e.g. writing, texting, typing), limitations to elective activities (e.g. playing instruments or sports, or participating in group activities like gym class or dance), worry or stress, and joint or muscle pain (which is clinically different from skin pain). Section 2.2.2.3 provides a more thorough examination regarding physical functioning in LS patients. Four specific items concerning every day and elective activities were as follows (1) *problems using my hands when I'm writing, texting, or typing*, (2) *problems using my hands when I'm writing, texting, or typing **for a long time***, (3) *problems doing active things like running, playing sports, or dancing*, and (4) *problems when I am doing fun things like painting or playing an instrument*. One item was created to quantify the patient's worry regarding their functioning; *worry about being able to do certain activities*, and three items for muscle/joint pain; (1) *aches in my joints (like knees, hips, fingers, toes, ankles, elbows)*, (2) *stiff joints*, and (3) *my muscles hurting*.

3.1.2.3 Theoretical Domain 3: Body image and social support

Domain 3 consisted of two subdomains; (1) appearance and body image and (2) peer relationships and social support. The goal of this domain was to identify both the negative effects of LS on physical appearance (i.e. does their skin appearance bother them or do they feel different from their peers) and to determine how peer interactions are driven by LS. These two subdomains were combined into one because during the focus groups a strong interaction was noted between a patient's feelings towards their body and their friendships. Additionally, two CDLQI items attempting to measure similar constructs loaded onto the same factor during an exploratory factor analysis with pediatric LS patients (Ardalan et al., *in submission*). See section 2.2.2.4 and 2.2.2.5 for more information.

Two items were generated to reflect appearance and body image; (1) *feeling embarrassed because of how my body looks*, and (2) *feeling different than other people because of my scleroderma*. An additional item asked respondents to indicate how often covering up their scleroderma *with things like long sleeves, long pants, makeup, or retainers* bothered them.

Four items attempted to quantify how much patients were bothered by negative peer interactions. These items were based upon the common focus group theme of patients' worry or negative feelings towards encountering people who do not already know their condition. Two items attempt to measure that information; (1) *feeling nervous when I am around new people who don't already know about my scleroderma*, and (2) *feeling upset when people ask questions about my scleroderma*. Two additional items were generated based on negative peer interactions, (1) *getting teased about the way I look*, and (2) *being bullied because of the way my skin, face, or body looks*.

3.1.2.4 Theoretical Domain 4: Medication attitudes and important side effects

As further discussed in section 2.2.2.6, there is a strong potential for serious negative side effects from the systemic medications commonly used to treat LS. The final theoretical domain attempted to identify the patient's attitude towards LS treatment and the extent to which the patient was suffering from several important side effects directly attributed to their medications. This domain is optional and can be added to the end of the main body of the survey for LS patients on medications at their clinical visit. Patients are especially unlikely to be on medications at their first visit to clinic and at follow-up visits for patients in remission.

Two items were designed to capture typical emotional responses to being on medications (1) *worry about medication side effects*, and (2) *feeling embarrassed that I need to take medications*. Nine additional items asked about general malaise or other significantly bothersome

side effects that patients directly attribute to their medications which include (1) *feeling sick right after I take my medications*, (2) *feeling hungry all the time*, (3) *stomach pain*, (4) *not feeling like eating*, (5) *feeling like they are in a fog or that it's hard to think clearly*, (6) *feeling tired*, (7) *having headaches*, (8) *gaining weight*, and (9) *vomiting*. In the focus groups, patients mentioned that even though the medication side effects were a major problem, the belief in the medications effectiveness helped tolerate them better. Thus, the final item asked patients if they agree with the statement: *I believe my medications will help me to get better*.

3.1.2.5 Theoretical factor structure

The theoretical factor structure of the LoSQI is shown in Figure 4. The structure was based on a bi-factor model (Chen, Sousa, and West, 2006; see section 3.3.4.2 for details).

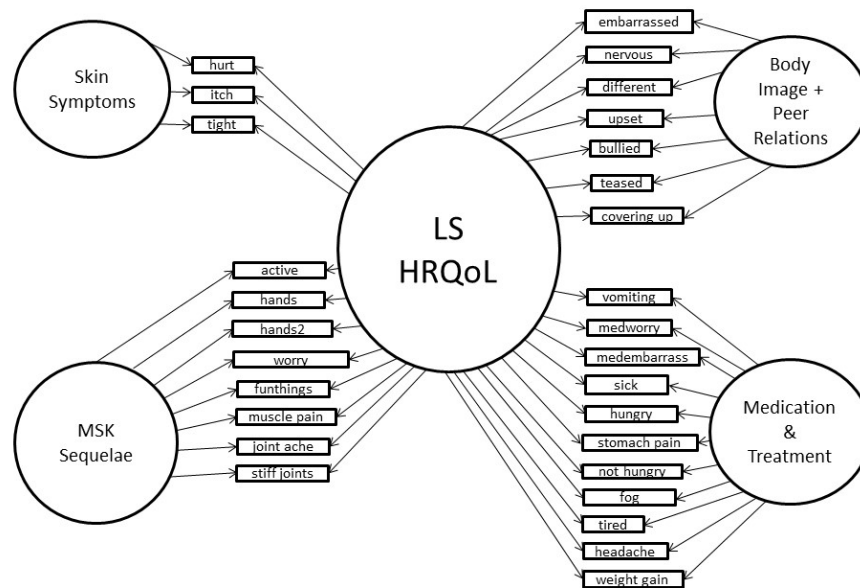


Figure 3. Theoretical factor structure of the Localized Scleroderma Quality of Life Instrument (LoSQI). LS = localized scleroderma; HRQoL = Health related quality of life; MSK = musculoskeletal.

3.1.3 Timeframe

For all items on the survey, patients were asked to reflect on how they felt during the *past seven days*. The intended use of the LoSQI, the duration and frequency of HRQoL concepts, and the characteristics of LS were all taken into account when determining the appropriate time frame to reference, aligning with the FDA's standards (2009). The purpose of the LoSQI is to accurately estimate HRQoL by obtaining a snapshot of the patient's LS life impact at the general time of their clinical visit. Increasing the recall period to a month would provide less reliable self-report, as it would be harder for children to reflect over 30 days compared with a week. LS disease features and treatment side effects often are long lasting and should not change significantly within the chosen timeframe, although the appropriateness of the timeframe was evaluated during the pilot study (see section 3.2).

3.1.4 Response scale

Response options are on a 4-point scale, as 4-options have been found to be ideal when surveying children (Borgers, Hox, Sikkel, 2004). Please note that the response options are NOT on a frequency scale (i.e. sometimes, often, very often), as the interest is not simply the amount of symptomatology present, but the extent to which the symptom *bothers* the patient.

0 - Does not bother me

1 - Bothers me a little

2 - Bothers me a medium amount

3 - Bothers me a lot

3.1.5 Age range

The LoSQI was initially evaluated in 8-20 year olds. As mentioned before, children as young as 5 years old can be capable of self-reporting HRQoL (Varni et al., 2007), and by 8-11 years old, the self-reports become even more reliable (Riley, 2004). In this study, children were required to both self-report on HRQoL and answer AVs. Children 8 years old and older are likely to have the cognitive ability to think abstractly, consider another's perspective, and empathize with the vignette characters (see section 2.3.3.6), although it is very likely older children will be better equipped cognitively than younger patients. Children diagnosed with LS and seen by pediatric rheumatologists often continue to fall under the care of the pediatric group after age 18, before being transitioned to an adult physician around age 20. Thus, the target age range extended through 20 years old.

Although the FDA recommends the use of narrow age range bands to assure the language of items is developmentally appropriate (2009), that approach was not used for the LoSQI. Currently, the proposed use of the LoSQI is to measure HRQoL in LS patients in clinic, and not to evaluate treatment options, which is the focus of the FDA Standards. The practical use of the LoSQI was determined to be more important at this point than potential advantages of narrow age-range bands. The development of items that could be understood across all age levels was of utmost importance during survey construction, and the readability of items and respondent understanding were extensively evaluated during the pilot study using stratified samples based on age (see section 3.2).

3.1.6 Scoring of theoretical domains

Item responses were scored on a 0-3 scale, with *doesn't bother me* resulting in '0' points and *bothers me a lot* '3' points. Subscale scores were calculated for each theoretical domain; domain 1 (skin sensations) ranged from 0 to 9 (3 items), domain 2 (physical functioning) ranged from 0 to 24 (8 items), and domain 3 (body image and social support) ranged from 0 to 21 (7 items).

3.1.6.1 Total score

The pilot LoSQI total score ranged from 0 to 54 and was calculated by adding up the three theoretical domain scores (excluding the medication subscale). Higher scores indicated more negative quality of life impact due to LS.

3.1.6.2 Scoring of optional medication subscale

The optional medication domain consisted of 11 items designed to detect the patients' attitudes towards taking their medication and medication side effects. The theoretical-domain score was calculated as above and range from 0 to 33. The appropriateness of how to report scores (separate subscale scores versus one total score) was examined during the field test (Section 3.3).

3.1.7 Anchoring vignettes

Anchoring vignettes were generated separately for each theoretical domain of the LoSQI (King et al., 2004), as domains were considered unidimensional. This approach was limited due to the unvalidated factor structure of the LoSQI and the diverse items under each theoretical domain. However, generating vignettes at the item level would increase the length of the survey

considerably. To ensure unidimensionality of the vignettes, four items, one from each domain, were selected as targets for AV sets. These target items were chosen based on their perceived importance; multiple children in the focus group transcripts mentioned each concept and they were discussed at length. For each target item, four vignettes were generated to represent scenarios at the four levels of the response options, resulting in 16 separate AVs. Whenever possible, the language of the AVs scenarios was taken directly from the focus group transcripts. At the time of survey development, researchers had not studied the use of AVs at all in children and the feasibility of this technique was unknown. A major goal of this study was to provide initial evidence on the use of AVs with children, but again, it was preliminary in nature and a comparison between item- and domain-level vignettes was beyond the scope of this project.

Two forms were created, one for self-identified females and a second for self-identified males (Appendix C). A gender-specific name list was generated from an online database that listed the most popular names by race from the 1990s (around when the targeted patients would have been born). The typical patient population for pediatric LS is majority white and non-Hispanic. However, common names for minority groups were also included to ensure that the AVs properly represented all subgroups (see Table 7). The names were applied to the AV items using a random number generator.

Table 7. Male and female names chosen for anchoring vignettes.

Female	Male
Abby	Anthony
Madison	Mike
Jessica	Joseph
Jasmine	Jordan
Christina	Christopher
Rachel	Jose
Jordan	Caleb
Kelly	James
Maria	Daniel
Kayla	Brandon
Alexandra	Matthew
Katie	William
Laila	Nathan
Sarah	Ethan
Aliyah	Tyler
Jennifer	Kevin

3.1.7.1 AV Set 1: Skin sensations

The first LoSQI item targeted for anchoring vignette development was related to *skin itch*. To indicate severity, this AV set described the physical act of scratching but also the emotional effects of itching (i.e. annoyance) and the influence on sleep (which was a common focus group complaint). The specific vignette scenarios related to each response option are listed below:

Does not bother her: Abby’s skin gets itchy where her scleroderma is, but she does **not** usually notice it.

Bothers her a little: When Madison’s scleroderma itches, it feels better right after she scratches it. Sometimes, she feels annoyed that she has to scratch it.

Bothers her a medium amount: Jessica’s skin is itchy during the day, but it gets especially itchy at night. On those nights, it makes it hard for her to fall asleep.

Bothers her a lot: Jasmine's skin itches almost every day and night. She scratches her skin most of the day and when it is time to go to bed, it keeps her awake for hours.

3.1.7.2 AV Set 2: Physical functioning

The target item for the second set of AVs was; *worry about being able to do certain activities because of my localized scleroderma*. Gym class was chosen as the specific activity because most children currently enrolled in school are required to take some type of physical education. Thus, this example was designed to be easily understandable and applicable to the majority of patients. The specific vignette scenarios related to each response option are listed below:

Does not bother her: Christina does **not** usually worry about going to gym class because of her scleroderma. She is comfortable with what her body can and cannot do and does **not** mind taking breaks if she needs to.

Bothers her a little: Rachel worries about going to gym class because of her scleroderma. She dislikes that she sometimes has to skip participating, but she is used to it at this point.

Bothers her a medium amount: Jordan worries about going to gym class because of her scleroderma. She sometimes **cannot** do the same activities that her friends and classmates are doing, and it makes her feel embarrassed to not participate.

Bothers her a lot: Kelly worries about going to gym class because of her scleroderma. She is worried about a lot of things; not being able to do the activities that everyone else is doing, accidentally hitting her skin and hurting herself, or having people ask her questions about why she's not participating.

3.1.7.3 AV Set 3: Body image and social support

The LoSQI target item for the third set of AVs was; *getting upset when people ask questions about my skin*. This item examined the emotional response to questions about the disease and its physical symptoms. The specific vignette scenarios related to each response option are listed below:

Does not bother her: Maria doesn't mind telling people about her skin and localized scleroderma. She knows how to explain it well, and understands that usually they are just curious.

Bothers her a little: Kayla does **not** like it when people ask questions about her skin, but she is used to explaining it so it does **not** upset her that much.

Bothers her a medium amount: Alexandra gets annoyed when people ask questions about her skin. She tries to explain her scleroderma to them but she sometimes gets mad.

Bothers her a lot: Katie does **not** like it when people ask her about her skin. It makes her feel sad and mad when they ask questions and it makes her uncomfortable to talk about it.

3.1.7.4 AV Set 4: Medication attitudes and side effects

The LoSQI target item for the optional medication domain was; *taking my medications makes me feel sick*. 'Feeling sick' was interpreted as nausea and fatigue, both common side effects of most systemic medications (like prednisone and methotrexate). This anchoring vignette also attempted to incorporate limitations to school and weekend activities as a way to indicate severity. The specific vignette scenarios related to each response option are listed below:

Does not bother her: Laila does **not** like her medications, but she usually feels fine after taking them.

Bothers her a little: Sarah feels nauseous after taking her medication(s), but when this happens she can still go to school and hang out with her friends like she usually does.

Bothers her a medium amount: After taking her medication(s), Aliyah feels nauseous and tired afterwards. It can be so bad that she has to miss school or skip fun activities on the weekend.

Bothers her a lot: Jennifer feels very tired and nauseous after taking her medication. She feels so sick on those days that she **cannot** go to school or hang out with her friends and family.

3.1.7.5 Instructions for anchoring vignettes

The literature indicates that the vignette assumption of response consistency is more likely to be met when respondents imagine themselves as the vignette character (Au & Lorgelly, 2013). Thus, the respondents were explicitly instructed to do so. Appendix C provides the administrative version of the AVs with corresponding instructions that was used in the pilot study.

3.1.8 Order of items and vignettes

The LoSQI items were organized within the survey by theoretical domain, while the order of the anchoring vignettes scenarios were randomly generated. All of the AV scenarios were administered to subjects prior to the LoSQI self-report items in order to calibrate the subjects' responses (Hopkins & King, 2010).

3.1.9 Background items

Three background items were included in the LoSQI (Appendix B). The first two background items asked patients how their disease or medication side effects were doing since their last visit to clinic (better/the same/worse). These two items were designed to represent the patient's opinion on any changes in the underlying construct measured by the LoSQI, and could be used in the future to examine responsiveness change.

The third background item was included to decrease the likelihood of patients completing the medication side effects subscale in error. Patients were asked if they were currently on systemic medications: if they indicated 'yes' they were prompted to continue with the scale, but if they indicated 'no' they were prompted to stop and return the survey to the administrator.

3.2 PILOT STUDY

3.2.1 Purpose

The main purpose of the pilot study was to obtain input from pediatric LS patients regarding the under- and over- representation of the included content domains, understandability and readability of the survey items, and appropriateness of the recall period. A secondary goal of this stage was to examine the basic feasibility of using anchoring vignettes with children. The vignettes were considered ‘feasible’ if children indicated understanding of the vignettes at the item-level and there was no evidence that the assumptions were violated (King et al., 2004; see section 2.3.3.2 for a more thorough discussion of the assumptions). Although parts of these assumptions can be checked quantitatively, qualitative methods are appropriate (King et al., 2004) and were used in this study.

3.2.2 Sample

Seventeen patients were recruited from the Children’s Hospital of Pittsburgh of UPMC scleroderma clinic from the National Registry of Childhood Onset Scleroderma (NRCOS). The NRCOS is one of the largest registries of its kind in the United States and has enrolled almost 400 patients since 2003. It includes extensive standardized, prospective data and biological sample collection. For this study, potential subjects were stratified into 3 age groups to ensure

enrollment included 8-10 year olds, 11-14 year olds, and 15-20 year olds (with a goal of 5 patients per age group).

3.2.3 Methods

The one-on-one qualitative interviews were informal and took place during the patient's regular clinic visit. In order to keep the environment as comfortable and familiar as possible, parents were present during the interviews but were asked to remain silent to allow their children to answer the items independently. Parents were invited to express their own opinions on the survey after their child. The facilitator used a general script to conduct the interviews (see Appendix A). The child was given the LoSQI and AVs to complete independently, and the facilitator documented the time it took to complete the survey. The facilitator wrote down the patient's answers in vivo, and immediately after the clinic visit, wrote notes on the patient's answers and their own observations from the meeting. The interviews were not recorded or videotaped to protect patient's privacy. Also, to limit the time of the interview, only a select number of items and vignettes were specifically discussed with each patient. The chosen items varied by patient using a spiral technique with three forms, and each item/vignette were discussed at least once in each age group (see Table 8).

Table 8. Spiral technique for item coverage for pilot study.

	Total number of items	Number of items discussed with each patient
LoSQI Items	18	6
Medication Subscale Items	12	4
Anchoring Vignettes	16	5 or 6

3.2.3.1 Administration

Standard administration of the LoSQI occurred during the patient's regularly scheduled visit to clinic, within a private room. Paper copies of the AV's and the LoSQI were provided and the patient instructed to answer in pen. Patients were presented with the AV's first and then the survey items. The respondent was told that they were able to ask the examiner questions as they completed the survey, although the examiner asked their parent to remain silent to protect the independence of the child's responses.

3.2.4 Analysis

Results from the pilot study were examined within each age range to determine functionality of the LoSQI for 8-10 year olds, 11-14 year olds, and 15-20 year olds. Major themes from the qualitative interviews were aggregated for both the LoSQI items and anchoring vignettes. The readability and understanding of items at each age level were reported, along with the appropriateness of the recall period. Common issues regarding under- and over-representation of the construct were documented, as well as if patients indicated duplicate or repetitive items.

The average times to complete the vignette section were reported along with the range of times over the entire sample. Overall, understanding and readability of items were concluded based upon answers to the pilot interview questions. In addition, responses concerning the understanding of the vignettes were compiled across subjects to determine if the vignette equivalence assumption was met. The assumption was considered met if "the level of the variable represented in any one vignette is perceived by all respondents in the same way" (King

et al., 2004), even if subjects chose different response options. The response consistency assumption, which indicates if individuals use the vignette response options similarly to their own self-ranking, was determined based on interview questions linking the conceptual vignette to the self-report item response.

3.2.4.1 Item revision

Based on the results from the pilot study, the LoSQI and AVs were revised prior to the field test. Items were candidates for revision or deletion based on the FDA Standards (2009) seen in Table 9. For the pilot study, evidence for revisions focused primarily on ‘clarity or relevance’. Items were candidates for deletion or revision if many patients asked for clarification while completing the survey, the item was reported to be ‘not relevant’ by most patients, or an item was frequently skipped. For the pilot study, response ranges were examined but since the sample size was small and LS patients can have very diverse symptoms, at this time items were not be deleted if a high percentage of patients respond at one of the extreme ends. Large-scale item revisions or major modifications to the survey format were confirmed by expert opinion and required submission of a modification to the IRB.

Table 9. Common reasons for changing an item during PRO Development from FDA Standards (2009; page 9).

Item Property	Reason for Change or Deletion
Clarity or relevance	<ul style="list-style-type: none"> ● Reported as not relevant by a large segment of the target population ● Generates an unacceptably large amount of missing data points ● Generates many questions or requests for clarification from patients as they complete the PRO instrument ● Patients interpret items and responses in a way that is inconsistent with the PRO instrument's conceptual framework
Response range	<ul style="list-style-type: none"> ● A high percent of patients respond at the floor (response scale's worst end) or ceiling (response scale's optimal end) ● Patients note that none of the response choices applies to them ● Distribution of item responses is highly skewed
Variability	<ul style="list-style-type: none"> ● All patients give the same answer (i.e., no variance) ● Most patients choose only one response choice ● Differences among patients are not detected when important differences are known
Reproducibility	<ul style="list-style-type: none"> ● Unstable scores over time when there is no logical reason for variation from one assessment to the next
Inter-item correlation	<ul style="list-style-type: none"> ● Item highly correlated (redundant) with other items in the same concept of interest
<i>Ability to detect change</i>	<ul style="list-style-type: none"> ● Item is not sensitive (i.e., does not change when there is a known change in the concepts of interest)
Item discrimination	<ul style="list-style-type: none"> ● Item is highly correlated with measures of concepts other than the one it is intended to measure ● Item does not show variability in relation to some known population characteristics (i.e., severity level, classification of condition, or other known characteristic)
Redundancy	<ul style="list-style-type: none"> ● Item duplicates information collected with other items that have equal or better measurement properties
Recall period	<ul style="list-style-type: none"> ● The population, disease state, or application of the instrument can affect the appropriateness of the recall period

3.3 FIELD TEST

3.3.1 Purpose

The main goal of the field test was to provide quantitative validity evidence for the LoSQI from multiple sources. The analysis included examination of patterns of missing or skipped items, reliability of scores, internal structure of the survey, convergent and divergent validity evidence, and test-retest reliability. The preferential reporting of scores was also explored (i.e. total score versus subscale scores, and how to incorporate the medication subscale). The secondary goal of the field test was to further examine the use of AVs within this population. Differences in ranking of the vignettes were examined and a statistical model used to determine if the AVs changed the conclusions provided by the unadjusted scores or provided further clarity into the construct of LS HRQoL.

3.3.2 Sample

Seventy-four pediatric LS patients aged 10-20 years old² were enrolled in the field test. Patients who participated in the pilot study could also participate in the field test. Patients were recruited from specialized scleroderma clinics at two diverse domestic sites: the Children's Hospital of Pittsburgh of UPMC (Pennsylvania) and the University of Texas Southwestern Medical Center

² The lower end of the age range was modified from 8 years old to 10, based on the pilot study (see Results Section 4.0 for more details).

(Texas). The sites house two of the largest localized scleroderma registries in the country, the National Registry for Childhood Onset Scleroderma (NRCOS) and the Morphea in Adults and Children (MAC) cohort. Together, the registries have enrolled over 900 patients and continue to actively enroll, making them uniquely qualified to obtain a large enough sample of pediatric patients with this rare condition. Patients were eligible for enrollment into the study if they were between the ages of 10 and 20 years old and had a physician confirmed diagnosis of localized scleroderma. As part of patients' participation in the registries, general demographics (e.g. age, gender, ethnicity/race) and disease history/symptomatology (e.g. disease subtype, age of diagnosis, age of disease onset, treatment history, disease duration, and extracutaneous manifestations) were collected from the patients' medical records. Approvals from the affiliated IRBs were sought prior to enrollment and consent/assent obtained.

3.3.3 Methods

3.3.3.1 Administration procedures

As in the pilot study, administration of the LoSQI was standardized and occurred during the patient's regularly scheduled visit to the clinic, within a private clinic room. Paper copies of the AV's and the LoSQI were provided and the patient answered in pen. The vignettes were administered prior to the survey items in order to calibrate responses and promote consistent definitions of the response scale (Hopkins & King, 2010). The respondents were told to ask the examiner any questions they had as they completed the survey, although the examiner asked any parent whom was present to remain silent to protect the independence of the child's responses. After the initial administration in the clinic, subjects were provided with a duplicate copy of the questionnaire to complete at home and mail back using a prepaid envelope, within 2 weeks. One

site offered patients the option of scanning and uploading the duplicate survey through their secure medical record system. Patients at both sites already completed additional HRQoL measures as part of their participation in the registries (see Section 3.3.3.2 for a specific list of questionnaires). The order of the LoSQI and the additional HRQoL measures were counterbalanced to control for order effects.

3.3.3.2 Additional assessment instruments

Patient-reported outcomes (PRO's)

The following additional HRQoL questionnaires were collected; the Childhood Dermatology Life Quality Index (CDLQ), the Childhood Health Assessment Questionnaire (CHAQ), the PedsQL Rheumatology Module, and the Methotrexate Intolerance Severity Score (MISS). The CDLQI (Lewis-Jones & Finlay, 1995) is a 10-item survey measuring how patients are affected by dermatological conditions. It includes six domains covering skin symptoms, leisure, school/holidays, personal relationships, sleep, and skin treatment. The CDLQI is commonly used with pediatric LS patients, despite no supporting validity evidence for this population. However, the psychometric properties of the survey have been extensively studied with 102 articles being published as of 2013 in 14 skin conditions (Salek, Jung, Brincat-Ruffini, MacFarlane, et al., 2013). It is commonly used with pediatric patients aged 3-16 years old.

The Childhood Health Assessment Questionnaire (CHAQ) is a 30-item survey measuring general physical functioning and includes three main domains; disability, discomfort, and pain (Singh, Arthreya, Fries, Goldsmith, 1994). It was designed to assess health status in children with juvenile rheumatoid arthritis and takes approximately 10 minutes to administer (Klepper,

2003). It has acceptable reliability (.96; Klepper, 2003) and validity evidence exists for patients aged 1-19 years old (parent proxy is available).

The PedsQL Rheum is a 22-item survey specifically designed for use with children who suffer from general rheumatic conditions (Varni, 1998). It includes five domains, pain and hurt, daily activities, treatment, worry, and communication, and has validity evidence for patients 8-18 years old (Varni, 1998). Mean scores are calculated for each domain and interpreted; there is no total score.

The Methotrexate Intolerance Severity Score (MISS) was designed for use with patients 2-18 years old, to quantify the extent of methotrexate (MTX) intolerance in patients with juvenile idiopathic arthritis (Bulatovic, Heijstek, Verkaaik, van Dijkhuizen, et al., 2011), psoriatic arthritis, and rheumatoid arthritis (Calasan, van den Bosch, Creemers, Custers, Heurkens, et al., 2013). The MISS has 5 domains and includes items about stomachaches, nausea, vomiting, sore mouth, and behavioral symptoms. At the time of the study, there was no published validity evidence for use of this survey with LS patients. This questionnaire was only provided to patients who were currently on MTX at the time of their clinic visit.

Finally, as part of their participation in the NRCOS, patients completed a 10cm visual analogue scale (VAS-pt) asking them to indicate their disease impact on a 0 (no problem) to 100 scale (very severe problem).

Parent-reported outcomes

Since parents are also a vital data source when examining quality of life in children, for subjects under the age of 18, parents were asked to report their child's disease severity on a 0 (no problem) to 100 (very severe problem) visual analog scale (VAS-par). In addition, parents completed the PedsQL Family Impact Module (PedsQL FIM) consisting of 36-items. The

PedsQL FIM is a multidimensional scale that measures physical, emotional, social, and cognitive functioning, communication, worry, daily activities, and family relationships. There is validity evidence for use of the PedsQL FIM in children with chronic health problems (Varni, Sherman, Burwinkle, Dickinson, & Dixon, 2004).

Physician-reported outcomes

Commonly used physician-scored outcome measures were collected. The Localized Scleroderma Cutaneous Assessment Tool (LoSCAT) consists of two parts and was used to objectively capture the patients' disease activity (the modified Localized Scleroderma Skin Severity Index; mLoSSI) and damage (Localized Scleroderma Damage Index; LoSDI) as per their treating physician. There is validity evidence for using the LoSCAT with these patients (Arkachaisri & Pino, 2008; Arkachaisri, Vilaiyuk, Li, et al., 2009; Arkachaisri, Vilaiyuk, Torok, & Medsger, 2010), and the mLoSSI was shown to be responsive to change in disease status (Kelsey & Torok, 2013). Also as part of the LoSCAT, physicians scored overall disease activity, damage, and severity using a 100-point global scale. Disease severity was of particular interest due to its potential association with HRQoL.

3.3.4 Analysis

Validity evidence for the survey and its proposed use were compiled using the framework provided by *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014).

3.3.4.1 Sample characteristics

Demographic information, including gender, race, age, LS subtype, and disease duration, were reported to and compared to published data on LS cohorts to ensure the representativeness of the sample to the general LS population.

3.3.4.2 Missing item analysis

Missing item analysis was performed to determine common patterns in missing/skipped items. For infrequently skipped/missed items (≤ 3 items per individual), imputation was done using the median value of that individual's score for other items in the respective theoretical domain. This allowed for the inclusion of the majority of subjects in analysis, ensuring the largest possible sample size.

3.3.4.3 Reading level

The reading level of the final self-report items and anchoring vignettes was reported to determine the appropriateness for 10-20 year olds. The Flesch-Kincaid Grade Level scale was used. This scale calculates the United States grade level of text based upon sentence length and word length (Flesch, 1948) and is used by the United States Department of Defense (DOD MIL-M-38784B).

3.3.4.4 Item level descriptive statistics

Frequencies and percentages for the 4-response levels were reported on the 21 items of the LoSQI so that distributions of responses could be examined. A skewed distribution with relatively high frequencies of '0' (not bothered) are expected, as this is commonly seen in pediatric LS cohorts (Baildam et al, 2011)

3.3.4.5 Internal structure

The internal structure of the survey was examined using exploratory factor analysis (EFA). Only items generated from the three main theoretical domains were used in the EFA model (items from the optional medication subscale were excluded). The items were dichotomized as ‘not bothered’ (0) or ‘at least a little bothered’ (scores of 1-3), due to the expected skewness of response patterns. The number of factors extracted was based on scree plots, high eigenvalues (>1.0), and factor interpretability (Costello & Osborne, 2005) and multiple solutions examined. With factor solutions that include more than 1 factor, oblique factor rotations were used if factors were correlated ($>.3$; Osborne, 2015). Items not loading strongly on any factor or loading significantly on multiple factors were considered for revision or elimination. The EFA was limited by the small sample size, and the repercussions of this are discussed in more detail in Section 5.3).

3.3.4.6 Descriptive Statistics for subscales and total scores

Based upon extracted factors and the final item set, relevant subscale and total scores were calculated and descriptive statistics reported for the entire sample (including central tendency, variability, and distributions of scores).

3.3.4.7 Reliability: Stability and Consistency

General reliability of the survey scores was computed using test-retest coefficients to compare the initial survey score to the duplicate score, for all duplicates returned within two weeks of initial administration. Coefficients were reported for all relevant subscale scores and the total score.

Cronbach's alpha coefficients were calculated and reported for the total score and subscale scores based upon factors extracted from the factor analysis solution. Stratified coefficient alpha for the total LoSQI score was also calculated because of the multiple theoretical domains included in the LoSQI (Rae, 2007). As a positively skewed distribution is expected among item responses, the items were also dichotomized, as in the factor analysis, and coefficient alpha recalculated. The effect of the skewed distributions on the coefficients is discussed in section 5.1.3.

Standardized errors of measurement (SEM) were calculated using both the test-retest reliability coefficients and Cronbach's alpha, for each subscale and total scores.

Item analysis was performed by examining the item-to-total score correlations and item-to-subscale score correlations, adjusted by removing each item from the score calculations.

3.3.4.8 Evidence in relation to other variables

Main convergent validity evidence was reported by correlating LoSQI scores with the most common survey currently used with LS patients, the CDLQI/DLQI. Ranges and other descriptive statistics were examined and compared between the LoSQI and CDLQI/DLQI. It was expected that the LoSQI scores would show a larger range of quality of life impact than the 10-item CDLQI/DLQI and the two scores would be moderately correlated. For the CDLQI/DLQI (as well as the CHAQ/HAQ), children and adults complete similar but different forms targeted specifically to their age range. For the CDLQI/DLQI, when completed by the same patients, scores on each form have found to be highly correlated, although scores on the adult version tend to be lower than on the child version (van Geel, Maatkamp, Oostveen, de Jong, et al, 2016). For this study, CDLQI and DLQI scores were examined separately. However, since correlations

were nearly identical between pediatric patients and adults, scores on either version of the survey were combined into one sample for analysis.

For the entire sample, LoSQI scores were compared with physician-scored assessments of disease activity, damage, and severity and parent-reported outcomes. Moderate to low correlations with the physician scored outcomes were expected based on results from previous research (Baildam et al., 2011) and conceptual differences in the content domains. The physician scored outcomes focus mainly on cutaneous features of LS, activity, and damage.

The relationship between the LoSQI scores and additional PRO's were examined. The subscale scores of the LoSQI were expected to correlate with similar domains from the PedsQL Rheumatology Module and total LoSQI scores were expected to correlate with the VAS-pt. The relationship between CHAQ/HAQ scores (measuring physical functioning) and LoSQI Domain 2 scores (physical functioning) were examined. For patients on systemic medications at the time of their clinic visit, LoSQI Domain 4 scores (Medication Side Effects) were correlated with total scores on the MISS.

3.3.4.9 Anchoring vignette analysis

Descriptive statistics for the anchoring vignettes were reported and the similarity of the rankings among respondents was discussed. Rank order within subjects was examined to determine the number of distinct rankings used by patients. Using the most basic statistical method, the B-scale, the vignette-adjusted total scores were compared to the raw scores. The B-scale was chosen over the C-scale due to a large frequency of ties in rankings (Wand, 2013). It also is important to note that there are no current recommendations for the sample size required for anchoring vignette models, although in the literature, sample sizes are usually large for the HOPIT (n 's > 2,000 in King et al., 2004 and King & Wand, 2007) and the latent-variable

approach ($n = 2965$ in Bolt et al. 2012). Only the non-parametric B-scale will be used with the field test data, as non-parametric statistics are often more robust with small sample sizes.

3.3.4.10 Software for analysis

For all basic analysis (i.e. descriptive analysis, correlations, reliability coefficients, etc.), SPSS 23.0 software (IBM Corp., 2015) was used. Exploratory factor analysis was performed using MPLUS (Muthén & Muthén, 2012). For anchoring vignette analysis, the ‘anchors’ package (Wand, King, & Lau, 2014) for R was employed (R Core Team, 2016).

3.4 SUMMARY OF METHODS

The purposes of this study were (1) to develop and provide validity evidence for the LoSQI, a disease-specific HRQoL survey for pediatric LS patients, and (2) to determine feasibility and usefulness of AVs in this population. The study procedures were divided into three separate stages; survey development, a qualitative pilot study, and a larger field-test. A summary of all study procedures is provided in Figure 5.

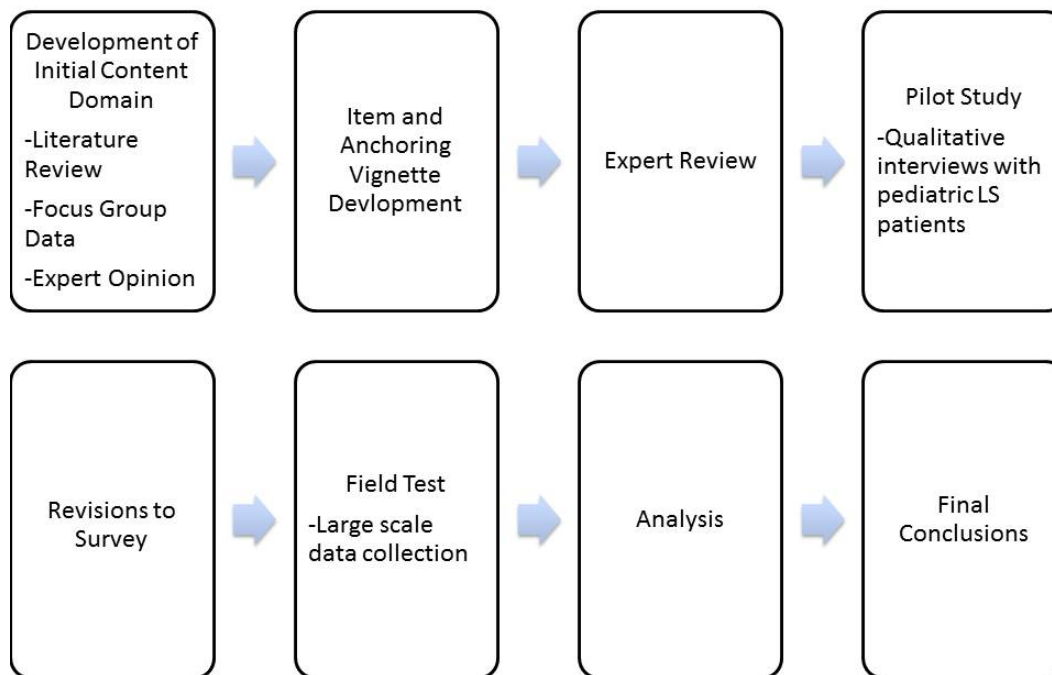


Figure 4. Summary of study procedures.

The content domain was developed based upon the published literature, data from past focus groups with pediatric LS patients, and expert opinions. Items were developed directly from the content domain and experts consulted before piloting. The pilot study collected information directly from pediatric LS stakeholders concerning the representativeness of the content domain, and basic issues like readability of items and the appropriateness of the recall period (see Table 10). The AV assumptions of response consistency and vignette equivalence were examined qualitatively. Once the survey was revised based upon the pilot study results, a field test gathered quantitative data from a larger sample of participants. A validity argument was built based upon internal consistency evidence, reliability coefficients, and convergent/divergent relationships

with other variables (Table 11). Responses to AVs were examined to determine if they imparted additional information regarding the content domain.

Table 10. Sources of data examined during the pilot study and field test.

Pilot Study	Field Test
<p>Localized Scleroderma Quality of Life Instrument (LoSQI)</p> <ul style="list-style-type: none"> • Time to survey completion • Understandability and readability of items • Adequate recall period • Content over- and under-representativeness 	<p>Localized Scleroderma Quality of Life Instrument (LoSQI)</p> <ul style="list-style-type: none"> • Descriptive statistics of items • Internal structure <ul style="list-style-type: none"> ○ Exploratory Factor Analysis • Reliability <ul style="list-style-type: none"> ○ Test-retest reliability ○ Internal consistency coefficients • Relationship with other variables <ul style="list-style-type: none"> ○ Other PRO's including the CLDQI ○ Parent-reported outcomes ○ Physician-reported outcomes
<p>Anchoring Vignettes (AVs)</p> <ul style="list-style-type: none"> • Time needed to complete AVs • Understandability and readability of AVs • Assumptions: <ul style="list-style-type: none"> ○ Vignette equivalence ○ Response consistency 	<p>Anchoring Vignettes (AVs)</p> <ul style="list-style-type: none"> • Descriptive statistics • Comparison of rankings over the sample • Adjusted scores

4.0 RESULTS

4.1 PILOT STUDY

4.1.1 Sample and administration

Participants were recruited and enrolled through the NRCOS, one of the largest registries of scleroderma patients in the country. Eighty-five percent of patients and their families were willing to participate (17/20). Of the three parent/patient dyads that chose not to participate, two reportedly were unable to stay after their clinic visit due to time constraints, and one parent did not think her younger boy would be able to comprehend the questions. The majority of the sample was female (13/17, 76%) and white (14/17, 82%), which is generally representative of LS (Murray & Laxer, 2002). Patient ages ranged from 8 to 18 years old; and since enrollment was stratified, there were 5-6 patients in each of the three age groups (8-10, 11-14, and 15-18 years old). All major subtypes were present in this sample with linear involvement being most common (10/17; 58.8%; Table 11). Twelve patients (70.6%) were on medication at the time of the study visit, and thus completed the LoSQI's optional medication subscale. Average disease duration for the sample was 4.6 years (time from diagnosis to the clinic visit) and ranged from less than one year to 9.4 years. The youngest age group had a shorter average disease duration when compared with the older age groups (Table 12).

Table 11. Frequency of LS classifications in the sample of patients enrolled in the pilot study.

LS Classification	n (%)
(1) Circumscribed morphea	3 (17.6)
(2) Linear scleroderma	
(a) Trunk/limbs	2 (11.8)
(b) Head	6 (35.3)
(3) Generalized morphea	2 (11.8)
(4) Pansclerotic morphea	1 (5.9)
(5) Mixed morphea	3 (17.6)

Table 12. Years from diagnosis of localized scleroderma to study visit in the pilot sample.

	<i>n</i>	<i>Mean</i>	<i>STD</i>	<i>Median</i>	<i>IQR</i>
8-10 year olds	5	3.07	3.0	2.06	1-6
11-14 year olds	6	4.12	3.0	3.18	2-7
15-18 year olds	6	6.39	3.8	8.31	2-9

On average, it took patients about 4.5 minutes to complete the LoSQI self-report items, including the medication subscale (*mean* = 4.5; range = 2-10 minutes; Table 13), and about a minute more to complete the anchoring vignettes (*mean* = 5.5; range = 4-8 minutes). Only five patients did not complete the medication subscale and their LoSQI times are reported separately in Table 13.

Table 13. Average time (in minutes) for patients to complete the quality of life instruments.

	Time to complete AVs		Time to complete LoSQI with Medication Subscale		Time to complete LoSQI without Medication* subscale	
	<i>n</i>	<i>Mean (STD)</i>	<i>n</i>	<i>Mean (STD)</i>	<i>n</i>	<i>Mean (STD)</i>
All	17	5.5 (1.59)	12	4.5 (2.47)	5	5.4 (2.9)
8-10 year olds	5	5.4 (2.07)	4	6.0 (4.43)	2	-
11-14 year olds	6	5.7 (1.37)	4	3.5 (1.00)	2	-
15-18 year olds	6	5.5 (1.64)	4	4.25 (2.63)	1	-

AVs = anchoring vignettes, LoSQI = Localized Scleroderma Quality of life Instrument, *due to very small numbers of patients, summary times to complete LoSQI without the medication subscale are not reported for each age group.

When times were broken down by age group (Table 13), the youngest group (8-10 year olds) was slower on average when completing the LoSQI items, yet faster or equal to their older counter parts at completing the AVs, which could indicate problems with understanding the vignettes (discussed further in Section 4.1.4). Overall, these times were faster than anticipated but not beyond expectations, as these patients are familiar with quality of life surveys through their enrollment in the NRCOS (which collects survey data at each clinic visit). However, this also led to some doubts emerging about patients truly reading the instructions. While watching them complete the form, some patients would start answering the questions so quickly that there was definitely not enough time for them to have thoroughly read the instructions. Again, this could be due to the fact that all these patients are very comfortable completing HRQoL surveys in clinic, but it could also be due to the potentially intimidating length of the instructions on both forms.

There were very few skipped/missing items or duplicate responses by the patients, probably due to the close supervision by the administrator. No items were skipped, but one

patient marked two answers for one AV scenario³. Final scores for all the domains and total scores for the LoSQI by age group can be found in Table 14 and Figure 6. Older patients (15-18 year olds) had the highest LoSQI total scores, indicating more HRQoL impact, with 11-14 year olds having the lowest (Table 14). However, 11-14 year olds had the highest median scores on the medication subscale when compared with the other two age groups (Table 14).

Table 14. Domain and total LoSQI scores for 17 patients enrolled in the pilot study.

	Skin Sensations	Physical Functioning	Body Image and Social Support	LoSQI Total*	Medication Subscale
	<i>Median (IQR)</i>	<i>Median (IQR)</i>	<i>Median (IQR)</i>	<i>Median (IQR)</i>	<i>Median (IQR)</i>
8-10 year olds	0 (0-1.5)	2 (0-3.5)	2 (0-4.5)	3 (2-8)	4.5 (2.25-6.75)
11-14 year olds	0 (0-0)	0 (0-.25)	0 (0-0)	0 (0-.25)	7 (2.5-10)
15-18 year olds	.5 (0-3.25)	6.5 (0-9.25)	3 (.75-8)	10.5 (2.25-25.25)	3 (1-9.5)
All patients	0 (0-1)	0 (0-5)	0 (0-3.5)	2 (0-8)	4.5 (2-9.25)

*LoSQI total score calculated by adding up Skin Sensation, MSQ Sequelae, and Body Image and Social Support subscale scores.

³ The anchoring vignette that was marked with two answers was completed by a subject in the youngest age group and is explored further in subsequent parts of section 4.1.

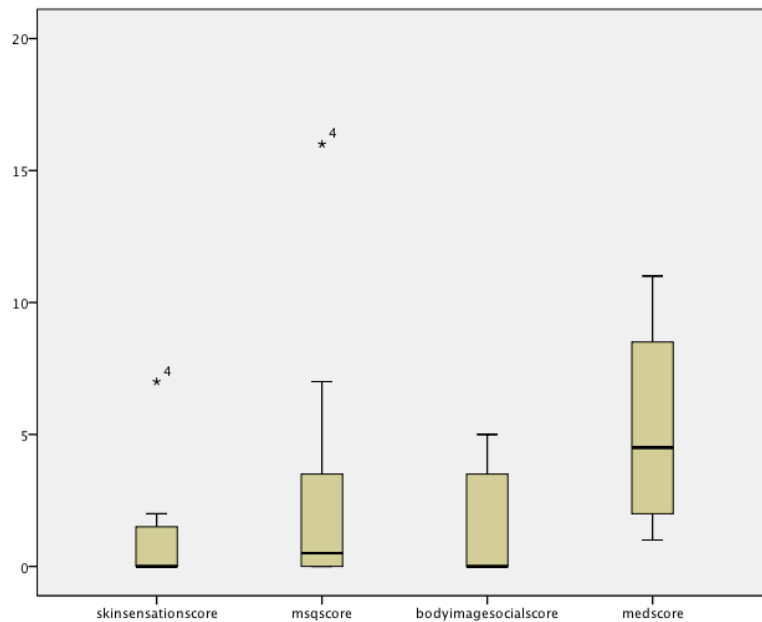


Figure 5. Distribution of domain scores for the LoSQI of 17 patients enrolled in the pilot study.

4.1.2 Pilot study results: 15-18 years old

The oldest age group included one male and five females ($n = 6$). Four of the patients were on medications at their clinic visit and completed the optional medication subscale.

4.1.2.1 LoSQI (15-18 yo)

Readability: The oldest age group had no problems reading the questions aloud when prompted. None of them stumbled significantly over any words or required help from the administrator or parent.

Questions about items: No patients in the oldest age group asked any questions or requested clarity about the LoSQI items or the corresponding instructions.

Understanding: Patients in this age group did not seem to have problems with understanding any LoSQI items from all three domains. When prompted, they were able to describe the item in their own words. For example, one patient viewed a skin sensation item as referring to a “burning” feeling (item: *painful skin where scleroderma is*; Skin Sensation Domain). Another patient said that she “pictured people mocking her because of her lesion” when I asked her to tell me what an item meant to her (item: *getting teased about the way I look because of my scleroderma*; Body Image and Social Support Domain). A male patient said that he thought about “being able to do stuff with his friends, like basketball” (item: *problems doing active things like running, playing sports, or dancing because of my scleroderma*; Physical Functioning Domain). In regards to another item (*worry about being able to do certain activities because of my scleroderma*; Physical Functioning Domain), a patient mentioned that she was particularly worried about sleepovers, as she had a very negative experience where her friends made her ‘feel like a zombie’ because of her disease. Most patients also responded that there was ‘nothing’ they would add to items to make the items more clear (one patient did have suggestions for improvement of items which are listed in the following sections).

Patients in this age group indicated that the response options provided were appropriate for the items. They were able to clearly articulate the response option they chose and why it was chosen. One patient mentioned that she experiences itchy skin, but it does not bother her because she is “used to it” (item: *itchy skin where my scleroderma is*; Skin Sensation Domain). While, another patient mentioned that she chose ‘bothers me a lot’, because she has stiff joints “all the time” (item: *stiff joints where my scleroderma is*; Physical Functioning Domain). Patients also were able to differentiate between symptoms that were and were not related to scleroderma. For example, one patient responded that she was ‘not bothered’ by problems using her hands (item:

problems using my hands when I write text, or type for a long time because of my scleroderma; Physical Functioning Domain). When asked about this answer, she said that sometimes she gets

tense in her hands and it bothers her, but that she does not believe it is related to her scleroderma.

Appropriateness of recall period: All patients in this age group indicated that if I asked them to think about the past 30 days, instead of the past 7, their answers would not have changed for all prompted items. However, for the patients with longer disease courses, all mentioned that their answers would have been very different if they were answering closer to when they were diagnosed or on medications like prednisone.

Under representation of the construct (missing elements): In order to determine content underrepresentation or missing items/domains, patients were asked, “Was there any important part of your life with LS that you feel was missing from the survey?” Half of the patients in this age group indicated that there was nothing else they would add about their life with LS, while the other three patients had specific concerns:

- One patient with wrist involvement was very distressed about the time LS took her out of school. All the doctor's appointments and occupational therapy after her initial diagnosis caused her grades to drop ‘a lot’ and she could not keep up with classwork. Her mom mentioned that she had to really advocate for her daughter at the school and let them know what was going on.
- A parent of another female patient with an *en coup de sabre* lesion mentioned that there were no questions about plastic surgery or the consideration of plastic surgery. She has found that her daughter was originally uninterested in this option, but as she had grown-up, she has started to consider it.

- A third female patient with a very visible *en coup de sabre* lesion wanted more open ended questions to be included in the survey. Particularly, she suggested that we ask about (1) how patients felt when they first got LS, (2) how life changes, (3) confidence and how that changes around other people, and (4) how do people act towards you about disease.
 - Interestingly, this patient also wrote in notes about some of her survey answers. On the AV page she wrote, “*I Do Not have Itchy skin*” (‘itchy skin’ underlined twice), and next to her answer for the ‘covering up my scleroderma’ item, she wrote “*very, very much*”. In her interview, she was also very descriptive in her answers and explained them in more detail than other patients in this age range.

Over representation of the construct (repetitive or duplicate items): Patients were asked, “did you feel like there were any questions that repeated too many times or asked the same thing?” to determine if there were any repetitive or duplicate items. Two patients mentioned potential repetitive LoSQI items. The first said that she considered bullying and teasing to be the same thing (both had negative connotations to her)⁴, and also she found ‘painful skin’ and ‘uncomfortably tight’ skin to be indistinguishable. When prompted about the ‘painful skin’ question, she said that she would clarify if the pain were more ‘sore’ or ‘stabbing’. The second patient mentioned that we asked a lot about ‘skin’. This was confusing to me at first, until I prompted further and realized that her disease damage was primarily related to hemi-facial atrophy. One side of her face was much smaller than the other (affecting both muscle and bone), and her actual skin involvement was minimal.

⁴ The male patient in this age group specifically mentioned that he was ‘not bullied but teased’ about his LS, although he said it didn’t bother him. My impression was that bullying had stronger negative connotations to him than teasing did.

Other Miscellaneous Feedback: One patient specifically confirmed that the two questions relating to problems with using hands (item 1: *problems using my hands when I write, text, or type because of my scleroderma*; item 2: *problems using my hands when I write, text, or type for a long time because of my scleroderma*) were different concepts. She stated that she typically uses her hands without issue but has significant problems after prolonged movements. Another patient mentioned that he would have liked five options for the item that asked if his localized scleroderma has improved since their last visit (3 options are provided: better/the same/worse).

4.1.2.2 Optional medication subscale (15-18 yo)

The results for the optional medication subscale were almost identical to the rest of the LoSQI items. However, one patient responded that she experienced ‘no’ side effects, but then had a score indicating relatively high rate of side effects compared with the pilot sample. When asked about specific symptoms, the patient mentioned that she was not sure they were related to her medications. Most of the ‘side effects’ she indicated could also be interpreted as general symptoms related to other conditions or even anxiety/depression; i.e. stomach pain, not feeling like eating, feeling in a fog, fatigue, and headaches.

4.1.2.3 AVs (15-18 yo)

Readability: The oldest age group had no problems reading the AVs aloud when prompted.

Questions about AVs: Only one patient asked a question to the administrator about completing the AV’s. The patient asked, “The ones that don't apply to me, do I just mark down 'do not bother?’” The administrator responded by reiterating the instructions and the patient had no other issues.

Understanding: There was no indication from the interviews that the oldest group had trouble understanding the AVs. All six patients were able to describe what the AV meant to them in their own words, articulate how their experiences were similar or different to the person in the AV, and logically report how they came to their answer.

4.1.3 Pilot study results: 11-14 years old

The middle age group included six patients (5 females and 1 male). One female patient finished the survey but had to leave before the interview could be completed. Thus, the validity of her answers was not verified and her results are not included in the qualitative analysis. Out of the five patients that were interviewed, 3 were on medication at the time of their visit and completed the medication subscale.

4.1.3.1 LoSQI (11-14 yo)

Readability: Patients did not have trouble reading the LoSQI items out loud. However, one patient was a much slower reader than the others (12 year old female). She struggled with the word ‘scleroderma’ and ‘retainers’, however, although her reading was prolonged, she was able to eventually read all words and her understanding of questions seemed acceptable.

Questions about items: Only one patient asked a specific question about the LoSQI items and it was in regards to the medication subscale. The patient was not currently on medications, and wanted to verify with the administrator she was stopping in the correct place.

Understanding: The patients in this group had very little reported HRQoL involvement (Table 14). Their answers related to understanding of the items were also a lot shorter than their older peers, but they were still able to indicate understanding. For example, one patient said that she

thought of “people making fun of me” while answering the item regarding teasing (item: *getting teased about the way I look because of my scleroderma*), while another patient mentioned that she was ‘not bothered’ by people asking questions about her skin but actually enjoyed answering them. A third patient reported that she interpreted an item as “not wanting to take medication because people might make fun of her” (item: *feeling embarrassed that I need to take medications*).

The rates of HRQoL involvement were very low in this age group and thus I was unable to prompt about their response options in as much detail as the oldest age group. However, when asked to explain their answers, most patients were able to report if the symptom had bothered them in the past or never had bothered them, providing further support that they understood the symptoms described by the items and were able to accurately choose the most applicable answer choice.

Appropriateness of recall period: All patients reported that their answer would not have changed if asked to recollect over the past 30 days instead of the past 7.

Under representation of the construct (missing elements): One patient mentioned that the LoSQI did not ask about problems or fears relating to blood work, lab draws, or skin biopsies, but these were bothersome to him. None of the other patients had suggestions for the LoSQI.

Over representation of the construct (repetitive or duplicate items): One female patient mentioned that she viewed teasing and bullying as being similar. No other patients indicated repetitive or duplicate items.

Other Miscellaneous Feedback: One patient mentioned that she thought back to the vignettes when considering her answer to an item (item: *problems doing fun things like painting or playing*

an instrument), indicating that the AVs might help to internally define the response scales for patients.

Another female patient responded very positively to all aspects of the survey tool, however, her mother mentioned that the patient tends to ‘aim to please’ and that things bother her at home much more than she lets on in clinic. For example, the mother said that the patient has mentioned her skin itches at night and requested a new bathing suit that covered more of her skin this year. Her remarks fit well with my impressions of the patient, whom I felt was trying to ‘play good’ and respond in the way she thought I would most like. The mother mentioned that it might be more accurate for her daughter to have a journal where she marks down when symptoms are bothering her in real time, so that she can better remember them during her visit to clinic.

4.1.3.2 Optional medication subscale (11-14 yo)

The results regarding the optional medication subscale were almost identical to the LoSQI items. However, it was again apparent that some patients were reporting symptoms that were unrelated to the medications. One patient indicated that she had headaches, but when probing further, her mother said that she did not believe they were fully related to the LS medications.

4.1.3.3 AVs (11-14 yo)

Readability: One patient was a significantly slow reader compared with the other participants in this age group (as mentioned above). She struggled with the names in the AVs (specifically Aliyah and Jasmine), but was able to read the rest of the questions aloud without assistance.

Questions about AVs: Only one patient in this age group asked for clarification regarding the AVs instructions. She asked, “So I pick how she would feel or how I would feel if I was her?”

Understanding: There was no indication from the interviews that this group had trouble understanding the AVs. All five patients were able to describe what the AV meant to them in their own words, articulate how their experiences were similar or different to the person in the AV, and logically report how they came to their answer. For example, one patient stated that she was more like the person described in the AV than different because people ask her about her disease, but it ‘doesn’t bother her’ (Maria/Daniel).

Content over-representativeness: None of the patients in this age group mentioned that the AVs were repetitive.

4.1.4 Pilot study results: 8-10 years old

The youngest sample included 5 patients; 3 females and 2 males. All but one of the patients were on medications. There was much more variability regarding reading level and understanding in this group. Particularly the male and youngest patients had many more questions regarding both the LoSQI and AVs, some trouble with reading certain words (including scleroderma), and indicated of a lack of understanding of the AVs. For the youngest patients (8-years-old), the interview was shortened; only half of the questions in each spiral were asked. In addition, some of the questions were modified to be more appropriate for this age group. For example, “can you tell me what question [1] means to you?” was not clearly understood by the younger participants. However, they were able to describe *how* they answered the question and *why* they chose that answer, or how the person described in the AV was alike or different from them.

4.1.4.1 LoSQI (8-10 yo)

Readability: The two 10-year-old girls in this sample had no trouble reading the LoSQI items, while the 8-year-old girl struggled with the word ‘scleroderma’. After asking for clarification, she replaced it with the word ‘skin’ and had no further troubles reading the items. The 8-year-old male asked his mother to read the questions aloud, which meant that mom also prompted a few of his answer choices. He had trouble with certain words like “aches”. Although the 10-year-old male also asked his mother to read the questions for him, it seemed to be related to fatigue as he had no trouble reading them aloud once prompted (his mother declined to do so).

Questions about items: Most patients in this age group asked at least one question, and some required additional help and support with answering the LoSQI items. The two 10-year-old girls were the most advanced, with only one patient asking if she had to complete the ID/date sections at the top of the form. The 8-year-old female patient asked for clarification of the word ‘scleroderma’ by asking, “does this mean my skin?” Once she received an affirmative answer, she had no other problems.

The youngest male said that the ‘instructions were confusing’, but could not clearly point to which part he did not understand. He also indicated that he could not comprehend certain questions (for example; item: *problems using my hands when I write, text, or type for a long time*). The other male patient had a lot of comments about anything that could relate to feminine gender roles; like dancing as an activity and wearing make-up (“boys don’t wear make-up”), although he did not ask specific questions that indicated a lack of understanding.

Understanding: Patients in this age group were mostly able to understand the LoSQI items. The two older females had no issue describing the concepts behind each LoSQI item; they both were asked about ‘feeling different’ because of their LS, and had very divergent responses (item:

feeling different than other people because of my scleroderma). One patient said that she does feel different but that she actually enjoys that feeling; while the other patient said that she feels a little different but it does not bother her too much (although she tries to keep her lesions hidden when possible). The youngest female was also able to articulate her understanding of the LoSQI. For one item, she mentioned that she has no problems doing active things because she “can move and it doesn’t hurt” (item: *problems doing active things like running, playing sports, or dancing because of my scleroderma*).

The older male patient was also able to articulate his understanding of the items and clearly describe how he filled out the LoSQI form. He said that he chose the response option ‘bothers me a little’ for one item because his skin around his lesion “feels weird” (item: *uncomfortably tight skin where my scleroderma is*). His linear lesion is also very visible when he wears shorts, so he indicated that he was a little embarrassed about how his body looks and also is bothered a little when people ask questions about his leg (items: *feeling embarrassed because of how my body looks; feeling upset when people ask questions about my scleroderma*). He told the administrator that this bothers him less now than it has in the past, as he performed well in his favorite sport this summer, which built up his confidence.

The youngest male patient was not able to clearly indicate his understanding of the items. For example, he said that he was unsure if his skin was itchy although he responded ‘bothers me a little’ to the LoSQI item (item: *itchy skin where my scleroderma is*). He also was very short with his answers, and it was unclear to the administrator if he was tired after his long visit to clinic (the interview took place almost three hours after his appointment began), shy, or was cognitively unable to understand the items.

Appropriateness of recall period: For the patients who indicated understanding of the items, all patients said their answers would stay the same if asked to reflect over the past 30 days instead of the past 7.

Under representation of the construct (missing elements): None of the five patients mentioned that an important part of their life with LS was missing from the survey.

Over representation of the construct (repetitive or duplicate items): One patient said that the 3 skin symptoms items were repetitive. The same patient reported that item: *problems using my hands when I write, text, or type because of my scleroderma* was the same as the item: *problems using my hands when I write, text, or type because of my scleroderma for a long time*.

Other Miscellaneous Feedback: For most patients, the interview took place after their normal visit. Although this could be tiring, the delay of the interview seemed to cause significant distress and respondent fatigue in the youngest patients. Particularly, the youngest male patient was seen 3 hours after his appointment time and his interview was the most disjointed and hard to perform. Both males in this group also indicated their (relatively good-natured) frustration with the length of the survey tool and the multiple questions being asked by the administrator.

4.1.4.2 Optional Medication Subscale (8-10 yo)

While there were no indicated problems with understanding of the medication subscale for the female patients, the youngest male indicated that he did not know what ‘medication side effects’ meant. His mom said that they do not usually discuss his medications, so that he has no clear vocabulary for that concept. He did mention that weight gain and headaches bothered him, but then also said he did not feel ‘bad’ after taking his medications, so there was also some disconnect between the interview and his responses, perhaps in what symptoms the patient linked to his

medications. This could mean that the temporal relationship between some of these side effects is not entirely clear to patients.

After probing into his answers more deeply, it was found that the older male patient was reporting symptoms that were unrelated to his medications. Specifically, he chose the response option ‘bothers me a lot’ regarding item: *feeling hungry all the time*. When the administrator asked if he thought this was related to his medications, he said, “I’m just growing.”

4.1.4.3 AVs (8-10 yo)

Readability: The two older females had no problems reading the AVs, while the younger female had trouble with the word ‘scleroderma’ (as described above). Contrarily, the younger male patient had a very hard time reading the AVs. When completing the form, he had his mother read him the instructions and requested that she also help by reading the vignettes. He had trouble with the second AV in particular (James), and had his mom read it aloud to him twice before answering. He also had trouble with the names on the AVs (i.e. ‘Caleb’). He mentioned that completing the form felt ‘like a test’; even though the researcher stressed multiple times that it was not a measure of his abilities.

The older male’s response to the survey administration was similar to the other male. He also asked if his mom could read him the AVs and had trouble with the James AV. After attempting to read it, he said, “I don’t understand this one. I know the words but it doesn’t make sense”.

Questions about AVs: One of the 10-year-old females asked if she needed to answer the AVs “how *she* would feel or how I would feel?” The 8-year-old female asked if each of the girls in the AVs was ‘different’, or if they were all the same person. She also asked what ‘medication’ meant, but seemed to understand ‘medicines’ better. The 8-year-old male had his mom read

some of the AVs twice, while the 10-year-old male had a number of questions. He asked, “Why does it say ‘Mike’s scleroderma’?” and mentioned “at least its all boys”.

Understanding: The two older female patients had no trouble indicating their understanding of the AVs. One patient reported that the AV regarding Kelly’s fear about gym class would bother Kelly ‘a lot’ because “she is afraid and doesn’t want to hurt”. The patient also was able to tell the administrator that this was very different from her experience, as the patient is not afraid of hitting her lesion. The other 10-year-old also indicated general understanding. For example, she said that unlike ‘Jessica’, who is bothered a lot by itchy skin, her skin doesn’t itch.

It was unclear if the youngest female understood all the concepts in the AVs. Although she was able to indicate general understanding for two vignettes, she had a hard time describing ‘Christina’s’ situation (vignette: *Christina does not usually worry about going to gym class because of her scleroderma. She is comfortable with what her body can and cannot do and does not mind taking breaks if she needs to*). After probing further, this could be either due to the length of this vignette or because the patient has experienced no functional limitations from her disease. This patient also had a duplicate answer for one of the vignettes.

The older male patient did not seem to understand the vignettes at all. He had a lot of questions about them while completing the section, marked them all as either ‘a little bit’ or ‘not at all’, and when asked could not put himself in the described person’s shoes. For example, the administrator asked him how he answered a specific vignette related to worry about gym class. The patient replied, “I don’t know.” When probing further, the administrator asked how the patient would feel if he had to skip gym class and the patient replied that he “never had to skip gym class” and then moved onto a tangent and said, “I don’t like it when my gym teacher won’t let me drink water during class.”

The 8-year-old male seemed to generally understand the AVs, as his answers were appropriate on the first page of the form and varied over all 4-response options. However, on the second page his answers were primarily at the two extreme ends of the response scale; ‘does not bother’ and ‘bothers him a lot’. This likely indicates respondent fatigue towards deciding between the 4 response levels, especially since the patient expressed his tiredness during the visit. This also could indicate a lack of understanding of the response scale, or a more extreme response style. When asking the patient to describe his answers in more detail, the patient could indicate how he was alike or different to the vignette character.

Content over-representativeness: Two patients said that the itchy AVs were repetitive, and one also mentioned that the survey talked a lot about ‘sleep trouble’.

4.1.5 Anchoring vignettes rank order and assumptions

In general, the two older cohorts of patients had no trouble indicating understanding of the anchoring vignettes. However, since the youngest group did indicate multiple problems with understanding and readability, the researchers determined that further examination of AVs should be limited to patients who are 11 years old and older. Thus, the ranked ordering of anchoring vignette results were examined for only 11-18 year olds ($n = 12$).

The first two sets of vignettes corresponding to the theoretical domains of skin sensations and physical functioning generally rank ordered as expected (Table 15). However, for the vignettes generated from the body image and social support domain, vignettes ‘Alexandra’ and ‘Katie’ were problematic with higher frequency of lower rankings. For example, ‘Katie’ was generated to correspond to the most extreme response option (‘3’ a lot bothered), but almost half

the patients ranked the vignette as a ‘1’ (a little bothered). Also, all vignettes from the medication side effects domain did not order as expected and were candidates for revision.

Table 15. Frequencies for responses to anchoring vignettes in the pilot study.

Skin sensations	Abby/Anthony n(%)	Madison/Mike n(%)	Jessica/Joseph n(%)	Jasmine/Jordan n(%)
Does not	8 (66.7)			
A little	4 (33.3)	9 (75.0)	2 (16.7)	1 (8.3)
Medium		3 (25.0)	2 (16.7)	
A lot			8 (66.7)	11 (91.7)
Physical Functioning	Christina/Christopher n(%)	Rachel/Jose n(%)	Jordan/Caleb n(%)	Kelly/James n(%)
Does not	8 (66.7)			
A little	2 (16.7)	7 (58.3)		1 (8.3)
Medium	2 (16.7)	4 (33.3)	7 (58.3)	4 (33.3)
A lot		1 (8.3)	5 (41.7)	7 (58.3)
Body image and social support	Maria/Daniel n(%)	Kayla/Brandon n(%)	Alexandra/Matthew n(%)	Katie/William n(%)
Does not	10 (83.3)	1 (8.3)	1 (8.3)	
A little	2 (16.7)	10 (83.3)	3 (25.0)	5 (41.7)
Medium		1 (8.3)	7 (58.3)	2 (16.7)
A lot			1 (8.3)	5 (41.7)
Medication side effects	Laila/Nathan n(%)	Sarah/Ethan n(%)	Aliyah/Tyler n(%)	Jennifer/Kevin n(%)
Does not	2 (16.7)	3 (25.0)		
A little	9 (75.0)	8 (66.7)		
Medium	1 (8.3)	1 (8.3)	3 (25.0)	4 (33.3)
A lot			9 (75.0)	8 (66.7)

Overall, there was no indication that the vignette assumptions were violated. Vignette equivalence holds if “the level of the variable represented in any one vignette is perceived by all respondents in the same way” (King et al., 2004), even if subjects chose different response

options. Different patients indicated similar considerations when answering the vignettes. For example, two patients in the oldest age group were asked about the same vignette (Alexandra/Matthew). One patient said that the vignette character was different than him because the vignette character is not bothered by questions about his skin, but the patient doesn't like to talk about it. A different patient said that she was similar to the character because does not get mad when people ask question; she knows how to word her answer to they understand it. Both patients indicated that they were using their own experiences to respond to the vignette. This interaction was similar to interview responses to other vignettes.

The response consistency assumption, which indicates if individuals use the vignette response options similarly to their own self-ranking, was determined based on interview questions linking the conceptual vignette to the self-report item response. For example, one patient indicated that the individual described in the vignette (*unwanted questions about their skin*, Katie/William) was different from her because the vignette person did not like telling people about her LS, but "I don't mind". For the same vignette set, another patient indicated that sometimes people ask about her skin, but she does not get upset about it. When asked why they responded the way they did, another patient said, "if I was her, I would be bothered". None of the responses indicated that patients were ranking the vignettes differently than the self-report items, even if they were not ranked the same.

4.1.6 Revisions based on pilot data

Based on the above-mentioned results, the following revisions were made to the LoSQI:

- The instructions were simplified as much as possible for clarity and to encourage patients to read them.

- Other names for localized scleroderma were added to the instructions. Specifically, Parry Romberg syndrome was mentioned so that patients with primary hemi-facial atrophy would feel more included.
- The low end of the targeted age range for the large scale data collection was changed to 10 years old instead of 8 due to lack of understanding of the anchoring vignettes by the youngest age group.
- Minor formatting changes were made to remove lines between items, add more white space, and synchronize formatting between the vignettes and self-report items.
- Both the AVs and self-report items were combined into one PDF for ease of administration, although 2 separate versions were created for children self-identifying as male or female.

4.1.6.1 LoSQI revisions

The following changes were made specifically to the LoSQI self-report items:

- Based upon specific suggestions by patients, three items were added for a total of 21 questions on the LoSQI, not including the medication subscale.
 - A fourth item was added to Domain 1 (skin sensations) that read, “my skin feeling different *where my scleroderma is*.” This item was included because it was noted that during the focus groups, some patients reported that their skin did not necessarily ‘hurt’ or ‘itch’, but felt ‘different’ at the site of their lesions.
 - Based on pilot test interviews, an additional item was included to measure the amount patients were bothered by time taken off of school for doctor’s appointments, therapy, and feeling sick (as related to LS); “having to miss school for doctor’s appointments, therapy, or feeling sick *because of my scleroderma*.”

- Also based on pilot interviews, an item was included to measure the distress/worry about blood draws/needle sticks as related to their LS; “getting my blood drawn *because of my scleroderma.*”
- A qualifying statement was also added to all the medication subscale items that remind patients to *only* consider symptoms specifically related to their medications.

Although debated, the bullying and teasing items were ultimately not combined. Although a few female patients thought these two items were repetitive, one male reportedly saw a clear difference between the two. The items related to skin symptoms were also not modified, despite one patient indicating that she would like more specifics related to ‘stabbing’ pain versus ‘soreness’.

4.1.6.2 Anchoring vignette revisions

Minor changes were made to the wording of six AV’s (Table 16); changes in text (i.e. revisions/deletions) are underlined. Most changes were prompted due to unexpected ordering of vignettes in the pilot study (see Section 4.1.5). However, the longest vignette in the survey was also shortened, as there was evidence for confusion; multiple patients asked for clarification.

Table 16. Anchoring vignettes used in the pilot study and their revised versions.

Domain	Order (0-3)	Reason for revision	Pilot Version of Vignette	Revised Version of Vignette
2	3	Confusion indicated in pilot interviews	Kelly worries about going to gym class because of her scleroderma. She is worried about a lot of things; not being able to do the activities that everyone else is doing, <u>accidentally hitting her skin and hurting herself, or having people ask her questions about why she's not participating.</u>	Kelly worries about going to gym class because of her scleroderma. She is <u>very</u> worried about not being able to do the activities that everyone else is doing and is afraid of accidentally hitting her skin and hurting herself.
3	3	Ordering issue	Katie does not like it when people ask her about her skin. It makes her feel sad and mad when they ask questions and it makes her uncomfortable to talk about it.	Katie does not like it when people ask her about her skin. It makes her feel sad and mad when they ask questions <u>and it makes her very upset.</u>
4	0	Ordering issue	Laila does not like her medications, but she usually feels fine after taking them.	Laila <u>usually feels fine after taking her medicines. On the days that she takes them, she can go to school and hang out with her friends like she always does.</u>
4	1	Ordering issue	Sarah feels nauseous after taking her medication(s), but when this happens she can still go to school and hang out with her friends like she usually does.	Sarah <u>sometimes feels sick to her stomach</u> after taking her medication(s), but when this happens she can still go to school and hang out with her friends like she usually does.

Table 16 continued

4	2	Ordering issue	After taking her medication(s), Aliyah feels nauseous and tired afterwards. It can be so bad that she has to miss school or skip fun activities on the weekend.	After taking her medication(s), Aliyah feels <u>sick to her stomach</u> and tired. <u>It can last awhile and sometimes</u> she has to miss school or skip fun activities on the weekend.
4	3	Ordering issue	Jennifer feels very tired and nauseous after taking her medication. She feels so sick on those days that she cannot go to school or hang out with her friends and family.	Jennifer <u>always</u> feels tired and sick to her stomach after taking her medication. She feels so sick on those days that she cannot go to school or hang out with her friends and family.

Theoretical domains: 1 = Skin sensations, 2 = physical functioning, 3 = Body image and social support, 4 = medication subscale. Ordering: 0 = does not bother, 1 = a little, 2 = a medium amount, 3 = a lot. Major changes in text are underlined.

4.1.7 Summary

Overall, the pilot study results supported the understanding and readability for the LoSQI items, the appropriateness of the recall period, and overall content representativeness for all age groups. The researcher saw very refined answers from the oldest age group, some potential issues with ‘playing good’ in the middle age group, and problems with understanding and respondent fatigue in the youngest age group (especially for the male patients). Three items were added to the LoSQI based upon patient suggestions. Understanding and readability of the anchoring vignettes was not totally clear for the youngest patients, and thus, the age range for large-scale data collection was modified. Wording of the anchoring vignettes was also changed based upon the rank ordering of items in the two oldest age groups. In all age groups, patients reported

symptoms on the medication subscale that were unrelated to the medications and thus, a qualifying statement was added to each of the medication items to indicate that their answers should be specifically related to their LS medications. Appropriate revisions were made to the LoSQI and anchoring vignette sections prior to the field test. The revised version of the LoSQI was 6 pages and can be found in Appendix D.

4.2 FIELD TEST

4.2.1 Field test sample

SPSS 23.0 software (IBM Corp., 2015) was used to examine the demographic characteristics of the sample. Both sites, UT Southwestern and Children's Hospital of Pittsburgh of UPMC, recruited patients for the study. Twenty-one patients were recruited from UT Southwestern (28%) and 53 patients from Children's Hospital of Pittsburgh of UPMC (72%). The majority of the sample identified as female (54; 73%) and white (67, 91%), which is generally representative of LS (Murray & Laxer, 2002). The patients ranged in age from 10 to 20 years old with an average of 16 years (SD = 3). All subtypes were represented in the sample, with the most common subtype being linear LS (Table 17). The frequency of linear subtype in this sample was generally representative of pediatric LS based upon a large, international study of 750 children (Zulian, Athreya, Nelson, Laxer, et al., 2006). Linear scleroderma of the head was more frequent in this sample than in other LS cohorts, most likely due to the specialized clinic located at the Pittsburgh site. Disease duration varied for the sample and ranged from less than one year to 13 years with an average of 5.6 years (SD = 3.5). Please note that in some cases, only the diagnosis

year was known. For these patients, the midpoint of the month/day was used for the disease duration calculation (i.e. if the month was unknown, June of that year was specified, if the date was unknown, the 15th of the month was specified).

Table 17. Rates of disease subtype in the study sample versus a large international sample of 750 children with LS.

LS Classification	Study sample n (%)	International Sample of 750 children ⁺ %
(1) Circumscribed morphea	9 (12)	Not reported
(2) Linear scleroderma*		
(a) Trunk/limbs	38 (53)	65
(b) Head	22 (31)	Not reported
(3) Generalized morphea	9 (12)	7
(4) Pansclerotic morphea	1 (1)	Not reported
(5) Mixed morphea	3 (4)	15

*Seven patients had linear scleroderma of both the trunk/limbs and head. ⁺Based on large multicenter study of 750 pediatric LS patients by Zulian, Athreya, Laxer, Nelson, et al., 2006

4.2.2 Missing item analysis

SPSS 23.0 software (IBM Corp., 2015) was used to examine the existence and pattern of missing/skipped data. Out of the 74 total patients who were administered the LoSQI, three had missing or skipped items (4%; NRCOS283, 2-005, 2-016). Two patients skipped or missed 1 item each, with the third patient skipping/missing 3 items. Forty-two patients completed the optional medication subscale, with only one patient skipping/missing one item. Scores for items that were missing were imputed using the median value of the individual scores for the respective theoretical domain.

In addition to the items generated from the theoretical domains, the LoSQI included three background items, which were more likely to be skipped. For example, the item that asked how the patient's scleroderma has been since their last visit to clinic was skipped 9 times (9/74, 12%). The same question on the medication subscale was skipped 6 times (6/42, 14%). The question that indicated if patients should continue with the medication subscale or stop (i.e. asks if they are on systemic medications), was skipped 7 times (7/74, 10%). Patients that skipped these questions were still included in analysis, as these items do not contribute to any of the intended scores.

4.2.3 Reading level

According to the Flesch-Kincaid Grade Level score, the reading level of the self-report items was 5.5 and 4.8 for the anchoring vignettes. This indicates appropriate wording for children in 5th or 6th grade (about 10-11 years old).

4.2.4 Item level descriptives

SPSS 23.0 software (IBM Corp., 2015) was used to obtain the item-level descriptives. The majority of patients reported that the symptom described in each item did 'not bother' them (Table 18). However, for all questions, at least some patients indicated they were bothered by the symptom. The most frequent items ranked as 'a little bothered' or greater were item 13 (feeling embarrassed about how my body looks, 40%), item 16 (feeling upset when people ask questions about my skin, 35%), and item 4 (my skin feeling different, 33%). Stiff and achy joints (items 11 & 12) also seemed to bother patients more frequently than some of the other symptoms (32%).

At their study visit, a little more than half the patients indicated their scleroderma was ‘the same’ as their last visit, with only 5 patients indicating their disease had gotten worse (Table 19).

Table 18. Item level frequencies for responses on the LoSQI (n = 74).

Num	Item Stem	Not bothered	A little	A medium amount	A lot
Domain 1: Skin sensations		n(%)	n(%)	n(%)	n(%)
1	Itchy skin	54 (73)	13 (18)	4 (5)	3 (4)
2	Painful skin	62 (84)	7 (10)	1 (1)	4 (5)
3	Uncomfortably tight skin	57 (77)	11 (15)	2 (3)	4 (5)
4	My skin feeling different	48 (65)	18 (24)	5 (7)	3 (4)
Domain 2: Physical Functioning		Not bothered	A little	A medium amount	A lot
5	Problems doing active things like running, playing sports, or dancing	57 (77)	8 (11)	3 (4)	6 (8)
6	Problems using my hands when I do things like write, text, or type	64 (87)	6 (8)	1 (1)	3 (4)
7	Problems using my hands when I do things like write, text, or type for a long time	62 (84)	4 (5)	4 (5)	4 (5)
8	Worry about being able to do certain activities because of my scleroderma	58 (78)	7 (10)	3 (4)	6 (8)
9	Problems when I am doing fun things like painting or playing an instrument	63 (85)	6 (8)	4 (5)	1 (1)
10	My muscles hurting	55 (74)	8 (11)	5 (7)	6 (8)
11	Aches in my joints	48 (65)	14 (19)	6 (8)	6 (8)
12	Stiff joints	49 (66)	16 (22)	5 (7)	4 (5)
Domain 3: Body image and social support		Not bothered	A little	A medium amount	A lot
13	Feeling embarrassed because of how my body looks	44 (60)	18 (24)	5 (7)	7 (10)
14	Feeling nervous when I am around new people who don't already know about my scleroderma	54 (73)	11 (15)	3 (4)	6 (8)
15	Feeling different than other people	51 (69)	12 (16)	4 (5)	7 (10)

16	Feeling upset when people ask questions	48 (65)	21 (28)	1 (1)	4 (5)
17	Covering up my scleroderma with things like...	55 (74)	8 (11)	4 (5)	7 (10)
18	Getting teased about the way I look	65 (88)	5 (7)	1 (1)	3 (4)
19	Being bullied about the way I look	68 (92)	3 (4)	-	3 (4)
	Additional items added after pilot	Not bothered	A little	A medium amount	A lot
20	Having to miss school for doctor's appts, therapy, or feeling sick	55 (74)	10 (14)	5 (7)	4 (5)
21	Getting my blood drawn	51 (69)	10 (14)	5 (7)	8 (11)

Table 19. Descriptives for additional question (n = 61).

Item Stem	Better n(%)	Same n(%)	Worse n(%)	NA n(%)
Since my last visit to the doctor, my scleroderma is...	19 (26)	40 (54)	5 (7)	1 (1)

Forty-two patients completed the medication subscale. The most bothersome items were 'feeling tired' (item 8), 'worry about medication side effects' (item 1), and stomach pain (item 5; Table 20). However, similarly to the items from the other domains, the majority of patients indicated that each symptom 'did not bother them'. This distribution was not surprising, as it is common for HRQoL measures in pediatric LS to be positively skewed (Baildam, Ennis, Foster, Shaw, Chieng, et al., 2011). Only 4 patients were bothered by 'feeling embarrassed about taking medications' (item 2), problems with vomiting (item 7), or 'feeling hungry all the time' (item 4). In addition, only two patients reported that their medication side effects had gotten worse since

their last visit, and most patients agreed/strongly agreed that their medications were helping them get better (Table 21).

Table 20. Descriptive statistics for Domain 4: Medication side effects (n = 42).

Num	Item Stem	Not bothered n(%)	A little n(%)	A medium amount n(%)	A lot n(%)
1	Worry about med side effects	29 (69)	10 (24)	1 (2)	2 (5)
2	Feeling embarrassed that I need to take medicines	37 (88)	3 (7)	2 (5)	-
3	Feeling sick right after I take my medicine	29 (71)	7 (17)	4 (10)	1 (2)
4	Feeling hungry all the time	37 (88)	2 (5)	1 (2)	2 (5)
5	My stomach hurting	29 (69)	9 (21)	2 (5)	2 (5)
6	Not feeling like eating	31 (74)	7 (17)	1 (2)	3 (7)
7	Vomiting	37 (88)	3 (7)	1 (2)	1 (2)
8	Feeling tired	27 (64)	7 (10)	4 (5)	4 (5)
9	Feeling like I am in a fog	33 (77)	6 (14)	2 (5)	1 (2)
10	Having headaches	31 (74)	5 (12)	3 (7)	3 (7)
11	Gaining weight	35 (83)	5 (12)	-	2 (5)

Table 21. Descriptives for additional questions related to Domain 4: medication side effects.

Item Stem	Better n(%)	Same n(%)	Worse n(%)	NA n(%)
Since my last visit to the doctors, my medication side effects are...(n = 36)	15 (42)	15 (42)	2 (6)	4 (11)
Item Stem	Strongly Agree	Agree	Disagree	Strongly Disagree
I believe that my medicine will help me get better. (n = 42)	22 (52)	16 (38)	4 (9)	-

4.2.5 Exploratory factor analysis (EFA)

An exploratory factor analysis was performed on the 21 items of the LoSQI using MPLUS (Muthén & Muthén, 2012). Item responses were dichotomized ('0' not bothered, '1' at least a little bothered) because of the skewed response distributions. Robust weighted least squares was used (WLSMV estimator). It is important to note that results from this analysis are exploratory in nature because of the small sample size ($n = 74$). Sample size recommendations for EFA vary greatly (Fabrigar et al, 1999), but researchers tend to agree that typically <100 subjects is not advised (Gorsuch, 1983), even when communalities are high and each common factor is overdetermined (>4 variables per factor; MacCallum et al., 1999). While not meant to be conclusive, the results of this factor analysis offer initial, although tentative, insight into the internal structure of the LoSQI.

Five extracted factors had eigenvalues above 1 (Table 22). Based upon the eigenvalues and the scree plot, (Figure 7) 1-, 2-, and 3-factor solutions were explored, and both the 1- and 2-factor solution reported. For the 2-factor solution, a geomin rotation was used because of

correlations between the factors, $r = .60$. The 3-factor solution was not easily interpretable and had negative residual variances, indicating potential over-factoring.

Table 22. Eigenvalues from the exploratory factor analysis of 21 LoSQI items ($n = 74$).

Factors	Eigenvalues	Ratio of Subsequent Eigenvalues
1	12.40	
2	2.68	.216
3	1.72	.642
4	1.32	.767
5	1.09	.826
6	0.91	.834

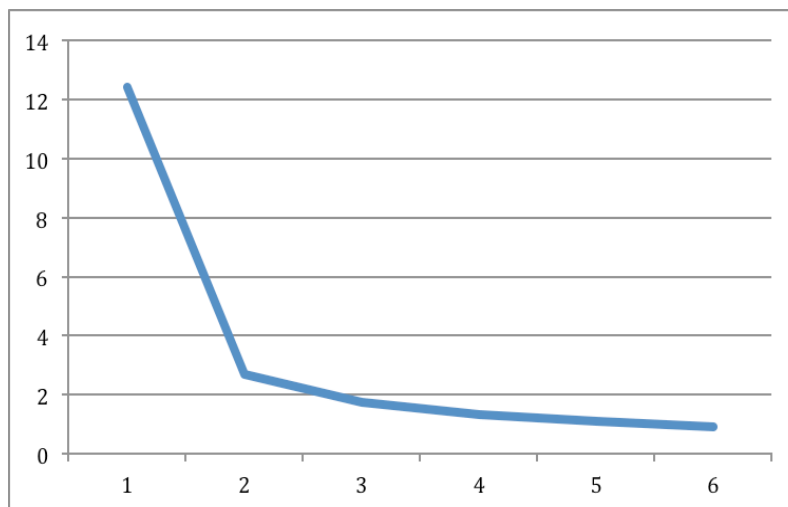


Figure 6. Scree plot for exploratory factor analysis of 21 items on the LoSQI ($n = 74$).

For the 1-factor solution, the first 20-items loaded saliently onto the factor and salient loadings ranged from .60 to .98 (Table 23). For the 2-factor solution, the first factor generally corresponded with items generated from the first two theoretical domains (skin sensations and physical functioning), as well as item 20 (*having to miss school for doctor's appts, therapy, or feeling sick*). Factor loadings ranged from .54 to 1.13 (Table 23). The second factor corresponded to theoretical domain 3 (body image and social support) and factor loadings ranged from .66 to 1.15. There were two loadings >1 for items 11 and 13. Factor loadings >1 are possible with correlated solutions, and are appropriate if the residual variances are positive (Joreskog, 1999; all residual variances were positive for both solutions). For both solutions, item 21 did not load strongly and was a candidate for deletion (loadings $<.4$, not shown in Table 23).

Table 23. Factor loadings for 1- and 2-factor solutions with oblimin rotations for 21 items of the LoSQL.

Num	Item Stem	1-Factor	2-Factor Solution	
		Solution	Factor 1	Factor 2
Domain 1: Skin sensations				
1	Itchy skin	.79	.83	-
2	Painful skin	.83	.84	-
3	Uncomfortably tight skin	.88	.79	-
4	My skin feeling different	.68	.54	-
Domain 2: Physical Functioning				
5	Problems doing active things like running, playing sports, or dancing	.83	.87	-
6	Problems using my hands when I do things like write, text, or type	.92	.77	-
7	Problems using my hands when I do things like write, text, or type for a long time	.91	.77	-
8	Worry about being able to do certain activities because of my scleroderma	.91	.91	-
9	Problems when I am doing fun things like painting or playing an instrument	.86	.93	-
10	My muscles hurting	.76	.79	-
11	Aches in my joints	.63	1.13	-
12	Stiff joints	.93	.91	-
Domain 3: Body image and social support				
13	Feeling embarrassed because of how my body looks	.98	-	1.15
14	Feeling nervous when I am around new people who don't already know about my scleroderma	.85	-	.98
15	Feeling different than other people	.80	-	.68
16	Feeling upset when people ask questions	.85	-	.80

Table 23 (continued).

17	Covering up my scleroderma with things like...	.87	-	.91
18	Getting teased about the way I look	.87	-	.71
19	Being bullied about the way I look	.82	-	.66
Additional items added after pilot				
20	Having to miss school for doctor's appts, therapy, or feeling sick	.60	.65	-
21	Getting my blood drawn	-	-	-

*factor loadings <.4 are suppressed.

Factor loadings were calculated again for all items after removing item 21 for both the 1-factor and 2-factor solution using MPLUS (Table 24). The final loadings were very similar to the 21-item solution. For the 2-factor solution, factors were again correlated ($r = .60$). The first factor contained 13 items and corresponded to the theoretical domains of skin sensations and physical functioning, and was termed the 'Pain and Physical Functioning Subscale'. The second factor contains 7 items and corresponded to the 3rd theoretical domain and was termed the 'Body Image and Social Support Subscale'. Loadings for items 11 & 13 were again >1.0 , and all residual variances were positive. Fit statistics for the 2-factor solution were slightly better than for the 1-factor solution; the chi-squared test for model fit was not significant and the RMSEA was < 0.05 (Table 25).

Table 24. Factor loadings from final solution of 20 LoSQI items (n = 74).

Num	Item Stem	1-Factor	2-Factor Solution	
		Solution	Factor 1	Factor 2
Domain 1: Skin sensations				
1	Itchy skin	.78	.83	-
2	Painful skin	.83	.86	-
3	Uncomfortably tight skin	.88	.80	-
4	My skin feeling different	.67	.52	-
Domain 2: Physical Functioning				
5	Problems doing active things like running, playing sports, or dancing	.83	.88	-
6	Problems using my hands when I do things like write, text, or type	.92	.76	-
7	Problems using my hands when I do things like write, text, or type for a long time	.91	.75	-
8	Worry about being able to do certain activities because of my scleroderma	.91	.91	-
9	Problems when I am doing fun things like painting or playing an instrument	.86	.95	-
10	My muscles hurting	.85	.80	-
11	Aches in my joints	.76	1.12	-
12	Stiff joints	.63	.91	-
Domain 3: Body image and social support				
13	Feeling embarrassed because of how my body looks	.93	-	1.14
14	Feeling nervous when I am around new people who don't already know about my scleroderma	.98	-	.98
15	Feeling different than other people	.85	-	.68
16	Feeling upset when people ask questions	.80	-	.81

Table 24 (continued)				
17	Covering up my scleroderma with things like...	.85	-	.91
18	Getting teased about the way I look	.86	-	.70
19	Being bullied about the way I look	.82	-	.66
Additional items added after pilot				
20	Having to miss school for doctor's apts, therapy, or feeling sick	.59	.63	-

Table 25. Fit statistics for the initial 1- and 2-factor solution of the 20-item LoSQI.

	One factor solution	Two factor solution
χ^2 test of model fit	254.25; $p < .001$	176.67; $p = .075$
RMSEA	0.082	0.048
CFI	0.963	0.989
TLI	0.958	0.986
SRMR	0.191	0.114

RMSEA = Root mean square error of approximation, CFI = comparative fit index, TFI = Tucker Lewis Index, SRMR = standardized root mean square residual.

4.2.5.1 Justification for score reporting

The factor analysis was exploratory due to the restrictions in analysis from the small sample size ($n = 74$) and skewed item distributions and subsequent dichotomization. Thus, both one and two factor solutions were reported. Based on the results of both solutions, the evidence for reporting a total score along with the two subscale scores was mixed. The dominant eigenvalue supported

the one factor solution and the use of a total score; the first eigenvalue was 4.6 times larger than the second. The fit statistics supported the two-factor solution over the one-factor solution, although, the correlation between factors was moderate ($r = .6$).

When scatterplots were created looking at the subscale scores and the total score, it looked as if the Pain and Physical Functioning Subscale had a slightly stronger relationship with the LoSQI total score when compared to the Body Image and Social Support Subscale (Figure 8). It also should be noted that the first subscale has more items ($i = 13$) than the second ($i = 7$). Validity evidence for both subscales and total scores will be examined in subsequent sections and the benefits and shortcomings of reporting subscale scores and along with a total score will be discussed more fully in Section 5.0.

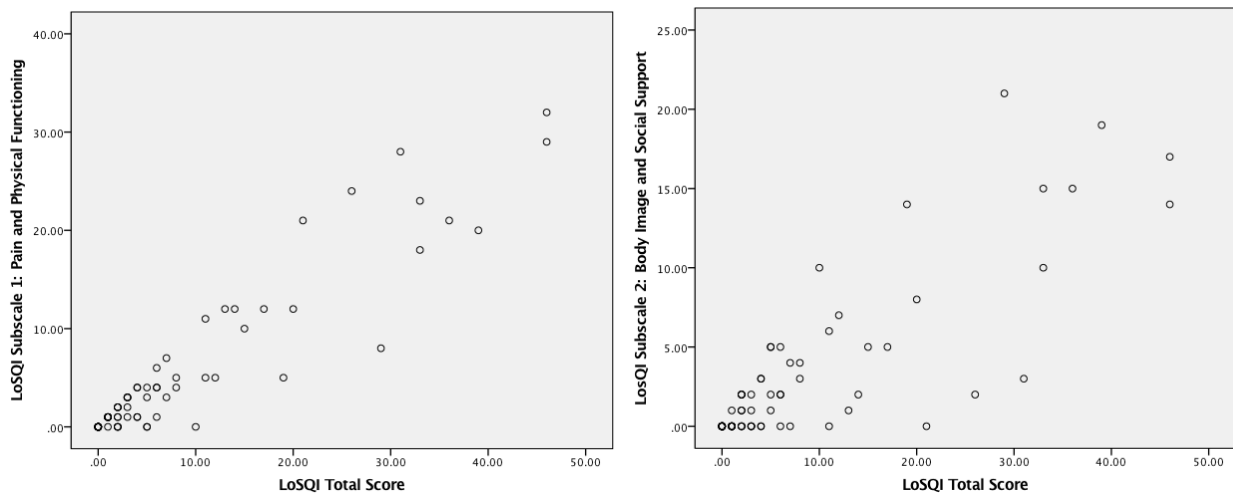


Figure 7. Scatterplots examining the relationship between the LoSQI total score and the subscale scores.

4.2.6 Subscale and total score descriptives

Distributions of scores for the two subscales and the total score were positively skewed with a relatively large frequency of the scores being 0. Not surprisingly, this resulted in SD larger than

the means for each score (Table 26). However, there was a relatively wide distribution of scores on the LoSQI with scores ranging up to 46 points (out of 63 possible points). There also was a small group of patients with very large subscale and LoSQI scores (>30, Figures 9 & 10). The medication domain followed a similar pattern with a smaller group of patients having higher scores than the rest of the group (> 10, Figure 11). This pattern has been seen in other published studies of HRQoL in pediatric LS (Baildam, Ennis, Foster, Shaw, Chieng, et al., 2011).

Table 26. Descriptive information for LoSQI total score and subscale scores (n = 74).

	Range	Mean	SD	Median	IQR	Score of 0 n(%)
Pain and Physical Functioning	0-31	5.2	7.9	1.5	5.25	32
Body Image And Social Support	0-21	3.0	5.0	1.0	4.0	49
LoSQI total score	0-46	8.3	11.6	3.0	10.0	23
Medication side effects (n = 42)	0-15	3.7	4.8	2.0	5.50	22

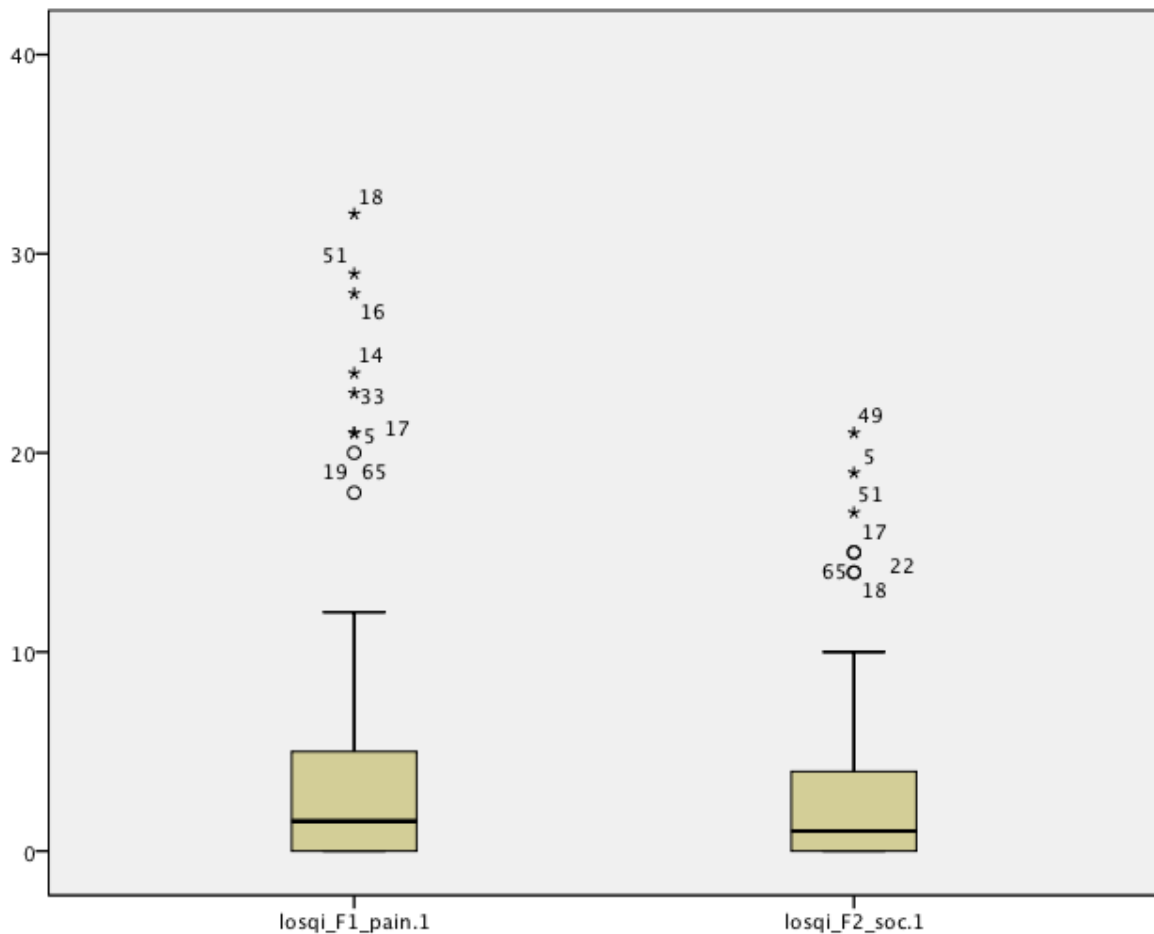


Figure 8. Boxplot showing distribution of scores on two LoSQI subscales (n = 74).

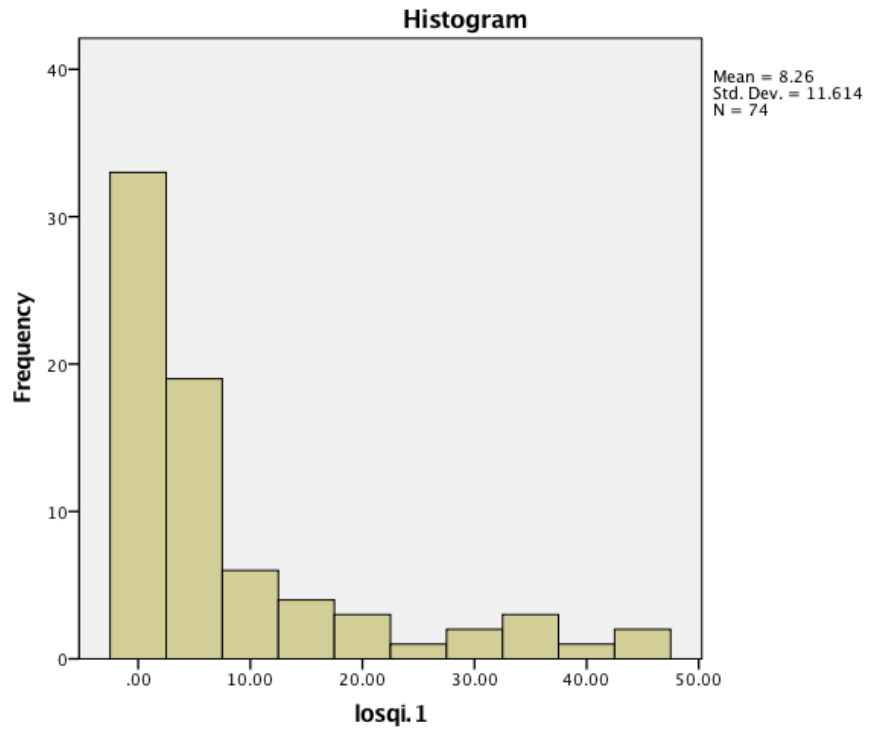


Figure 9. Distribution of total LoSQI scores (n = 74)

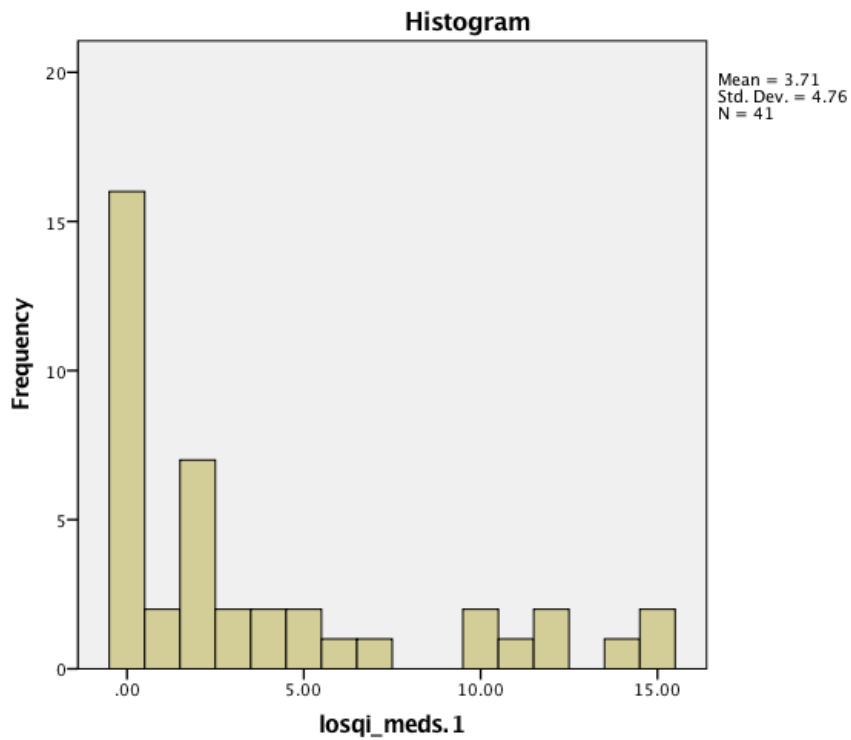


Figure 10. Distribution of scores on the medication subscale (n = 41).

4.2.6.1 Potential subgroups of interest

Differences in scores among potential subgroups and relationships to demographic variables were examined including age, disease duration, linear subtype, and facial involvement.

There was no relationship between age and LoSQI total scores (*Spearman's rho* = .008; Figure 12). There also was no relationship between disease duration and total LoSQI scores (*Spearman's rho* = .172, Figure 13).

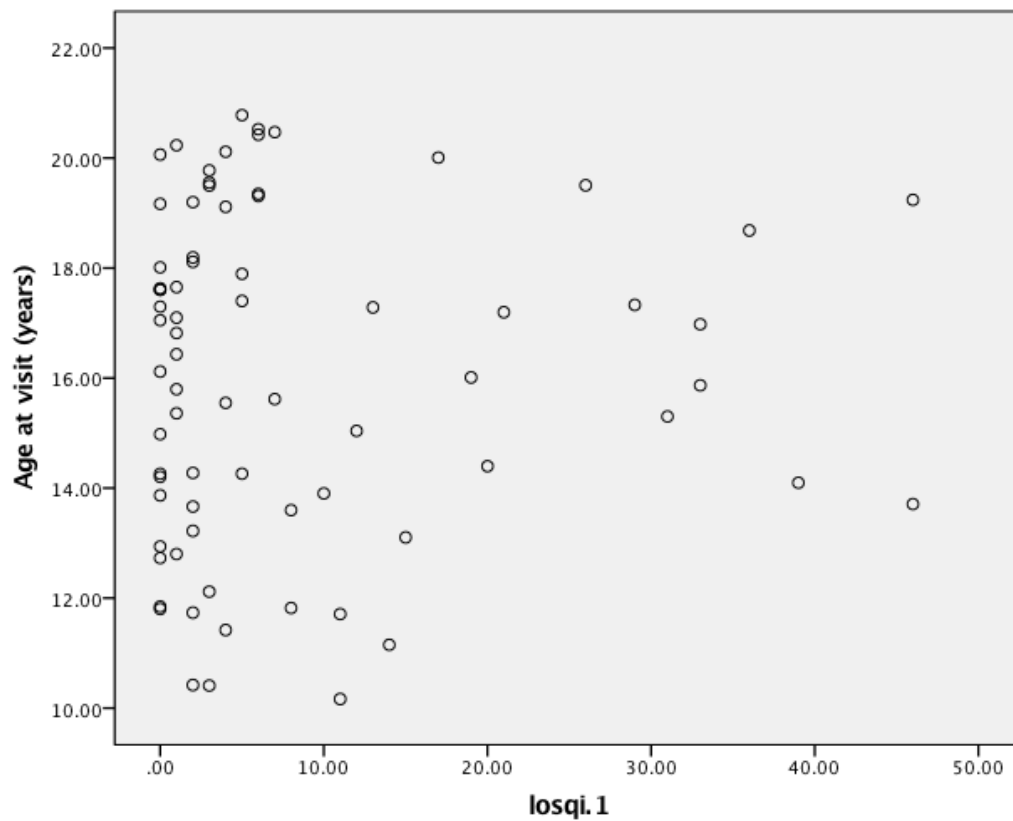


Figure 11. Scatterplot of relationship between age and LoSQI total score.

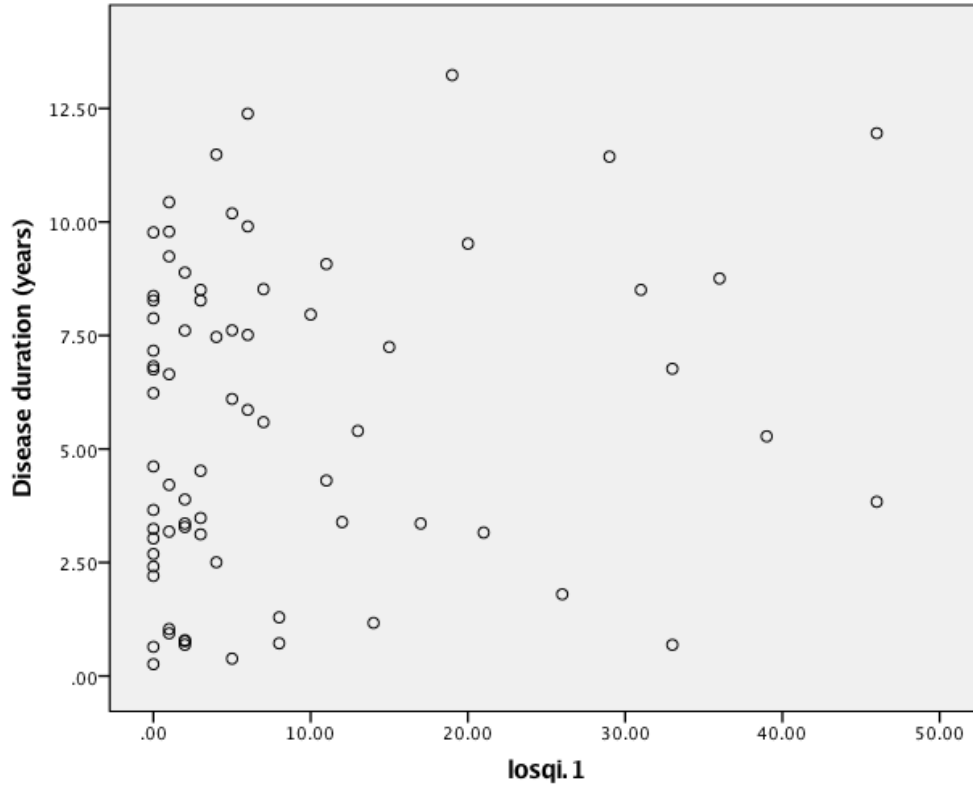


Figure 12. Scatterplot between disease duration in months and LoSQI scores.

There were significant differences in the distribution of LoSQI total scores between patients who had linear involvement of the trunk or limb and those that did not, *Mann-Whitney U* = 2.28, $p = .023$, with linear involvement leading to higher scores (median = 5, IQR = 15) than those without linear involvement (median = 2, IQR = 6.25). However, when looking at graphs of the distributions, these differences seemed less pronounced (Figure 14). In addition to the total score, there also were significant differences in the distribution of the Pain and Physical Functioning Subscale (Table 27), with linear involvement being associated with higher scores. This difference had the highest effect size ($r = .38$).

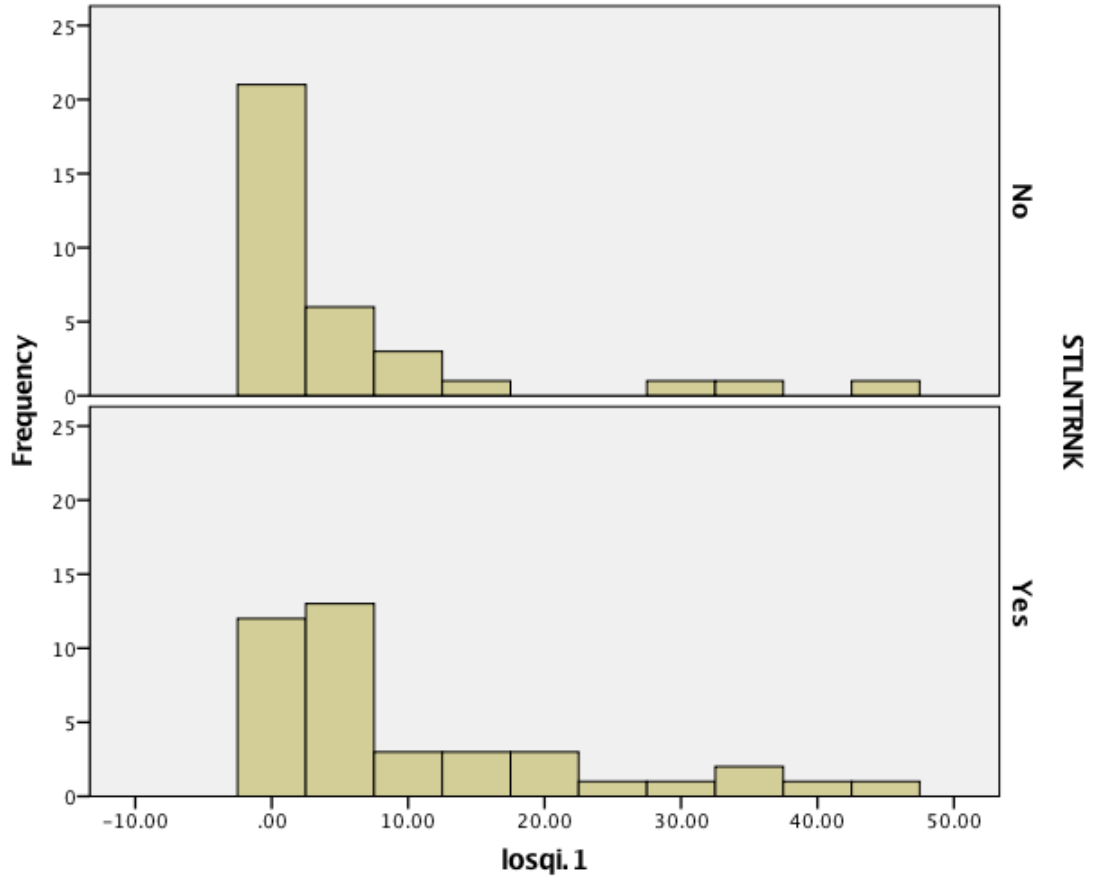


Figure 13. Distribution of LoSQI total scores for patients with and without linear involvement of the trunk/limb.

Table 27. Results of independent-samples Mann Whitney U tests comparing score distributions between patients with and without linear LS of the trunk/limb.

LoSQI score	Linear trunk/limb (n = 40)		Other subtype (n = 34)		<i>p</i>	Effect size*
	Median	IQR	Median	IQR		
Pain and Physical Functioning	4.0	11.0	0.5	3.0	.001	.38
Body image and social support	1.0	3.0	0.5	5.0	.913	.01
LoSQI Total Score	5.0	15.0	2.0	6.3	.023	.26
Medication side effects	3.5	9.0	1	2.5	.118	.24

*effect size is calculated as $r = Z / \sqrt{N}$;

There were no significant differences when comparing LoSQI subscale and total scores between LS patients with facial involvement and those without when using $\alpha = .05$ (Table 28). However, there were slight differences in scores on the Body Image and Social Support Subscale (effect size = .23; Figure 15).

Table 28. Results of independent-samples Mann Whitney U tests comparing score distributions between LS patients with and without facial involvement.

LoSQI scores	Facial Involvement (n = 23)		No Facial Involvement (n = 51)		<i>p</i>	Effect size
	Median	IQR	Median	IQR		
Pain and Physical Functioning	1.0	8.0	2.0	5.0	.527	.07
Body Image and Social Support	2.0	5.0	0.0	3.0	.052	.23
LoSQI Total Score	2.0	14.0	4.0	11.0	.874	.01
Medication side effects	2.0	9.0	2.0	4.8	.814	.04

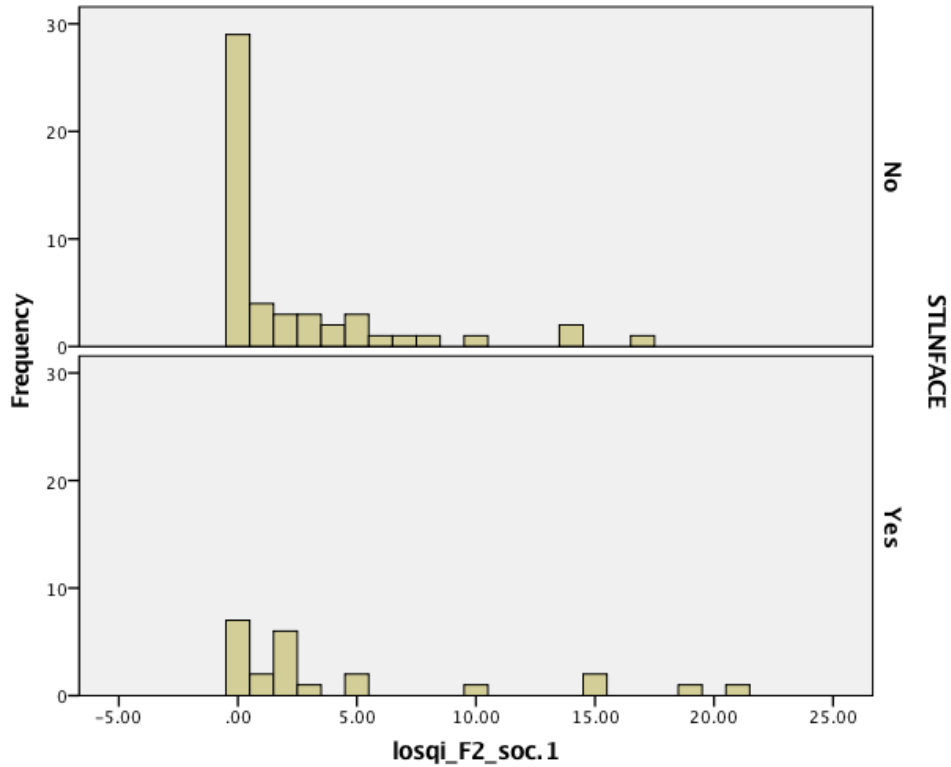


Figure 14. Distribution of the Body Image and Social Support Subscale of the LoSQI for patients with and without facial involvement.

4.2.7 Reliability

SPSS 23.0 software (IBM Corp., 2015) was used to evaluate the stability and consistency of the LoSQI scores.

4.2.7.1 Stability

Duplicate surveys were given to patients for completion within 2 weeks of initial administration. Thirty subjects returned the duplicate survey (30/74, 41%), of those, 20 completed the medication subscale (20/42, 48%). The average time between administrations of the survey was 4.8 days (SD = 4 days) and all duplicate surveys were returned within two weeks. Overall test-

retest reliability was highest for the medication subscale, but was $>.7$ for the total score (Table 29).

Table 29. Test-re-test reliability of the LoSQI (n = 30).

Score	<i>rho</i> *
Pain and Physical Functioning	.65
Body Image and Social Support	.66
LoSQI Total Score	.73
Medication side effects (n = 20)	.77

*Spearman's correlations were used because of the non-normal distribution of the data.

4.2.7.2 Consistency

Cronbach's alpha⁵ was calculated for each the two LoSQI subscales as well as the total score (Table 30). The alpha reflected high internal consistency for all three domains and the total score. The optional medication subscale was completed by approximately half of the sample, and had lower internal consistency than the other scales, although it was still above .8 (Table 30).

Table 30. Internal consistency coefficients for LoSQI domains and total score (n = 74).

	Number of items	Cronbach's alpha
Pain and Physical Functioning	13	.93
Body Image and Social Support	7	.93
LoSQI Total Score	20	.94
Domain 4: Medication side effects (n = 41)	11	.82

⁵ Please note that the effect of the skewed response distributions for items on Cronbach's alpha are discussed in more detail in Section 5.1.2.

Items were also dichotomized, as in the factor analysis, and coefficient alpha recalculated. Although slightly lower, the alpha values are comparable with those from the full response scale (Table 31).

Table 31. Cronbach's alpha for full response scales and dichotomized item responses.

	Number of items	Full response scale	Dichotomized response scale
		Cronbach's alpha	Cronbach's alpha
Pain and Physical Functioning	13	.93	.89
Body Image and Social Support	7	.93	.89
LoSQI Total Score	20	.94	.92

Stratified coefficient alpha was calculated based using equation 17, which incorporates the reliability and variances of each subscale.

$$r_{strat,\alpha} = 1 - \frac{\sum_i \sigma_i^2 (1-r_i)}{\sigma_c^2} \quad [17]$$

Where $r_{strata,a}$ is the reliability of the composite score, r_i is the reliability of stratum i , σ_i^2 is the variance of stratum i , and σ_c^2 is the variance of the composite scores. Stratified coefficient alpha for the total LoSQI score, using the two subscales, was high at $r_{strata,a} = .95$.

4.2.7.3 Standard error of measurement

Standard errors of measurement (SEM) were calculated for each score using both the test-retest reliability coefficients and Cronbach's alpha (Table 32). When using the test-retest coefficients, the SEM for the subscale scores were 5.00 for the Pain and Physical Functioning Subscale and

2.92 points for the Body Image and Social Support Subscale. When using Cronbach’s alpha, the SEM for the subscale scores were smaller, 2.09 and 1.33 points, respectively.

Table 32. Standard errors of measurement for LoSQI subscale and total scores using test-retest coefficients and Cronbach’s alpha.

	Total Points possible	SEM Test-retest	SEM Cronbach’s Alpha
Pain and Physical Functioning	39	5.00	2.09
Body Image and Social Support	24	2.92	1.33
LoSQI Total Score	60	6.03	2.84
Medication Side Effects	33	2.28	2.02

4.2.7.4 Item-total score correlations

Item-to-total score correlations were calculated for the two subscale scores and for the total LoSQI score. Scores were adjusted by removing the item in question from the calculation. Spearman’s rho was used due to the non-normal distribution of scores. All correlations were positive and showed discrimination above .4 (Table 33).

Table 33. Item-to-total score Spearman's correlations for the three LoSQI domains and the total LoSQI score.

Item Number	Pain and Physical Functioning	Body Image and Social Support	LoSQI
1	.59		.56
2	.54		.51
3	.64		.61
4	.51		.56
5	.65		.62
6	.42		.43
7	.47		.47
8	.67		.66
9	.54		.53
10	.67		.63
11	.68		.58
12	.55		.50
13		.82	.59
14		.76	.63
15		.68	.64
16		.69	.60
17		.73	.59
18		.53	.47
19		.46	.42
20	.46		.42

Item-to-total score correlations were also calculated for the Medication Side Effects Subscale using the same method described above (Table 34). Three items had correlations with the medication subscale score below .4: item 2: “Feeling embarrassed that I need to take medicines”, item 4: “Feeling hungry all the time *because of my medicine*” and item 11 “Gaining weight *because of my medicine.*”

Table 34. Item-to-total score correlations for LoSQI domain 4 (medication side effects).

Item	Medication Side Effects
1	.48
2	.31
3	.69
4	.12
5	.65
6	.57
7	.42
8	.64
9	.61
10	.58
11	.32

4.2.8 Relationships with other variables

SPSS 23.0 software (IBM Corp., 2015) was used to estimate correlations between the LoSQI scores and relevant outcome measures.

Sixty-one patients completed either the CDLQI or the DLQI and scores ranged from 0 to 21 points (*mean* = 2.7, *STD* = 4.5, *median* = 1, *IQR* = 3). The distribution of scores was positively skewed, with 41% of the respondents having a score of 0 ($n = 25$). The comparison between the LoSQI and the DLQI/CDLQI was of particular interest, as the DLQI/CDLQI is currently the most commonly used HRQoL PRO in pediatric LS. The CDLQI/DLQI score correlated strongly with the LoSQI total score ($rho = .81$; Figure 16), and more moderately with each subscale (Pain and Physical Functioning $rho = .71$, Body Image and Social Support $rho = .66$).⁶ However, the range of LoSQI scores was much wider compared with the range of DLQI/CDLQI scores (Figure 17).

⁶ Please note that the adult version (DLQI) and pediatric version (CDLQI) were examined separately for correlational analysis, and the correlations for the different versions of the survey were very close to the correlations when combined (i.e. CDLQI with total LoSQI = .830, DLQI with total LoSQI = .833, combined = .831). Thus, the CDLQI and DLQI scores are reported in combination.

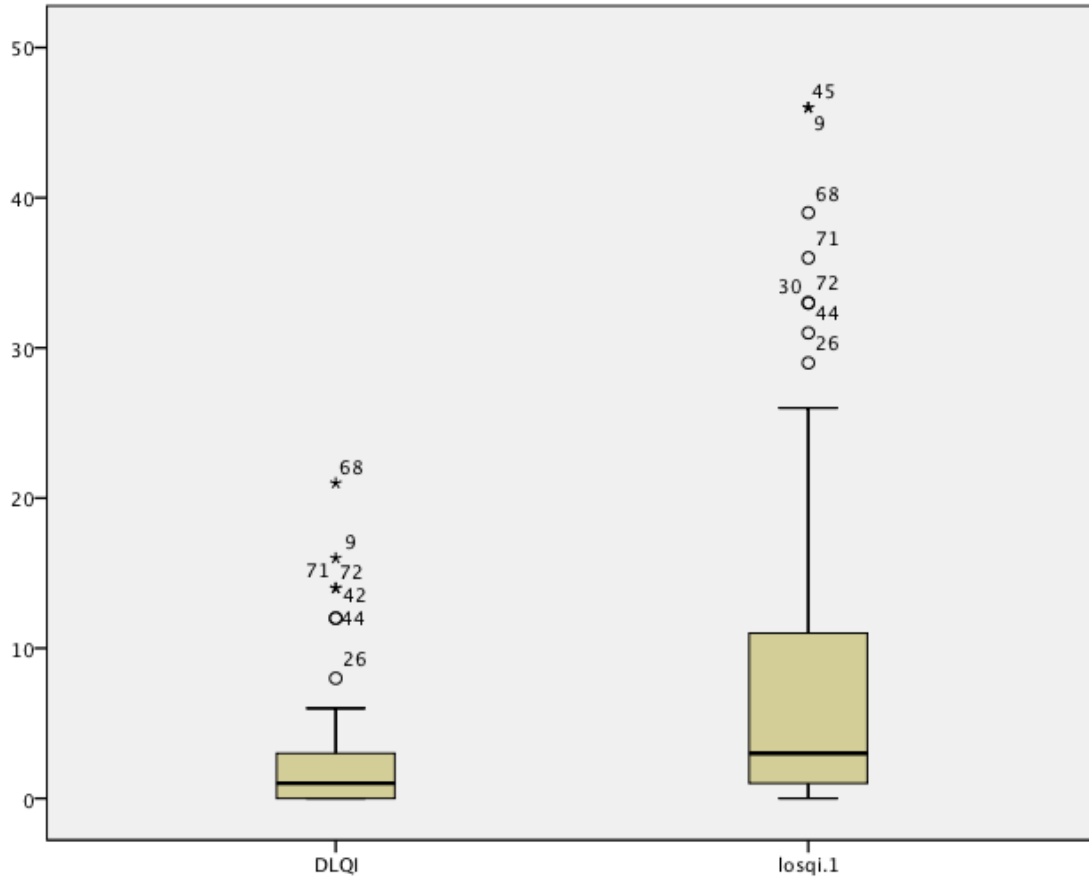


Figure 16. Distribution of scores on the DLQI/CDLQI and LoSQI for sample of pediatric LS patients.

Only a small subset of patients completed all of the outcome measures ($n = 24$, 32%). Based upon this sample, the LoSQI total score and subscales correlated moderately to strongly with the patient scored visual analogue scale (VAS-patient; Table 35). The LoSQI total score and subscales did not correlate with physician scored measures of activity (mLoSSI or PGA-activity). However, both the LoSQI total score and the CDLQI correlated significantly with the physician scored global assessments of disease damage and severity (PGA-damage and PGA-severity), although coefficients were low to moderate. The CDLQI had similar relationships with the patient and physician scored outcome measures as the LoSQI subscales (Table 35).

Table 35. Relationships between the LoSQI total score, the CDLQI, and other patient, parent, and physician reported outcomes for patients with complete data ($n = 24$; Spearman's rho).

	Subscale 1	Subscale 2	LoSQI	CDLQI
Patient Reported Outcomes				
CDLQI/DLQI	.77*	.73*	.90*	
VAS-patient	.74*	.55*	.72*	.62*
Physician Reported Outcomes				
mLoSSI (activity)	.31	.08	.23	.15
LoSDI (damage)	.21	.01	.21	.26
PGA-activity	.07	.03	.01	-.10
PGA-damage	.46*	.45*	.60*	.49*
PGA-Severity	.49*	.44*	.61*	.52*

Subscale 1 = Pain and Physical Functioning, Subscale 2 = Body Image and Social Support.

*significant at the .05 level.

Due to the small number of patients completing all outcome measures, the correlations were also examined using pair-wise deletion (Table 36). Overall, correlations between the LoSQI scores and the physician-scored outcome measures were fairly stable, and the pattern of significance for all correlations remained the same when compared with the smaller sample. However, the differences between samples were largest for the correlations between the LoSQI scores and the patient outcomes (Table 35 versus Table 36; differences $>.10$ highlighted in grey in Table 36). When including more patients, the correlations became smaller. For example, the correlation between the Pain and Physical Functioning Subscale and the VAS-pt, dropped from .74 to .58. For patients who were younger than 18 years old, their parent completed two HRQoL measures, the VAS-par and the PedsQL FIM. The LoSQI scores and the CDLQI correlated significantly with the parent-reported measures, although coefficients were low to moderate.

Table 36. Relationships between the LoSQI total score, the CDLQI, and other patient, parent, and physician reported outcomes for all patients with the outcome (Spearman’s rho).

	n	Subscale 1	Subscale 2	LoSQI	n	CDLQI
Patient Reported Outcomes						
CDLQI/DLQI	61	.71*	.66*	.81*		
VAS-patient	47	.58*	.38*	.54*	43	.58*
Physician Reported Outcomes						
mLoSSI (activity)	56	.17	-.15	.06	50	.01
LoSDI (damage)	56	.29	.17	.26	50	.18
PGA-activity	34	.05	-.02	-.00	29	-.06
PGA-damage	34	.43*	.44*	.54*	29	.45*
PGA-Severity	34	.43*	.42*	.54*	29	.47*
Parent Reported Outcomes (<18 y.o.)						
VAS-parent	34	.50*	.41*	.57*	31	.65*
Peds QL Family Impact	45	-.65*	-.37*	-.59*	40	-.67*

Subscale 1 = Pain and Physical Functioning, Subscale 2 = Body Image and Social Support.
*significant at the .05 level. Major changes in correlations (>.10) from Table 37 are highlighted in grey.

The LoSQI subscales correlated as expected with the corresponding subscales on the Peds QL Rheumatology Module (Table 37). All correlations were moderate and negative, as the Peds QL Rheum is scored in the opposite direction as the LoSQI; higher scores indicate less HRQoL impact. The Pain and Physical Functioning Subscale also correlated with the CHAQ (.48, $p = .12$, $n = 12$) and the HAQ (.79, $p = .01$, $n = 9$), although sample sizes were very low. In addition, the Medication Side Effects Subscale correlated moderately with total scores on the MISS (.58, $p < .01$, $n = 30$), indicating some overlap in content.

Table 37. Correlations between domain scores on the LoSQI and subscale scores on the Peds QL Rheumatology Module (n = 55).

LoSQI Subscales	Peds QL Rheum Subscales	Spearman's <i>rho</i>
Pain and Physical Functioning	Pain	-.82*
	Activity	-.56*
Body Image/Social Support	Communication	-.56*
	Worry	-.48*
Mediation side effects	Treatment	-.68*

*significant at alpha = .05.

4.2.9 Anchoring vignette results

4.2.9.1 Descriptives and missing values over entire sample

SPSS 23.0 software (IBM Corp., 2015) was used to evaluate the missing/skipped anchoring vignettes and report the response distributions.

One patient did not complete any of the vignettes but other missing and skipped items were limited to one vignette (vignette 5). The subject with missing values was excluded from analysis. Descriptive values for anchoring vignette responses are shown in Table 38. There was variability in vignette rankings, and most vignettes were scored over the entire range of rankings.

The two most 'bothersome' vignettes for domain 4 were scored similarly despite revisions to these vignettes after the pilot study. Please note that Table 38 is limited, as it does not show the ordering of vignettes *within each patient*.

Table 38. Descriptive statistics for the four vignette sets (n = 73).

Vignette	Does not bother	A little	A medium amount	A lot
	n(%)	n(%)	n(%)	n(%)
Theoretical domain 1: Skin sensations				
Self-report item: <i>itchy skin</i>	54 (73)	13 (18)	4 (5)	3 (4)
16. Abby's skin gets itchy where her scleroderma is, but she does not usually notice it.	24 (32)	44 (60)	3 (4)	2 (3)
1. When Madison's scleroderma itches, it feels better right after she scratches it. Sometimes, she feels annoyed that she has to scratch it.	11 (15)	41 (55)	20 (27)	1 (1)
11. Jessica's skin is itchy during the day, but it gets very itchy at night. On those nights, it makes it hard for her to fall asleep.	8 (11)	3(4)	24 (32)	38 (51)
6. Jasmine's skin itches almost every day and night. She scratches her skin most of the day and when it is time to go to bed, it keeps her awake for hours.	8 (11)	4 (5)	8 (10)	53 (72)
Theoretical domain 2: Physical functioning				
Self-report item: <i>Worry about being able to do certain activities because of my scleroderma</i>	58 (78)	7 (10)	3 (4)	6 (8)
4. Christina does not usually worry about going to gym class because of her scleroderma. She is comfortable with what her body can and cannot do and does not mind taking breaks if she needs to.	60 (81)	10 (14)	1 (1)	2 (3)
9. Rachel worries about going to gym class because of her scleroderma. She dislikes that she sometimes has to skip participating, but she is used to it at this point.	7 (10)	42 (57)	19 (26)	5 (7)
12. Jordan worries about going to gym class because of her scleroderma. She sometimes cannot do the same activities that her friends and classmates are doing and it makes her feel embarrassed to not participate.	6 (8)	11 (15)	27 (37)	29 (39)
2. Kelly worries about going to gym class because of her scleroderma. She is very worried about not being able to do the activities that everyone else is doing and is afraid of accidentally hitting her skin and hurting herself.	10 (14)	9 (12)	19 (26)	35 (47)

Table 38 (continued).

Theoretical domain 3: Body image and social support				
Self-report item: <i>Feeling upset when people ask questions</i>	48 (65)	21 (28)	1 (1)	4 (5)
5. Maria does not mind telling people about her skin and localized scleroderma. She knows how to explain it well, and understands that usually they are just curious.	57 (77)	8 (11)	3 (4)	4 (5)
13. Kayla does not like it when people ask questions about her skin, but she is used to explaining it so it does not upset her that much.	14 (20)	44 (60)	15 (20)	-
7. Alexandra does not like it when people ask questions about her skin. She tries to explain her scleroderma to them but she sometimes gets mad.	9 (12)	12 (16)	30 (41)	22 (30)
10. Katie does not like it when people ask her about her skin. It makes her feel sad and mad when they ask questions and it makes her very upset.	8 (11)	10 (14)	18 (24)	37 (50)
Theoretical domain 4: Medication side effects				
Self-report item: <i>Feeling sick right after I take my medicine</i>	29 (71)	7 (17)	4 (10)	1 (2)
8. Laila usually feels fine after taking her medicines. On the days that she takes them, she can go to school and hang out with her friends like she always does	61 (82)	8 (11)	2 (3)	2 (3)
15. Sarah sometimes feels sick to her stomach after taking her medication(s), but when this happens she can still go to school and hang out with her friends like she usually does.	10 (14)	37 (50)	24 (32)	2 (3)
3. After taking her medication(s), Aliyah feels sick to her stomach and tired. It can last awhile and sometimes she has to miss school or skip fun activities on the weekend.	6 (8)	6 (8)	16 (22)	45 (61)
14. Jennifer always feels tired and sick to her stomach after taking her medication. She feels so sick on those days that she cannot go to school or hang out with her friends and family.	7 (10)	1 (1)	19 (26)	46 (62)

4.2.9.2 Rank ordering

The ‘anchors’ package (Wand, King, & Lau, 2014) for R (R Core Team, 2016) was employed to determine the rank ordering of vignettes for this sample. As the vignettes were created using a target item from the four theoretical domains, the analysis of vignette sets follow that pattern. The ordering of vignettes within each patient for each vignette set was first examined independent of the self-response items. Vignettes within each set were designed to roughly correspond to the 4-point response scale. For the subsequent sections, numerals 1-4 refer to the four vignettes within each set; with ‘1’ indicating that the was designed to represent ‘not bothered’, ‘2’ as ‘bothered a little’, ‘3’ as ‘bothered a medium amount’, and ‘4’ as ‘bothered a lot’. The patients’ responses to the vignettes will be called ‘the ranking’ in the following sections.

Theoretical domain 1: Skin sensations

Sixty-seven patients responded with at least two distinct vignette rankings and 55 had no violations of the intended order. A majority of patients gave the vignettes in this set the same ranking (Table 39; Figure 18), termed ‘ties’. Only 2 individuals (n = 73) ranked all four vignettes as designed. However, despite the ties, most patients generally ranked the vignettes in order (i.e. 4 higher than 3, 2 higher than 1, see Table 40). Vignettes 1 and 2 ranked lower than vignettes 3 and 4 for 70-85% of the sample. However, there were still some problematic rankings, as vignette 3 ranked below vignette 4 only 27% of the time.

Table 39. Most common ranking patterns for vignettes in set 1 (skin sensations).

Pattern	Frequency	Proportion	Number of distinct rankings
{1,2},{3,4}	23	0.31	2
1,2,{3,4}	13	0.18	3
1,{2,3},4	8	0.11	3
{1,2,3,4}	7	0.09	1
{1,2},3,4	7	0.09	3
{1,2},4,3	2	0.03	3
1,2,3,4	2	0.03	4

1 = vignette 16, 2 = vignette 1, 3 = vignette 11, and 4 = vignette 6.

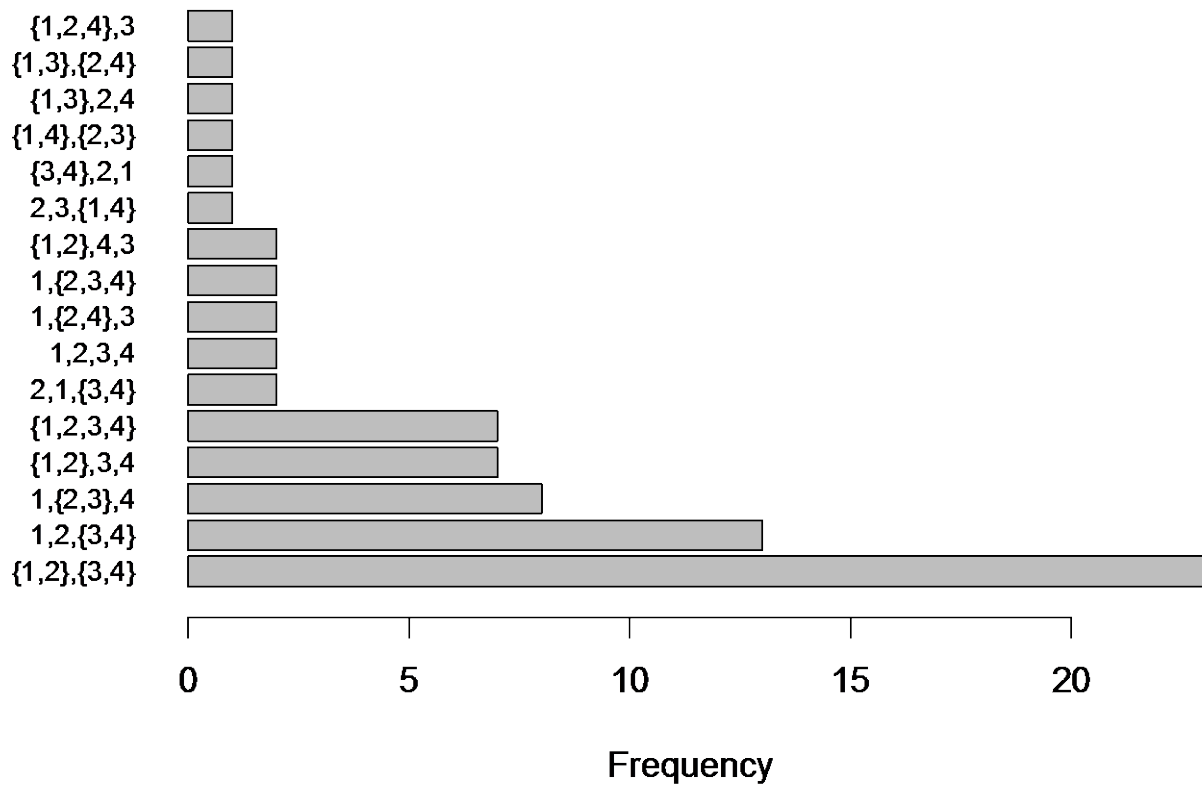


Figure 17. Ranking patterns for vignettes in set 1 (skin sensations).

Table 40. Proportion of cases that a vignette was ranked less than another vignette for set 1 (skin sensations).

Intended order	Vignette #	<1	<2	<3	<4
1	16	-	.40	.85	.85
2	1	.05	-	.72	.80
3	11	.03	.04	-	.27
4	6	.01	.03	.08	-

Theoretical domain 2: Physical functioning

For theoretical domain 2, physical functioning, sixty-six subjects had at least two distinct vignette rankings, and 45 had no violations of the intended order. There were also a large number of ties in the data (Table 41). Interestingly, vignette 3 ranked more ‘bothersome’ than vignette 4 for about 23% of the subjects (Figure 19; Table 42).

Table 41. Most common ranking patterns for vignettes in set 2 (physical functioning).

Pattern	Frequency	Proportion	Number of distinct rankings
1,2,{3,4}	23	.31	3
{1,2,3,4}	8	.11	1
1,2,3,4	7	.09	4
1,{2,3,4}	6	.08	2
1,{2,4},3	6	.08	3
1,{2,3},4	5	.07	3
1,2,4,3	4	.05	4

1 = vignette 4, 2 = vignette 9, 3 = vignette 12, and 4 = vignette 2.

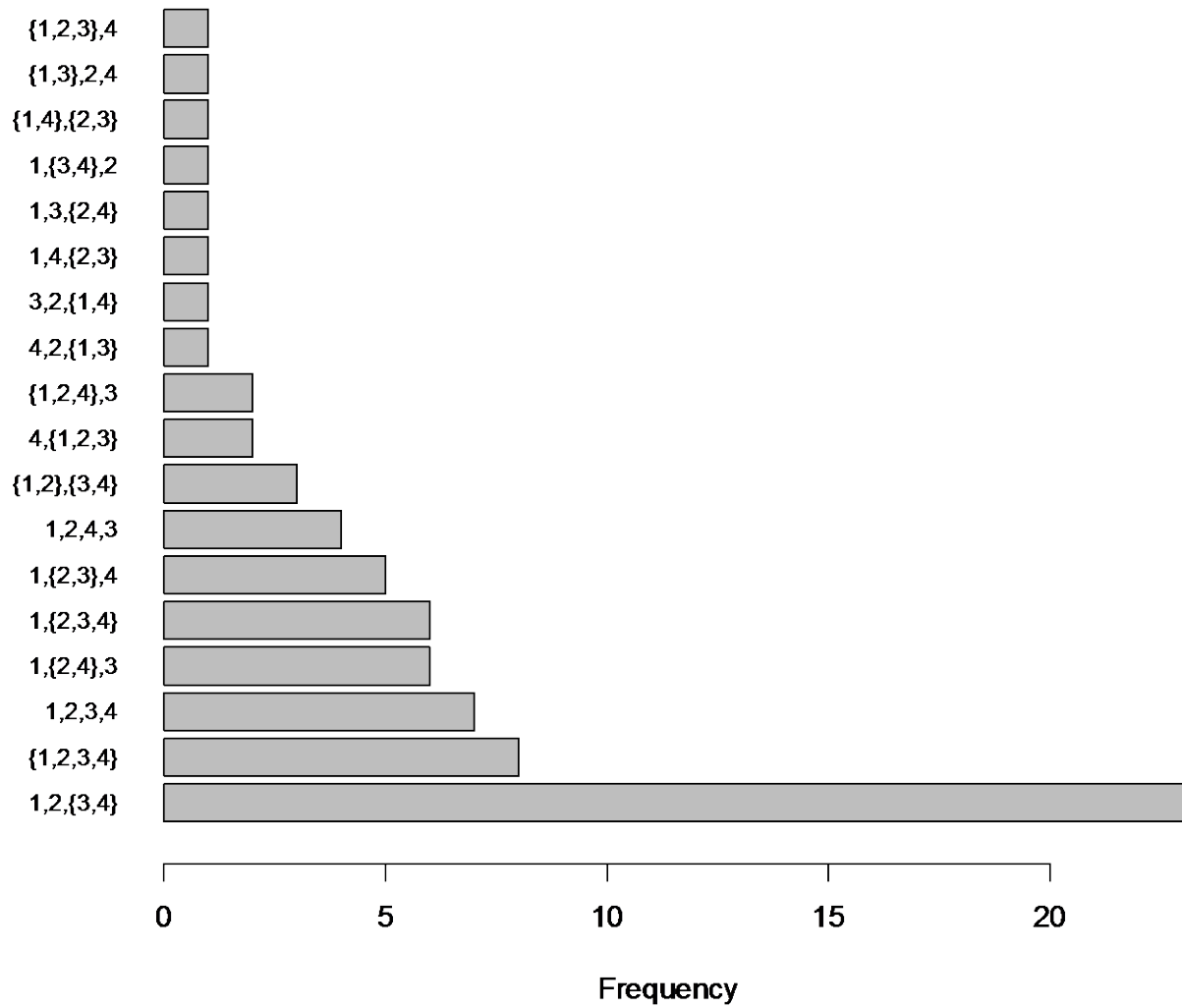


Figure 18. Ranking patterns for vignettes in set 2 (physical functioning).

Table 42. Proportion of cases that a vignette was ranked less than another vignette for set 2 (physical functioning).

Intended order	Vignette #	<1	<2	<3	<4
1	4	-	.76	.81	.80
2	9	.03	-	.62	.61
3	12	.01	.05	-	.22
4	2	.04	.08	.23	-

Theoretical domain 3: Body image and social support

For theoretical domain 3, body image and social support, sixty-three subjects responded with at least two distinct vignette rankings and 48 had no violations of the intended ordering. Vignettes in this set ordered relatively well with 12 subjects ranking them as designed (Table 43; Figure 20), although there were still significant ties. Vignette 4 ranked above vignettes 3 in almost 40% of subjects (Table 44).

Table 43. Most common ranking patterns for vignettes in set 3 (body image and social support).

Pattern	Frequency	Proportion	Number of distinct rankings
1,2,{3,4}	19	.26	3
1,2,3,4	12	.17	4
{1,2,3,4}	9	.13	1
1,{2,3},4	9	.13	3
{1,2},{3,4}	3	.04	2
{1,2},3,4	2	.03	3
{1,4},{2,3}	2	.03	2

1 = vignette 5, 2 = vignette 13, 3 = vignette 7, and 4 = vignette 10.

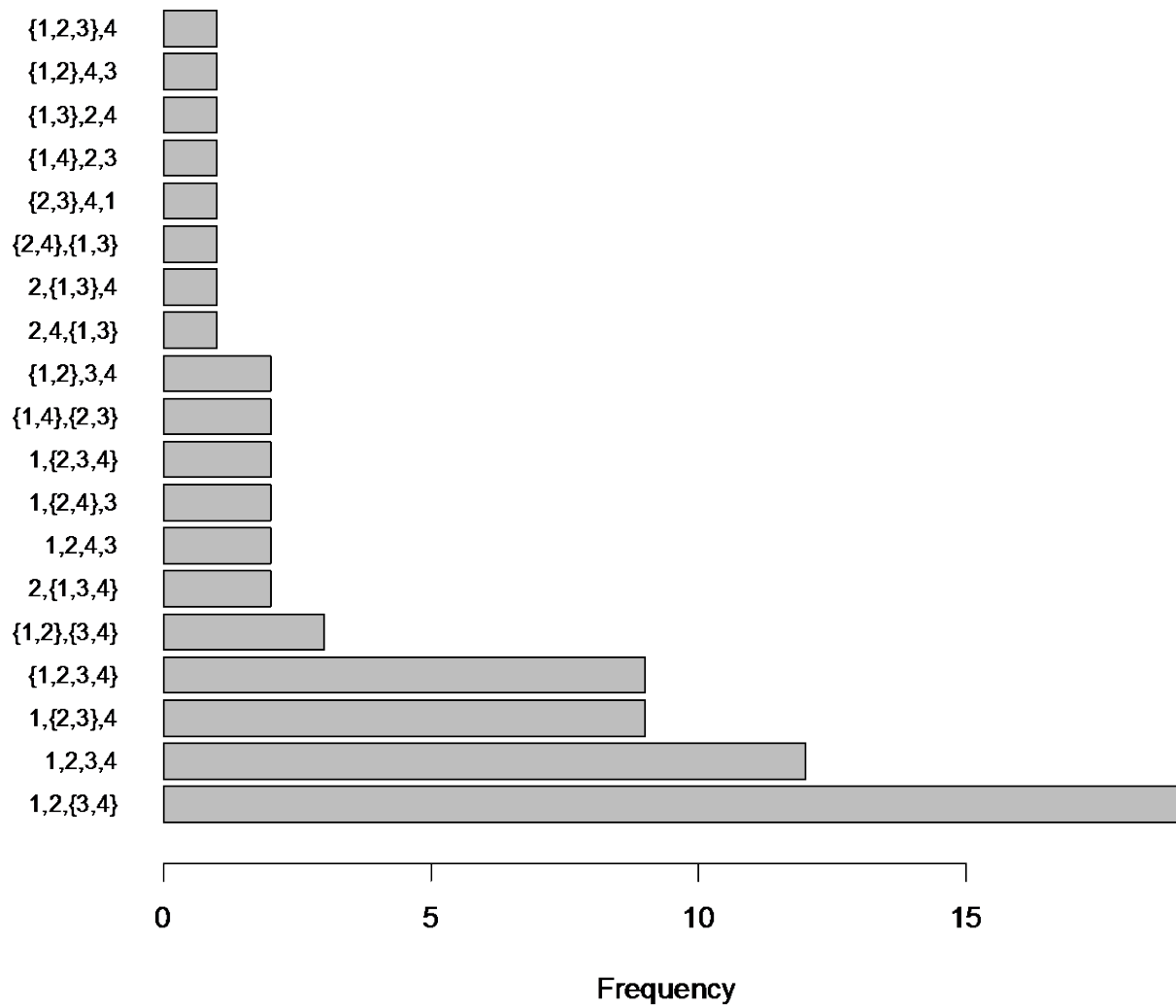


Figure 19. Ranking patterns for vignettes in set 3 (body image and social support).

Table 44. Proportion of cases that a vignette was ranked less than another vignette for set 3 (body image and social support).

Intended order	Vignette #	<1	<2	<3	<4
1	5	-	.69	.76	.76
2	13	.08	-	.65	.76
3	7	.01	.01	-	.38
4	10	.04	.04	.14	-

Theoretical domain 4: Medication side effects

For theoretical domain 4, sixty-eight cases had at least 2 distinct vignette responses and 54 had no violations of intended ordering. There were still a large number of ties in the rankings; with half the sample ranking the two ‘most bothersome’ vignettes the same (Table 45, Figure 21). When vignettes were not tied, patients still seemed to have trouble discerning between vignette 4 and vignette 3, as vignette 3 was only ranked below in 16% of cases (Table 46).

Table 45. Most common ranking patterns for vignettes in set 4 (medication side effects).

Pattern	Frequency	Proportion	Number of distinct rankings
1,2,{3,4}	37	.50	3
{1,2},{3,4}	7	.09	2
{1,2,3,4}	6	.08	1
1,2,4,3	5	.07	4
1,{2,3},4	4	.05	3
1,2,3,4	4	.05	4
1,{2,3,4}	2	.03	2

1 = vignette 8, 2 = vignette 15, 3 = vignette 3, and 4 = vignette 14.

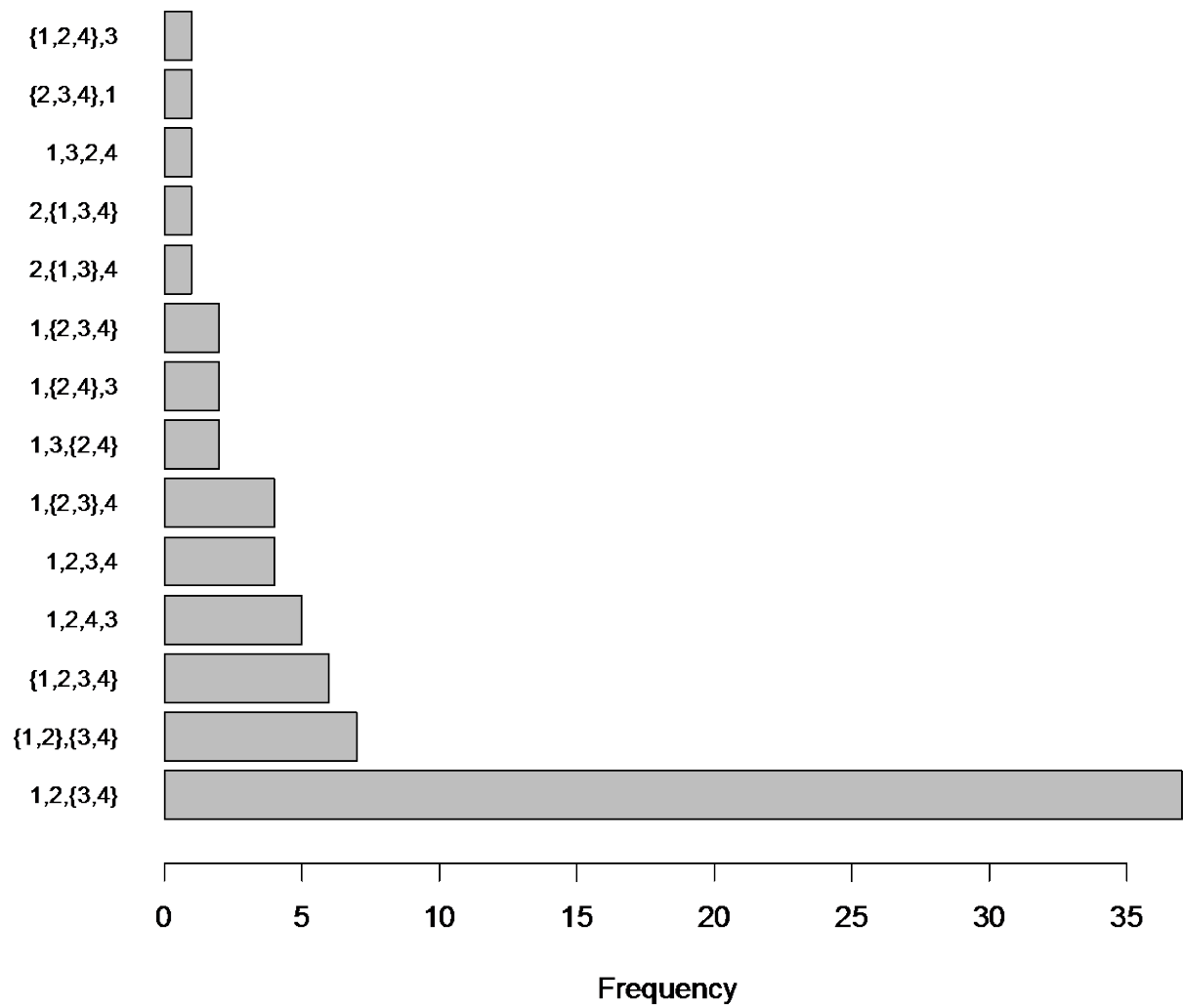


Figure 20. Ranking patterns for vignettes in set 4 (medication side effects).

Table 46. Proportion of cases that a vignette was ranked less than another vignette for set 4 (medication side effects).

Intended order	Vignette #	<1	<2	<3	<4
1	8	-	.77	.88	.88
2	15	.04	-	.78	.81
3	3	.01	.04	-	.16
4	14	.01	.00	.11	-

4.2.9.3 Vignette-adjusted LoSQI scores

One of the benefits of anchoring vignettes is the calculation of an ‘adjusted’ item score for the associated self-report item based upon the subject’s ranking of the vignettes (King et al., 2004). Adjusted scores on each target item were examined using the B-scale, due to high frequency of ties in the dataset (Wand, 2013). Each target-item response was considered to be ‘ordinal’ and median values compared to determine if differences exist in average scores after adjusting for the vignette rankings. Table 47 shows the medians for the four LoSQI items and the medians after adjustment using anchoring vignettes. There was no change between the raw median and adjusted median for the ‘itchy skin’ item. However, there was an increase in the median for the target item in other the three domains after adjusting for vignette rankings. This increase was for a half a point for the item associated with the Physical Functioning theoretical domain and 1 point for the items associated with the theoretical domains of Body Image and Social Support and Medication Side Effects. For a response scale of only 4 points, an increase in 1 point per item could be considered significant. For example, if these adjustments held for all items within the Body Image and Social Support Subscale, it could increase the subscale score by 7 points

(well above the SEM). Although these results are interesting, it is important to note that this analysis is likely limited due to the small sample size. For conclusions to be valid, they should be replicated in larger samples.

Table 47. Comparison between the raw medians and adjusted medians for items with associated anchoring vignettes using the B-Scale (Wand, 2013).

Theoretical Domain	Self-report item	Raw Median	Adjusted Median ¹
Skin sensations	<i>1. Itchy skin</i>	0.0	0.0
Physical Functioning	<i>8. Worry about being able to do certain activities because of my scleroderma</i>	0.0	0.5
Body Image and Social Support	<i>16. Feeling upset when people ask questions</i>	0.0	1.0
Medication Side effects	<i>Meds3. Feeling sick right after I take my medicine</i>	0.0	1.0

¹Medians calculated using B-scale due to frequency of ties, and allocating interval ranks using ‘minimal entropy’ method.

5.0 DISCUSSION

5.1 DISCUSSION OF PRIMARY RESEARCH QUESTIONS

The goal of this project was to develop a survey that accurately and reliably measures quality of life in pediatric LS patients. Validity evidence was gathered from three phases of research; content development and item generation, a qualitative pilot study, and a quantitative field-test. Evidence for the primary and secondary research questions are discussed in the following sections.

5.1.1 RQ 1: What was the validity evidence to support the definition of the content domain for the LoSQI?

Evidence to support the validity of the content domain is discussed in relation to the first two phases of research: development of the LoSQI and the pilot study.

The LoSQI was developed using guidance from multiple standards including those from the PROMIS network, the FDA, the ISOQOL, and *The Standards* (AERA, APA, & NCME, 2014). The main initial conceptual areas were identified and developed based upon transcripts from focus groups with pediatric LS patients, the target population. A thorough literature review was conducted and the most frequently used tool, the 10-item CDLQI, was examined in depth. Salient items from that measure were used as potential theoretical domains for the LoSQI (for

example: social support and skin sensations). LoSQI items were generated from the theoretical conceptual framework and went through multiple rounds of review before being piloted in a diverse sample of pediatric LS patients.

The pilot study supported the definition of the construct. There were only minor indications of construct under-representation, and three additional items were added to address those issues. The new items asked about the patient's skin feeling 'different', missing school, and having to get blood drawn. During the pilot interviews, patients did not indicate major issues with repetitive items (construct over-representation).

The pilot confirmed some of the unique design aspects of the LoSQI. For example, patients specifically confirmed that the two questions relating to problems with using hands (item 1: *problems using my hands when I write, text, or type because of my scleroderma*; item 2: *problems using my hands when I write, text, or type for a long time because of my scleroderma*) were different concepts. There also was support that 'teasing' and 'bullying' were two separate concepts, with bullying having more negative connotations.

In the pilot study, the researcher became aware of the different ways patients refer to their disease, which was heavily influenced by their main symptoms. For example, a patient with mainly hemi-facial atrophy (en coup de sabre subtype) had little to no skin involvement, and was confused at why the tool asked about 'her skin'. To be more inclusive to all patients, other names for localized scleroderma were added to the instructions after the pilot. Specifically, Parry Romberg syndrome was mentioned so that patients, like the one mentioned above, with primary hemi-facial atrophy would feel more included.

The Medication Side Effects Subscale was explored separately than the other theoretical domains because patients were not all on systemic medications. Some were newly diagnosed.

Some were on alternative treatments, like phototherapy or topical medications, and some had well-controlled disease. In the pilot study, some patients had very high scores on this subscale. However, when pressed, they indicated that their symptoms might not be related to their treatment. Most of the ‘side effects’ of systemic medications are also general symptoms related to other conditions or even anxiety/depression; i.e. stomach pain, not feeling like eating, feeling in a fog, fatigue, and headaches. This poses a problem with the validity of the content domain. Changes made to the instrument after the pilot addressed this issue by adding a qualifying statement to all items that reminded patients to *only* consider symptoms specifically related to their medications.

Based upon the pilot study results, edits and revisions were made to the LoSQI items, instructions, and formatting. These changes intended to further support the validity of the content domain.

5.1.2 RQ 2: What was the evidence to support the reliability of the LoSQI scores in this sample?

Both stability and consistency were quantitatively examined during the field test. Test-retest coefficients between the survey administered in clinic and a duplicate survey that was sent home with patients provided support for stability. Duplicate surveys were included in analysis if they were returned within two weeks of the initial administration. Cronbach’s alpha coefficients provided support for internal consistency and were calculated using both the raw responses and dichotomized data, due to the skewed response patterns. Standard errors of measurement (SEM) were calculated for each subscale and the total score using both types of reliability coefficients.

The test-retest coefficients for the subscales were moderate ($>.6$ for the main two subscales, and $>.7$ for the total score and medication subscale). This was surprising, as there were indications that HRQoL should be stable in this population for short time periods. Patients in the pilot study said that their answers would not have changed if they were asked to think back over the past 30 days. However, patients reported that their answers would be significantly different when they were first diagnosed and/or on higher doses of systemic medication (prednisone was particularly mentioned as having a strong negative impact).

There are some arguments in the literature that HRQoL has both ‘state’ and ‘trait’ components (Sprangers & Schwartz, 2008). Measures that tap into the ‘state’ aspects of HRQoL would be more likely to fluctuate. The window for returned duplicate surveys was also two weeks, so patients could have been thinking back over a different seven days for each survey. During those two weeks, it is possible that something could have happened or changed in their life to affect the reported scores. Alternatively, the administrator was not present while patients completed the second survey, which could also have impacted their answers. Patients might have less pressure to ‘play good’ while not in clinic, or they might have been more distracted when filling out the form at home.

There was stronger support for the internal consistency of the LoSQI and subscale scores. Cronbach’s alpha coefficients were very high when calculated using both the 4-point scale and when responses were dichotomized ($=>.9$ for both subscales and total score). The stratified version was also very high ($>.9$).

The researcher considered if the skewness in response patterns for the LoSQI items could potentially affect the accuracy of the Cronbach’s alpha coefficients. It has been found that including highly skewed items along with normally distributed items will *decrease* estimated

reliability when the number of response categories is small and inter-item correlations are weak (Enders & Bandalos, 1999). In this case, inter-item correlations were variable, ranging from .11 to .75, with higher correlations between items on each subscale (Appendix E). Furthermore, all items within the LoSQI had homogeneously skewed response distributions. Another study found that coefficient alpha consistently *underestimated* reliability when used with negatively skewed Likert scale responses (skewness = -2; Zumbo, Gadermann, & Ziesser, 2007), and another showed that the effect of item distributions on reliability was “surprisingly small” (Feldt, 2009). In this light, it is more likely that the high coefficients seen in this sample were appropriate, or even underestimated, and thus, supports the reliability of the LoSQI and its two subscales.

It is important to note, however, a very high coefficient alpha could be problematic in another way. Coefficients $>.9$ could potentially indicate repetitive items (Streiner, 2003); although, none of the LoSQI inter-item correlations were above .8 (Appendix E). An argument could be made to use the standard error of measurements (SEM) calculated from the test-retest reliability coefficients, which in this case, are more conservative than those using coefficient alpha. Either way, the SEM's seemed reasonable, with the total LoSQI having SEMs of 6.0 and 2.8 out of a possible total score of 60 points (for test-retest and alpha, respectively).

For the two subscales of the LoSQI and the total score, item-to-total score correlations were also appropriate (all above .4 and positive). For the medication subscale, item-to-total score correlations were not as ideal. There were a few items that had very low correlations with the total score, indicating that there might be some issues with internal consistency for that subscale.

5.1.3 RQ 3: What is the internal and external validity evidence to support the LoSQI score interpretations?

Based upon *The Standards*, both internal and external validity are required to support the proposed interpretation of scores (AERA, APA, & NCME, 2014). Based on data from the field test, an exploratory factor analysis provided initial support for the internal structure of the survey and is discussed in more detail in the following sections. The utility of reporting a total score along with the subscale scores was debated. External validity evidence was supported by correlations between LoSQI scores and physician, parent, and additional patient reported outcomes.

5.1.3.1 Internal validity evidence

Exploratory factor analysis and score reporting

The results of the factor analysis offer initial, although unconfirmed, insight into the internal structure of the LoSQI due to the very small sample enrolled into the field test ($n = 74$). Due to this significant limitation, both one-factor and two-factor solutions were reported. For both solutions, item 21 (*blood draws*) did not load saliently on the factor(s) and was deleted from the score calculations.

There may be utility in reporting two subscale scores. For the two-factor solution, items from the theoretical domains of Skin Sensations and Physical Functioning loaded together onto the first factor, along with the item related to ‘missing school due to illness’. This factor was termed the ‘Pain and Physical Functioning Subscale’. Although the inclusion of item 20 (missing school) was surprising, conceptually, it fit well within this subscale. Missing school is likely to

be related to physical limitations and/or pain. The second factor consisted entirely of items generated from the theoretical domain of Body Image and Social Support. It is also important to note that no items loaded saliently onto multiple factors.

There was mixed support for reporting a total score along with the two subscale scores. There was one dominant eigenvalue (the ratio of the second eigenvalue to the first was only .22), indicating a total score might be appropriate. However, for the two-factor solution, the correlation between factors was not as strong as expected ($r = .6$). When scatterplots were created looking at the potential subscale scores and the total score, the Pain and Physical Functioning Subscale had a slightly stronger relationship with the LoSQI total score when compared to the Body Image and Social Support Subscale. Still, it is important to note that the first subscale included more items.

As the evidence is mixed and the factor analysis exploratory due to the limitations in sample size and skewed response distributions, at this point, the researcher would recommend reporting a total score along with both subscale scores. A total HRQoL score would be attractive to clinicians; it would provide a more straightforward conceptualization of HRQoL and easily identify patients with very high scores. However, the use of both subscale scores in clinical trials or when making statistical comparisons between subgroups of patients might allow for more sensitivity in determining the effect on the components of HRQoL. Regardless, further study is needed with a larger sample size to confirm the internal structure of the LoSQI.

5.1.3.2 External validity evidence

Before discussing the external validity evidence in detail, it is important to note that all patients did not complete all of the outcome measures. This issue was likely due to the administrative set-up of the study; patients completed the LoSQI items under a new IRB protocol, but the

administration of the other outcomes were covered under each site's unique registry protocol. Thus, there were some differences in data collection. This posed problematic to analysis; the calculated correlation coefficients were not completely stable when using different samples (pair-wise versus list-wise deletion). What was stable was the relative *pattern* of relationships seen in each table (Table 35 & 36). Although detailed claims about the strength of the correlation coefficients cannot be confidently made at this time, the prevailing pattern is interesting and discussed in more detail in the following sections.

Correlations with physician-scored outcomes

Overall, the LoSQI scores (including subscales and the total score) did not correlate with the mLoSSI, LoSDI, or the PGA-activity. This was not surprising, as weak relationships between these outcomes and HRQoL measures have been seen in other studies with these patients (Arkachaisri et al., 2010; Kelsey & Torok, 2013). Patients likely view the disease impact beyond simply cutaneous effects, which are specifically measured by the components of the LoSCAT. The LoSQI did show stronger relationships with the PGA-severity and damage, which are more subjective measures scored by the treating physician and typically take the entire patient experience into account. The relative disagreement with physician measures was expected, as the LoSQI does not ask about the *extent* of symptoms, but about how much the symptoms *bother* the patient. It would be possible for a patient with severe disease, as per their physician, to have a low LoSQI score (and vice versa).

Correlations with parent-reported outcomes

For patients < 18 years old, LoSQI scores showed significant relationships with parents' scores on both HRQoL measures: the VAS-par and the PedsQL FIM. This supports the convergent

validity of the LoSQI, as the domains of the parent-reported measures are conceptually related to the content of the LoSQI. In particular, the eight domains of the PedsQL FIM have a lot of overlap with the LoSQI domains, and include items on physical and social functioning. The agreement with the VAS-par, which is an overall score on a 0-100 scale, likely indicates that parents conceptualize the HRQoL impact of their disease similarly to their children. However, these relationships are not strong enough to indicate that parents are adequate proxies for their children.

Correlations with other patient-reported outcomes

The LoSQI subscale scores correlated as expected with other PRO's, providing further external validity evidence. Overall, relationships between the LoSQI scores and the CDLQI were significant and strong, and were significant and moderate to strong with the VAS-patient. In addition, the LoSQI Pain and Physical Functioning Subscale correlated strongly with the Peds QL Rheum Pain score and moderately with the Activity score. The Body Image and Social Support Subscale correlated moderately with the Communication score. The Pain and Physical Functioning Subscale correlated moderately with the CHAQ and but highly with the HAQ. This could mean that adult patients, who fill out the HAQ, have a larger range of physical limitations than pediatric patients (as correlations are affected by narrow ranges of scores). This relationship has also been seen in some published studies, where adults report more quality of life limitations than children (Das et al., 2014). However, these correlations in particular were based off of very small numbers of patients.

Despite issues with the content domain of the Medication Side Effects Subscale, the score correlated significantly with total scores on the MISS and the 'Treatment' score of the Peds QL Rheum, indicating some overlap in content. The MISS was specifically designed as a

multidimensional scale to detect treatment burden of methotrexate, a common systemic medication. The relationship of the MISS scores to the LoSQI subscale could indicate that the Medication Side Effects Subscale is also measuring a multidimensional construct. This is not surprising, as the items within the subscale capture information on side effects from multiple systemic medications as well as psychosocial aspects of treatment burden (i.e. feeling embarrassed about taking medicines).

5.1.4 RQ 4: How did the LoSQI perform compared to the Children's Dermatology Life Quality Index (CDLQI)?

The LoSQI scores correlated strongly with scores on the CDLQI/DQLI ($\rho > .75$ for both subscales and the total score), indicating an overlap in content between the two surveys. This was generally expected, as some items from the CDQLI were used to generate the content domains of the LoSQI. The CDLQI also correlated with the other PRO's similarly to the LoSQI (Tables 35 & 36). Interestingly, the CDLQI was specifically designed as a measure of HRQoL for cutaneous conditions, but the cutaneous measures (mLoSSI/LoSDI) were not related to the CDLQI any more than they were the LoSQI. In the researcher's opinion, none of the differences in correlation coefficients between the LoSQI and CDLQI/DLQI with the other outcomes were large or seemed clinically relevant.

The 10-item CDLQI is attractive to clinicians because it is short and quick to administer. However, the LoSQI did not have significantly longer administration time than the CDLQI, making it a feasible alternative for use in clinics. There also was a larger range of scores for the 20-item LoSQI when compared to the 10-item CDLQI. This likely represents the collection of more nuanced information about HRQoL impact. For example, patients with uncomfortable skin

sensations have four questions to answer on the LoSQI, compared with only one on the CDLQI. This allows a physician to identify the exact sensations that are bothering the patient, as well as their intensity.

During the course of the study, there was additional information that supported use of the LoSQI over the CDLQI. In the pilot study, patients indicated the questions from the LoSQI were more applicable to their life than the ‘normal’ HRQoL forms they were asked to complete (including the CDLQI). Also, quite a few parents mentioned the positive aspects of having a survey tool designed specifically for their child’s disease. There seemed to be value and buy-in for the LoSQI by the patients and their families, above and beyond that for the CDLQI.

5.2 DISCUSSION OF SECONDARY RESEARCH QUESTIONS

5.2.1 RQ 5: What validity evidence is there to support using the LoSQI anchoring vignettes with pediatric patients?

Like the LoSQI items, the anchoring vignettes were generated from four theoretical domains that were based upon transcripts from focus groups with pediatric LS patients. The vignettes went through multiple rounds of expert review before being piloted in a representative sample of LS patients. Based upon the pilot study interviews, there were no indications that the older patients had trouble understanding the AVs. The older patients, who were >10 years old, were able to describe what the AV meant to them in their own words, articulate how their experiences were similar or different to the person in the AV, and logically report how they came to their answer. One patient mentioned that she thought back to the vignettes when considering her answer to a

self-report item (item: *problems doing fun things like painting or playing an instrument*), indicating that the AVs might help to internally define the LoSQI response scales for patients.

There were some issues that affected the validity of anchoring vignettes in the youngest group of patients, who indicated multiple problems with understanding and readability. Thus, the quantitative examination of AVs was limited to patients who were 10 years old and older. There also were some issues with the vignette rankings within each set, discussed in the following sections.

5.2.2 RQ 6: Did patients rank the LoSQI anchoring vignettes the same way or were there differences?

In general, the vignettes did not rank outside of the intended order, however, ties in the rankings within each set were very prevalent. Around 8-10% of patients assigned all four vignettes the same ranking (for all four vignette sets). This is a major problem with the validity of the vignettes, and limits their utility for adjusting the self-report item scores. It also was common for patients to assign the first two vignettes and the last two vignettes the same ranking, providing only two distinct vignette responses. For the skin sensation vignette set, about a third of the sample ranked the vignettes in this way.

Vignettes designed to rank as ‘a medium amount’ and ‘a lot’ bothersome seemed to cause the most issues for patients. The Body Image and Social Support set was the best functioning vignette set but only had 38% of the sample ranking the vignette designed to rank as ‘a medium amount’ as lower than ‘a lot’, while the physical functioning set only had 22% of patients ranking those two vignettes in the correct order.

The large frequency of ties in rankings could be due to a number of reasons. Adolescents might be unable to cognitively distinguish between the intended subtleties in the vignette descriptions and artificially dichotomize the rankings. It also is possible that the vignettes were not written as clear as necessary or that the underlying constructs are not as variable as expected. This could also bring under question the validity of the response scale, since most patients also ranked the self-report items as being ‘not bothersome’. Perhaps collapsing categories or using a dichotomized response scale would be a more appropriate way to measure the underlying construct. However, the ordinal 4-point scale allows for better identification of a smaller group of patients with very high scores, which is clinically important.

5.2.3 RQ 7: Did adjusting for patients’ response styles using anchoring vignettes change the LoSQI scores significantly? If not, did the anchoring vignette results impart any new information concerning the content domain?

One of the benefits of anchoring vignettes is the calculation of an adjusted score on self-response items based upon the subject’s ranking of the vignettes (King et al., 2004). When items on the LoSQI were adjusted, the largest median change was found for the item generated from the theoretical domains of Body Image and Social Support, which increased from 0 to 1 when adjusted for responses on vignettes. An increase in 1 point could potentially be significant if this relationship held over the entire subscale range, which for Body Image and Social Support could increase the LoSQI subscale and total scores by 7 points.

The self-report item that corresponded to the Body Image and Social Support vignette set was, *‘feeling upset when people ask questions’*. This self-report item bothered about a third of patients at least ‘a little’, which made it one of the most variable items. For the corresponding

vignettes, the most frequent response pattern (26%) was ranking vignette 1 below 2, and 2 below 3 and 4, which were tied. Other frequent response patterns were ranking all four vignettes in order as designed (17%) and ranking all four vignettes the same (13%). This diverse set of patterns is strange, and it is hard to hypothesize why these patterns occur within the same sample. In future studies, additional qualitative interviews would be especially helpful to query patients about their rankings in more detail. In addition to being conceptual confusing, the ties pose a problem to the different statistical models. The C-scale was not used for this reason, as the B-scale is more helpful when frequent ties are present (Wand, 2013). The other statistical models, the HOPIT and the latent variable approach, were not used on this data due to the very small sample size. This limits the conclusions that can be drawn from the anchoring vignette analysis and is discussed in more detail in the next section (5.3).

5.3 LIMITATIONS

The main limitation of this project was the small sample size; only 74 patients were enrolled in the field test. This is low for validity studies and very low for some of the quantitative techniques like factor analysis and anchoring vignette analysis. Unfortunately, this sample size was expected for pediatric LS, as the incidence rate is itself very low (Murray & Laxer, 2002). To optimize enrollment for this study, patients throughout the disease course were recruited, including very established patients who were considered to be in clinical ‘remission’. A longer time from diagnosis could be associated with lower HRQoL impact due to long-term treatment and time to adapt to any functional limitations. In the future, it would be important to enroll more treatment-naïve patients who are newly diagnosed so that the changes in scores over the course of the

disease could be examined. A larger-multicenter study should also be prioritized so that a large enough sample size can be used to replicate the quantitative analysis.

In addition to the small sample size, responses to the items were highly skewed. Although typical in LS samples, the distribution limited the quantitative analysis of results. It is possible that this skewness represents a floor effect in survey scores, and that the response scale could be improved; perhaps by collapsing the more extreme categories in future versions of the LoSQI. Regarding analysis of the field test, item scores were dichotomized for the EFA and calculation of reliability coefficients. When scales are dichotomized, there is a potential loss of information, and it is possible that another type of dichotomization (i.e. 0/1 versus 2/3) could result in different solutions. The dichotomization that was chosen optimized the number of responses in the second group, although there were still some very low frequencies. It is hard to know if the skewness in item responses is due to truly low HRQoL impact in these patients (which, in the researcher's opinion, is likely) or also related to an artificial floor affect because of the way HRQoL is traditionally measured. It could be helpful to collect additional information about a patient's positive coping mechanisms alongside the LoSQI, to capture a more complete picture of their experience.

Although patients skipped very few of the self-response items and vignettes, the background questions on the LoSQI were frequently skipped. This limited the additional analyses that could be run, including responsiveness to change.

There were two minor limitations that should be mentioned regarding the sample of LS patients obtained for the pilot and field studies. First, the frequency of linear scleroderma of the head/neck was higher in the samples for the pilot and field test than in other published cohorts of pediatric LS. The experience of a patient with facial involvement could be very different than

patients with lesions on other parts of their body, even among patients with linear subtype. Although the experiences of these patients are valuable, a larger scale or multisite validation study is necessary to make sure the findings replicate. The second issue was that the sample skewed slightly older than most pediatric samples, with about 30% being >18 years old. One site predominantly sees adult patients, which could have affected this sample characteristic. The LoSQI was designed to be used in patients up to 20 years old, which is typically when they are transferred to adult physicians. However, it is likely that older adult patients require some modification of items to fully capture their mature lifestyles (i.e. romantic relationships, sexual issues, workplace problems).

There was one additional issue that might affect the validity of score interpretations. During the pilot study, there was evidence that some patients were ‘playing good’. ‘Playing good’ was the tendency to respond to the scale in a way that would please the physician or administrator. For example, one parent challenged a patient’s answers after she completed her qualitative interview. The patient had responded that none of the symptoms ‘bothered her’, but the mother said that the patient indicated in her behavior at home that some of the symptoms did in fact bother her. The researcher had a similar impression after interviewing some of the other patients in the 11-14 year old age group, which was also the age group with the lowest LoSQI scores. Although interesting, it is unlikely that this was a pervasive issue, as patients in this age group were very forthcoming about their HRQoL impact in the past. However, instead of using a survey tool in clinic, HRQoL impact might be more accurately measured in this age group using other data collection methods that limit the role of the administrator and increase recall; i.e. using electronic devices that prompt patients to answer LoSQI questions at different times of the day while outside of the clinic.

The final issue with LoSQI scores was related to the Medication Side Effects subscale. The items on this subtest were not included in the exploratory factor analysis because of the restricted sample size (not everyone was on systemic medications at the time of their visit). There also were some low item-to-total score correlations for a few of the items related to general symptomatology. This echoed problems from the pilot study; patients were indicating higher scores on this subscale but, during the interview, said the symptoms were not related to their medications. However, the subscale scores did correlate well with the MISS and ‘Treatment’ score of the Peds QL Rheum, indicating that the items were responded to similarly across scales and could be reflective of a multi-dimensional construct. Another consideration not addressed in this study was potential measurement error introduced because of the diverse administration of systemic medications. Methotrexate is often administered once a week, either by subcutaneous injection or by mouth, while prednisone and Cellcept have different administration methods. To address this, in the future, this subscale could be revised to limit responses to right after the medication is administered, which might result in ‘cleaner’ scores. Before use in clinical trials, it is recommended that this subscale and corresponding items be studied further to determine adequate validity.

5.4 FUTURE DIRECTIONS

There are a few important directions for future research. First, a multicenter study should be prioritized to obtain a larger sample of patients with this rare disease so that the internal structure of the LoSQI, and other results, could be confirmed. In addition, responsiveness to change must be examined to ensure that the LoSQI scores change along with changes in HRQoL or disease

status. The next step after the scores are shown to be responsive is to establish a minimal clinically important difference (MCID). It would be most helpful to look at the changes in scores once systemic treatment is established and the disease goes into remission. Enrolling treatment-naïve pediatric LS patients at diagnosis when the disease is at its worst would be important. Differential item functioning (DIF) analysis using important subgroups of interest would also be valuable.

It also would be helpful to develop an adult version for patients >20 years old so that both pediatric and adult populations could be combined; allowing for larger sample sizes in clinical trials. For pediatric patients, a parental version of the LoSQI would also be an interesting project, as during the pilot study, many parents mentioned their own concerns.

While initial validity and reliability evidence was provided for the LoSQI, the Medication Side Effects Subscale needs to be examined in more detail before use, as some item-to-total score correlations were problematic and the validity evidence not as solid. In the pilot study, many patients indicated that they were reporting symptoms that were not related to their medication. Although the items were qualified in the field test, it is hard to determine if the revisions solved the issue without further study. It is also likely that the medication domain is multidimensional and influenced more strongly by mood, anxiety, depression, developmental level, etc.

Another interesting direction would be looking specifically at the capture of HRQoL in younger patients. In the pilot study, patients who were 8-10 years old indicated problems with understanding and readability of both the vignettes and self-report items. For the field test, the sample was limited to older patients, but it is still important to capture the experience of these younger patients. There are a number of ways that the LoSQI could be adapted to better suit the

cognitive needs of patients <10 years old; pictures of response scale options could be used, the items could be further simplified to reflect a lower reading level, or administrators could be trained on how to assist these patients in a neutral way. It also might be advantageous to tailor vignettes specifically to younger children, as there were many reported issues in the pilot study.

Anchoring vignettes showed some initial promise for use with pediatric populations. The older patients were able to articulate their responses and explain how the vignette character was different or alike to their experience. However, the small sample size available with pediatric LS patients limited the options for statistical analysis, and the 4-point Likert scale showed some negative properties (high frequency of ties). To determine if anchoring vignettes are truly useful for pediatric patients, different samples should be examined.

5.5 RECOMMENDATIONS FOR PRACTITIONERS

There are a number of recommendations that stem from the observed challenges with this project. When validating future HRQoL surveys in pediatric, rare disease populations, practitioners and clinicians should consider the following issues. The biggest limitation with this population was obtaining a large enough sample. For other studies using similar diseases, it would be important to involve as many different sites as possible, focusing on those with large specialized clinics or established research registries. It also would be beneficial to keep the enrollment period open for an extended period of time, to optimize the sample collection. The other main limitation that was seen in this project was the skewness of item response distributions. During the pilot, practitioners should be sure to query patients about the item responses to see if a floor effect is likely to occur, if response categories should be collapsed, and

if the score distributions accurately reflect the underlying level of the construct. If the distributions are ‘truly’ skewed in the population then non-parametric methods should be utilized to minimize bias in results, and multiple analysis methods performed to confirm findings. One of the strengths of this project was the thoughtful and systematic way the content domain was defined. If there are limitations to the procured sample size, future surveys should follow this blueprint and be based upon multiple sources of data: content experts, focus groups with patients, and published literature on both the target disease and comparable conditions. Finally, the early and continued involvement of the patients and their families is instrumental to the process. Patients are the experts on their experiences and can clearly identify areas of impact that clinicians and researchers might overlook.

5.6 CONCLUSION

Overall, there was initial reliability and validity evidence for the use of the LoSQI to capture HRQoL in pediatric LS patients. Along with a total score, two subscale scores, (1) Pain and Physical Functioning and (2) Body Image and Social Support, were explored. Reliability and internal consistency evidence of both the subscales and total scores were acceptable. There was also practical evidence for the use of the LoSQI in clinical settings; the survey was relatively quick to administer (<15 minutes for all patients in the pilot study) and easily understood by patients, making it feasible and attractive to clinicians. During the qualitative interviews, patients indicated understanding of the items, asked minimal questions about the survey and the items themselves, and reported that the recall period seemed appropriate. There were no major problems with missing items, and having the administrators check over the form after completion

could further minimize the small frequency of skipped items. Additionally, during the pilot study, patients and their parents indicated positive support for the scale due to its disease-specific design. Overall, the initial internal and external validity evidence was limited by skewed item distributions and inconsistencies with data collection, but the information gathered was an auspicious first step. Based upon the pilot study and the field test, the LoSQI is a promising tool for capturing HRQoL in LS patients ages 10-20. Larger multi-centered studies should be pursued to gather a large enough sample size for confirmation of the psychometric evaluation.

The Medication Side Effects Subscale also merits more investigation. Initial validity evidence was mixed but, from a clinician's perspective, the subscale provided useful information. It is also likely that the scale has practical value in a clinical setting; patients might find it less burdensome than completing separate full-scale surveys on medication side effects and issues coping with treatment.

There was less evidence for use of anchoring vignettes in this population, as there was a high frequency of ties in rankings within vignette sets, limiting the statistical adjustment of scores. However, older patients (>10 years old) indicated their understanding of the vignettes and support for the assumptions. The adjusted scores, using the B-Scale, seemed to impart some interesting information about the Body Image and Social Support Subscale, which should be examined further in a larger sample.

APPENDIX A

QUALITATIVE INTERVIEW SCRIPT FOR PILOT STUDY

Facilitator (*addressing the patient*): Hi there, like we mentioned before when going over the consent form, you are taking part in this study because we want to make sure that our new survey is doing the best it can at capturing how your daily life is with localized scleroderma. Remember, when filling out this form, there are no right or wrong answers. As the expert on your own life, we want to know YOUR thoughts and feelings. There are two main parts to this survey, questions about your own life and feelings, and then questions about other kids with LS whose experiences could be worse or better than yours. We're going to ask you imagine yourself as that person, and then rank those questions as if you were them. Please let me know if you have any questions as you fill it out– especially if you find a question confusing or unclear. After you are done, we are going to talk about your answers and see if we can make this survey even better.

(*addressing the parent/guardian*): Mom/Dad, I'm going to ask you to stay as silent as possible while your child fills out the survey and answers my questions. After they are done, we will get your opinion on the survey. We know you have a lot of good insights into how your child has been affected by LS and want to get your opinion too. Do either of you have any questions before we begin?

Administers survey to patient.

Anchoring vignettes. Time to completion: _____ minutes.

General survey.

Time to completion: _____ minutes.

Questions the patient asks to facilitator **while taking survey:**

1. _____
2. _____
3. _____
4. _____

Questions the facilitator asks the patient **after survey is completed:**

Anchoring vignette questions

1. Can you read AV [1] aloud to me?

Goal: to determine readability of the AV.

2. Can you tell me about what AV [1] means to you?

Goal: to determine understandability of the AV.

3. Tell me how the person in question [1] is like you?

Goal: to determine if the child is able to imagine themselves as the vignette character.

4. Tell me how the person in question [1] is different from you?

Goal: same as above.

5. What were you thinking when you answered this question?

Goal: to determine response process.

6. Why did you answer this question as a [2]? (if additional prompts are needed)

Goal: to determine response process.

Item-related questions

1. Can you read question [1] aloud to me.

Goal: to determine readability of items.

2. Can you tell me what question [1] means to you?

Goal: to determine understanding of item.

3. What were you thinking when you answered this question? Why did you answer it as a [3]?

Goal: to determine understanding of response process.

7. In item [1], you ranked your own [itchiness] to be a [3], but for vignette [3] you ranked it as a [2]. Was there a difference in how you thought about these two questions?

Goal: Linking vignette question to the self-response item to check the response consistency assumption (defined as each individual uses the response options in the same way for vignettes as for their own ranking).

4. Was there anything that you would add to question [1] to make it clearer?

Goal: to evaluate completeness of question wording and applicability to multiple situations.

5. If I had asked you to think about the past month instead of the past week, would your answer have been different?

Goal: to determine appropriateness of recall period.

Overall questions

1. Was there any important part of your life with LS that you feel was missing from the survey?

Goal: to determine construct under-representation

2. What there a question on the survey that felt the most important to you or bothers you the most?

Goal: to determine importance of domains/items

3. Did you feel like there were any questions that repeated too many times or asked the same thing?

Goal: construct over-representation

APPENDIX B

**PILOT VERSION OF THE LOCALIZED SCLERODERMA QUALITY OF LIFE
INSTRUMENT**

Subject ID: _____

Date: ____/____/____

The Localized Scleroderma Quality of Life Instrument (LoSQI)

We are interested in learning about how localized scleroderma affects you. As the expert on your own life, we want to know YOUR thoughts and feelings, so remember, there are no 'right' or 'wrong' answers.

Before you start answering the questions, please think about your localized scleroderma. Localized scleroderma is sometimes also called morphea. It can change your skin in certain places, your body underneath the skin (like your muscles and bones) and even the size or shape of parts of your body.

	Better	The Same	Worse
Since my last visit to the doctor, my localized scleroderma is....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please answer the following questions as they relate to your localized scleroderma and the parts of your body where you have it. If the symptom described happens but it does not bother you, then please mark 'does not bother me'. You should also mark 'does not bother me', if you do not experience the symptom at all. If the symptom happens AND it bothers you, please mark the choice that best fits you: 'bothers me a little', 'bothers me a medium amount', or 'bothers me a lot'.

How much have the following things bothered you during the PAST SEVEN DAYS?

Skin Symptoms

In the <u>PAST SEVEN DAYS</u>	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
Itchy skin <i>where my scleroderma is</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Painful skin <i>where my scleroderma is</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uncomfortably tight skin <i>where my scleroderma is</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Physical Functioning

In the <u>PAST SEVEN DAYS</u>	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
Problems doing active things like running, playing sports, or dancing <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Problems using my hands when I write, text, or type <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Problems using my hands when I write, text, or type <u>for a long time</u> <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Worry about being able to do certain activities <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Subject ID: _____

Date: ____/____/____

In the PAST SEVEN DAYS	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
Problems when I am doing things fun things like painting or playing an instrument <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My muscles hurting <i>where my scleroderma is</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Aches in my joints (like knees, hips, fingers, toes, ankles, elbows) <i>where my scleroderma is</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stiff joints (like knees, hips, fingers, toes, ankles, elbows) <i>where my scleroderma is</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Body Image, Appearance, and Peer Relations

In the PAST SEVEN DAYS	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
Feeling embarrassed because of how my body looks <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling nervous when I am around new people who don't already know <i>about my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling different than other people <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling upset when people ask questions <i>about my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Being bullied because of the way my skin, face, or body looks <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Getting teased about the way I look <i>because of my scleroderma</i> .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Covering up my scleroderma with things like long sleeves, long pants, makeup, or retainers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Subject ID: _____

Date: ____/____/____

LoSQI: Medication Subscale

Now think about all the medications you take because of your skin, they could be pills that you swallow, shots that your parents or a nurse give you, or even light therapy. The questions below are only about medications that you drink, swallow, or get as a shot – not about creams or lotions.

Are you currently taking medications (NOT including creams/lotions) for your localized scleroderma?

NO (STOP HERE)

YES (PLEASE CONTINUE)

	Better	The Same	Worse	I have no side effects
Since my last visit to the doctor, my medication side effects are....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How much have the following things bothered you during the **PAST SEVEN DAYS**?

In the PAST SEVEN DAYS	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
Worry about medication side effects.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling embarrassed that I need to take medications.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling sick right after I take my medications.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling hungry all the time.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My stomach hurting.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Not feeling like eating.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling like I am in a fog or that it's hard to think clearly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feeling tired.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Having headaches.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gaining weight because of my medications.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vomiting when I take my medications.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How much do you **AGREE** with the following statement?

	Strongly Agree	Agree	Disagree	Strongly Disagree
I believe my medicines will help me to get better.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

APPENDIX C

PILOT VERSION OF ANCHORING VIGNETTES

Subject ID: _____ Date: ____/____/____

The Localized Scleroderma Quality of Life Instrument (LoSQI):

Vignettes (FEMALE)

This part of the survey describes different people with localized scleroderma (also called ‘morphea’) and the symptoms they had over the past seven days. Please read the examples and imagine that person is the same as you in every way; your age, your gender, your likes and dislikes. After you read an example, you will be asked to mark how often the symptom **bothers the person**. There are no right or wrong answers; we want to know YOUR opinion about living with localized scleroderma.

When answering the questions, please think about the example person’s localized scleroderma and the parts of her body where she has it. If the symptom happens AND you believe it would bother the person, please mark the choice that best fits: ‘*bothers her a little*’, ‘*bothers her a medium amount*’, or ‘*bothers her a lot*’. If you feel that the symptom would not bother her, then please mark ‘*does not bother her*’.

Please indicate how much the following symptoms bother the patient...

	Does Not Bother Her	Bothers Her A Little	Bothers Her a Medium Amount	Bothers Her A Lot
When Madison’s scleroderma itches, it feels better right after she scratches it. Sometimes, she feels annoyed that she has to scratch it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kelly worries about going to gym class because of her scleroderma. She is worried about a lot of things; not being able to do the activities that everyone else is doing, accidentally hitting her skin and hurting herself, or having people ask her questions about why she’s not participating.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
After taking her medication(s), Aliyah feels nauseous and tired afterwards. It can be so bad that she has to miss school or skip fun activities on the weekend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Christina does not usually worry about going to gym class because of her scleroderma. She is comfortable with what her body can and cannot do and does not mind taking breaks if she needs to.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The LoSQI Anchoring Vignettes © Zigler, Ardalan, Lane, Stone, Yu, & Torok, 2016

Subject ID: _____

Date: ____ / ____ / ____

	Does Not Bother Her	Bothers Her A Little	Bothers Her a Medium Amount	Bothers Her A Lot
Maria doesn't mind telling people about her skin and localized scleroderma. She knows how to explain it well, and understands that usually they are just curious.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jasmine's skin itches almost every day and night. She scratches her skin most of the day and when it is time to go to bed, it keeps her awake for hours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alexandra gets annoyed when people ask questions about her skin. She tries to explain her scleroderma to them but she sometimes gets mad.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Laila does not like her medications, but she usually feels fine after taking them.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rachel worries about going to gym class because of her scleroderma. She dislikes that she sometimes has to skip participating, but she is used to it at this point.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Katie does not like it when people ask her about her skin. It makes her feel sad and mad when they ask questions and it makes her uncomfortable to talk about it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jessica's skin is itchy during the day, but it gets especially itchy at night. On those nights, it makes it hard for her to fall asleep.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jordan worries about going to gym class because of her scleroderma. She sometimes cannot do the same activities that her friends and classmates are doing and it makes her feel embarrassed to not participate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kayla does not like it when people ask questions about her skin, but she is used to explaining it so it does not upset her that much.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jennifer feels very tired and nauseous after taking her medication. She feels so sick on those days that she cannot go to school or hang out with her friends and family.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sarah feels nauseous after taking her medication(s), but when this happens she can still go to school and hang out with her friends like she usually does.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Abby's skin gets itchy where her scleroderma is, but she does not usually notice it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The LoSQI Anchoring Vignettes © Zigler, Ardalán, Lane, Stone, Yu, & Torok, 2016

Subject ID: _____ Date: ____/____/____

The Localized Scleroderma Quality of Life Instrument (LoSQI):

Vignettes (MALE)

This part of the survey describes different people with localized scleroderma (also called ‘morphea’) and the symptoms they had over the past seven days. Please read the examples and imagine that person is the same as you in every way: your age, your gender, your likes and dislikes. After you read an example, you will be asked to mark how often the symptom bothers the person. There are no right or wrong answers; we want to know YOUR opinion about living with localized scleroderma.

When answering the questions, please think about the example person’s localized scleroderma and the parts of his body where he has it. If the symptom happens AND you believe it would bother the person, please mark the choice that best fits: ‘*bothers him a little*’, ‘*bothers him a medium amount*’, or ‘*bothers him a lot*’. If you feel that the symptom would not bother him, then please mark ‘*does not bother him*’.

Please indicate how much the following symptoms bother the patient...

	Does Not Bother Him	Bothers Him A Little	Bothers Him A Medium Amount	Bothers Him A Lot
When Mike’s scleroderma itches, it feels better right after he scratches it. Sometimes, he feels annoyed that he has to scratch it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
James worries about going to gym class because of his scleroderma. He is worried about a lot of things; not being able to do the activities that everyone else is doing, accidentally hitting his skin and hurting himself, or having people ask him questions about why he’s not participating.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
After taking his medication(s), Tyler feels nauseous and tired. It can be so bad that he has to miss school or skip fun activities on the weekend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Christopher does not usually worry about going to gym class because of his scleroderma. He is comfortable with what his body can and cannot do and does not mind taking breaks if he needs to.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The LoSQI Anchoring Vignettes © Zigler, Ardalan, Lane, Stone, Yu, & Torok, 2016

Subject ID: _____

Date: ____/____/____

	Does Not Bother Him	Bothers Him A Little	Bothers Him A Medium Amount	Bothers Him A Lot
Daniel doesn't mind telling people about his skin and localized scleroderma. He knows how to explain it well, and understands that usually they are just curious.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jordan's skin itches almost every day and night. He scratches his skin most of the day and when it is time to go to bed, it keeps him awake for hours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Matthew gets annoyed when people ask questions about his skin. He tries to explain his scleroderma to them but he sometimes gets mad.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nathan does not like his medications, but he usually feels fine after taking them.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jose worries about going to gym class because of his scleroderma. He dislikes that he sometimes has to skip participating, but he is used to it at this point.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
William does not like it when people ask him about his skin. It makes him feel sad and mad when they ask questions and it makes him uncomfortable to talk about it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Joseph's skin is itchy during the day, but it gets especially itchy at night. On those nights, it makes it hard for him to fall asleep.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Caleb worries about going to gym class because of his scleroderma. He sometimes cannot do the same activities that his friends and classmates are doing and it makes him feel embarrassed to not participate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brandon does not like it when people ask questions about his skin, but he is used to explaining it so it does not upset him that much.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kevin feels very tired and nauseous after taking his medications. He feels so sick on those days that he cannot go to school or hang out with his friends and family at all.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ethan feels nauseous after taking his medication(s), but when this happens he can still go to school and hang out with his friends like he usually does.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anthony's skin gets itchy where his scleroderma is, but he does not usually notice it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

APPENDIX D

REVISED VERSION LOSQI

Subject ID: _____

Date: ____ / ____ / ____

**The Localized Scleroderma Quality of Life Instrument (LoSQI):
Vignettes (FEMALE)**

This part of the survey describes different people with localized scleroderma (also called ‘morphea’) and the symptoms they had over the past seven days.

Please read the examples and imagine that person is the same as you in every way. After you read an example, you will be asked to mark if you think the symptom **bothers the person**.

There are no right or wrong answers; we want to know YOUR opinion about living with localized scleroderma.

Please tell us how much the person would be bothered...

	Does Not Bother Her	Bothers Her A Little	Bothers Her a Medium Amount	Bothers Her A Lot
1. When Madison’s scleroderma itches, it feels better right after she scratches it. Sometimes, she feels annoyed that she has to scratch it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Kelly worries about going to gym class because of her scleroderma. She is very worried about not being able to do the activities that everyone else is doing and is afraid of accidentally hitting her skin and hurting herself.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. After taking her medication(s), Aliyah feels sick to her stomach and tired. It can last awhile and sometimes she has to miss school or skip fun activities on the weekend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Christina does not usually worry about going to gym class because of her scleroderma. She is comfortable with what her body can and cannot do and does not mind taking breaks if she needs to.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Maria does not mind telling people about her skin and localized scleroderma. She knows how to explain it well, and understands that usually they are just curious.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Jasmine’s skin itches almost every day and night. She scratches her skin most of the day and when it is time to go to bed, it keeps her awake for hours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Alexandra does not like it when people ask questions about her skin. She tries to explain her scleroderma to them but she sometimes gets mad.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Subject ID: _____

Date: ____ / ____ / ____

	Does Not Bother Her	Bothers Her A Little	Bothers Her a Medium Amount	Bothers Her A Lot
8. Laila usually feels fine after taking her medicines. On the days that she takes them, she can go to school and hang out with her friends like she always does.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Rachel worries about going to gym class because of her scleroderma. She dislikes that she sometimes has to skip participating, but she is used to it at this point.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Katie does not like it when people ask her about her skin. It makes her feel sad and mad when they ask questions and it makes her very upset.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Jessica's skin is itchy during the day, but it gets very itchy at night. On those nights, it makes it hard for her to fall asleep.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Jordan worries about going to gym class because of her scleroderma. She sometimes cannot do the same activities that her friends and classmates are doing and it makes her feel embarrassed to not participate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Kayla does not like it when people ask questions about her skin, but she is used to explaining it so it does not upset her that much.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Jennifer always feels tired and sick to her stomach after taking her medication. She feels so sick on those days that she cannot go to school or hang out with her friends and family.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Sarah sometimes feels sick to her stomach after taking her medication(s), but when this happens she can still go to school and hang out with her friends like she usually does.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Abby's skin gets itchy where her scleroderma is, but she does not usually notice it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Subject ID: _____

Date: ____ / ____ / ____

The Localized Scleroderma Quality of Life Instrument (LoSQI)

We are interested in learning about how localized scleroderma affects you.

As the expert on your own life, we want to know YOUR thoughts and feelings, so remember, there are no 'right' or 'wrong' answers.

Localized scleroderma is sometimes called morphea, en coup de sabre, and Parry Romberg syndrome. It can change your skin in certain places, your body underneath the skin (like your muscles and bones) and even the size or shape of parts of your body.

Please answer the following questions **as they relate to your localized scleroderma and the parts of your body where you have it**.

Since my last visit to the doctor, my localized scleroderma is....	Better <input type="checkbox"/>	The Same <input type="checkbox"/>	Worse <input type="checkbox"/>	Not Applicable: 1 st Visit <input type="checkbox"/>
--	------------------------------------	--------------------------------------	-----------------------------------	--

How much have the following things bothered you during the PAST SEVEN (7) DAYS?

In the <u>PAST SEVEN DAYS</u>	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
1. Itchy skin <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Painful skin <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Uncomfortably tight skin <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. My skin feeling different <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Problems doing active things like running, playing sports, or dancing <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Problems using my hands when I do things like write, text, or type <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Problems using my hands when I do things like write, text, or type for a long time <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Worry about being able to do certain activities <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Subject ID: _____

Date: ____ / ____ / ____

In the PAST SEVEN DAYS.....	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
9. Problems when I am doing things fun things like painting or playing an instrument <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. My muscles hurting <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Aches in my joints (like knees, hips, fingers, toes, ankles, elbows) <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Stiff joints (like knees, hips, fingers, toes, ankles, elbows) <i>where my scleroderma is.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Feeling embarrassed because of how my body looks <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Feeling nervous when I am around new people who don't already know <i>about my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Feeling different than other people <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Feeling upset when people ask questions <i>about my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Covering up my scleroderma with things like long sleeves, long pants, makeup, or retainers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Getting teased about the way I look <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Being bullied about the way I look <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Having to miss school for doctor's appointments, therapy, or feeling sick <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Getting my blood drawn <i>because of my scleroderma.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Subject ID: _____

Date: ____ / ____ / ____

LoSQI: Medication Subscale

Now think about all the medicines you take because of your localized scleroderma.

The questions below are ONLY about medicines that you drink, swallow, or get as a shot – not about creams or lotions.

Are you currently taking medicine for your localized scleroderma (NOT including creams/lotions)?

NO (STOP here - you're done!)

YES (Keep answering!)

	Better	The Same	Worse	I have no side effects
Since my last visit to the doctor, my medicine side effects are....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How much have the following things bothered you during the PAST SEVEN DAYS?

In the PAST SEVEN DAYS.....	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
1. Worry about medicine side effects.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Feeling embarrassed that I need to take medicine.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Feeling sick right after I take my medicine.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Feeling hungry all the time <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. My stomach hurting <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Not feeling like eating <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Vomiting <i>when I take my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Feeling tired <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The LoSQI © Zigler, Ardalan, Lanc, Stone, Yu, & Torok, 2016

Subject ID: _____

Date: ____ / ____ / ____

In the PAST SEVEN DAYS.....

	Does Not Bother Me	Bothers Me A Little	Bothers Me A Medium Amount	Bothers Me A Lot
9. Feeling like I am in a fog or that it's hard to think clearly <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Having headaches <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Gaining weight <i>because of my medicine.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Finally, how much do you AGREE with the following statement?

	Strongly Agree	Agree	Disagree	Strongly Disagree
12. I believe my medicine will help me get better.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

APPENDIX E

INTER-ITEM CORRELATIONS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1																			
2	.58**	1																		
3	.57**	.62**	1																	
4	.53**	.34**	.56**	1																
5	.34**	.38**	.57**	.40**	1															
6	.31**	.27*	.47**	.33**	.53**	1														
7	.34**	.33**	.49**	.42**	.45**	.82**	1													
8	.47**	.60**	.69**	.56**	.67**	.49**	.59**	1												
9	.49**	.57**	.55**	.35**	.59**	.40**	.37**	.66**	1											
10	.57**	.60**	.65**	.37**	.54**	.34**	.36**	.60**	.60**	1										
11	.59**	.48**	.51**	.42**	.57**	.35**	.36**	.52**	.57**	.54**	1									
12	.38**	.27*	.40**	.35**	.47**	.27*	.29*	.34**	.43**	.47**	.67**	1								
13	.29*	.19	.32**	.38**	.22	.35**	.33**	.28*	.23	.38**	.16	.16	1							
14	.35**	.33**	.42**	.36**	.37**	.46**	.46**	.42**	.32**	.44**	.27*	.16	.79**	1						
15	.37**	.45**	.52**	.37**	.51**	.41**	.41**	.58**	.50**	.55**	.34**	.20	.67**	.72**	1					
16	.34**	.22	.42**	.41**	.31**	.27*	.27*	.41**	.27*	.34**	.32**	.36**	.64**	.65**	.57**	1				
17	.15	.20	.38**	.27*	.43**	.47**	.41**	.43**	.38**	.35**	.29*	.28*	.75**	.75**	.64**	.63**	1			
18	.30**	.51**	.50**	.20	.26*	.46**	.51**	.42**	.19	.45**	.21	.21	.54**	.58**	.52**	.43**	.51**	1		
19	.27*	.40**	.33**	.26*	.19	.31**	.29*	.23	.15	.40**	.24*	.31**	.51**	.47**	.52**	.34**	.47**	.67**	1	
20	.44**	.37**	.40**	.39**	.39**	.19	.27*	.53**	.24*	.49**	.33**	.17	.16	.24*	.41**	.33**	.21	.30**	.22	1

BIBLIOGRAPHY

- Ahmed, S., Berzon, R.A., Revicki, D.A., Lenderking, W.R., Moinpour, C.M., Basch, E., Reeve, B., Wu, A.W., International Society for Quality of Life Research. (2012). The use of patient-reported outcomes within comparative effectiveness research: Implications for clinical practice and health care policy. *Medical Care*, 50(12), 1060-70.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Angelini, V., Cavapozzi, D., & Paccagnella, O. (2011). Dynamics of reporting work disability in Europe. *Journal of the Royal Statistical Society*; 174(3), 621-638.
- Ardalan, K., Zigler, C.E., & Torok, K.S. (2016). Predictors of longitudinal quality of life in pediatric localized scleroderma. *Arthritis Care & Research*. DOI: 10.1002/acr.23101
- Ardalan, K., Switzer, G.E., Zigler, C.E., Hershey, N.M., & Torok, K.S. (submitted). Psychometric properties of the Children's Dermatology Life Quality Index in pediatric localized scleroderma.
- Arkachaisri, T., Vilaiyk, S., Torok, K.S., & Medsger, T.A. (2010). Development and initial validation of the Localized Scleroderma Skin Damage Index and Physician Global Assessment of disease Damage: A proof-of-concept study. *Rheumatology*, 49, 373-81.
- Arkachaisri, T., & Pino, S. (2008). Localized scleroderma severity index and global assessments: a pilot study of outcome instruments. *The Journal of rheumatology*, 35(4), 650-657.
- Arkachaisri, Thaschawee, Vilaiyuk, S., Li, S., O'Neil, K.M., Pope, E., Higgins, G.C., Punaro, M., Rabinovich, E.C., Rosenkranz, M., Keitz, D.A., Rosen, P., Spalding, S.J., Hennon, T.R., Torok, K.S., Cassidy, E., Medsger Jr., T.A., & the Localized Scleroderma Clinical and Ultrasound Study Group. The localized scleroderma skin severity index and physician global assessment of disease activity: a work in progress toward development

- of localized scleroderma outcome measures. *The Journal of Rheumatology* 36.12 (2009): 2819-2829.
- Au, N., & Lorgelly, P.K. (2013). Anchoring vignettes for health comparisons: an analysis of response consistency. *Quality of Life Research*, 23(6): 1721-31.
- Baildam, E.M., Ennis, H., Foster, H.E., Shaw, L., Chieng, A.S., Kelly, J., Herrick, A.L., & Richards, H.L. (2011). Influence of childhood scleroderma on physical function and quality of life. *Journal of Rheumatology*, 38(1), 167-73.
- Beyer, J.E., McGrath, P.J., & Berde, C.B. (1990). Discordance between self-report and behavioral pain measures in children aged 3-7 years after surgery. *Journal of Pain and Symptom Management*, 5(6), 350-356.
- Bode, R.K., Klein-Gitelman, M.S., Miller, M.L., Lechman, T.S., & Pachman, L.M. (2003). Disease activity score for children with juvenile dermatomyositis: Reliability and validity evidence. *Arthritis Care & Research*, 49(1), 7-15.
- Bolt, D.M. & Johnson, T.R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Measurement*, 33(5), 335-352.
- Bolt, D.M., Lu, Y., & Kim, J. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, advanced online publication.
- Bootomley, A., Flechtner, H., Efficace, F., Vanvoorden, V., Coens, C., Therasse, P., Velikova, G., Blazeby, J., Greimel, E., & the European Organization for Research and Treatment of Cancer (EORTC) Data Center and Quality of Life Group. (2005). Health related quality of life outcomes in cancer clinical trials. *European Journal of Cancer*, 41(12), 1697-709.
- Borgers, N., Hox, J., & Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and Quantity*, 38, 17-33.
- Buckley, J. (2008). Survey context effects in anchoring vignettes.
URL <http://polmeth.wustl.edu/media/Paper/surveyartifacts.pdf>
- Bulatovic, M., Heijstek, M.W., Verkaaik, M., Pieter van Dijkhuizen, E.H., Armbrust, W., Hoppenreijns, E.P.A., Kamphuis, S., Kuis, W., Egberts, T.C.G., Sinnema, G., Rademaker, C.M.A., & Wulfraat, N.M. (2011). High prevalence of methotrexate intolerance in juvenile idiopathic arthritis: Development and validation of a methotrexate intolerance severity score. *Arthritis & Rheumatism*, 63(7), 2007-2013.
- Calasan, M.B., van den Bosch, O.F.C., Creemers, M.C.W., Custers, M., Heurkens, A.H.M., van

- Woerkom, J.M., & Wulffraat, N.M. (2013). Prevalence of methotrexate intolerance in rheumatoid arthritis and psoriatic arthritis. *Arthritis Research & Therapy*; 15(R217), 1-5.
- Carroll, J.B. (1997). Psychometrics, intelligence, and public perception. *Intelligence*; 24(1), 25-52.
- Calvert, M., Blazeby, J., Altman, D.G., Revicki, D.A., Moher, D., & Brundage, M.D. (2013). Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *Journal of the American Medical Association*; 309(8), 814-822.
- Chen, F., Sousa, K.H., & West, S.G. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*; 41(2), 189-225.
- Chevalier, A., & Fielding, A. (2011). An introduction to anchoring vignettes. *Journal of the Royal Statistics Society*; 174(3): 569-74.
- Christen-Zaech, S., Hakim, M.R., Asfar, F.S., & Paller, A.S. (2008). Pediatric morphea (localized scleroderma): Review of 136 patients. *Journal of the American Academy of Dermatology*, 59(3), 385-96.
- Condie, D., Grabell, D., & Jacobe, H. (2014). Comparison of outcomes in adults with pediatric-onset morphea and those with adult-onset morphea: A cross-sectional study from the morphea in adults and children cohort. *Arthritis & Rheumatology*; 66(12), 3496-504.
- Costello, A.B. & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1-9.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 443-507). Washington, DC: American Council on Education.
- Das, S., Berstein, I., & Jacobe, H. (2014). Correlates of self-reported quality of life in adults and children with morphea. *Journal of the American Academy of Dermatology*, 70(5), 904-10.
- DeMars, C.E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*; 13, 354-378.
- Deshpande, P. R., Rajan, S., Sudeepthi, B. L., & Abdul Nazir, C. P. (2011). Patient-reported outcomes: A new era in clinical research. *Perspectives in Clinical Research*, 2(4), 137-144. <http://doi.org/10.4103/2229-3485.86879>
- DeWalt, D.A., Rothrock, N., Yount, S., & Stone, A.A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, 45(5:1), S12-S21.

- Dowd, J.B., & Todd, M. (2011). Does self-reported health bias the measurement of health inequalities in U.S. adults? Evidence using anchoring vignettes from the health and retirement study. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*; 66(4), 478-89.
- Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostralateral prefrontal cortex. *Developmental Cognitive Neuroscience*, 10, 57-76.
- d'Uva, T.B., Lindeboom, M. O'Donnell, O., & van Doorslaer, E. (2011). Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. *Journal of Human Resources*; 46(4): 875-906.
- Enders, C.K. & Bandalos, D.L. The effects of heterogeneous item distributions on reliability. *Applied Measurement in Education*, 12(2), 133-150.
- Ennis, H., Herrick, A.L., Baidam, E.M., & Richards, H.L. (2012). Children's and parents' beliefs about childhood onset scleroderma are influenced by child age and physical function impairment. *Rheumatology*, 51(7), 1331-3.
- Fabrigar, L.R., Wegener, D.T., MacCallum, R.B., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3): 272-299.
- Falanga, V., Medsger Jr., T.A., Reichlin, M., & Rodnan, G.P. (1986). Linear scleroderma: Clinical spectrum, prognosis, and laboratory abnormalities. *Annals of Internal Medicine*, 104(6), 849-59.
- Feldt, L.S. (2009). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education*, 6(1), 37-48.
- Fett, N., & Werth, V.P. (2011). Update on morphea: Part 1. Epidemiology, clinical presentation and pathogenesis. *Journal of the American Academy of Dermatology*, 64(2), 217-228.
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, 32(3), 221.
- Foster, J.M., van Sonderen, E., Lee, A.J., Sanderman, R., Dijkstra, A., Postma, D.S., & van der Molen, T. (2006). A self-rating scale for patient-perceived side effects of inhaled corticosteroids. *Respiratory Research*, 7(131), 1-15.
- Ghetti, S., DeMaster, D.M, Yonelinas, A.P., & Bunge, S.A. (2010). Developmental differences in medial temporal lobe function during memory encoding. *The Journal of Neuroscience*, 30(28), 9548-9556.
- Gibbons, R.B, Clark, D.C, Cavanaugh, S.V., & Davis, J.M. (1985). Application of modern

- psychometric theory in psychiatric research. *Journal of Psychiatric Research*; 19(1), 43-55.
- Gibbons, R.D., Immekus, J.C., & Bock, R.D. (2007). The added value of multidimensional IRT models. *Multidimensional and hierarchical modeling monograph, 1*.
- Gibbons, R.D., Rush, A.J., & Immekus, J.C. (2009). On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. *Journal of Psychiatric Research*; 43, 401-410.
- Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Grabell, D., Hsieh, C., Andrew, R., Martires, K., Kim, A., Vasquez, R., & Jacobe, H. (2014). The role of skin trauma in the distribution of morphea lesions: a cross-sectional survey of the morphea in adults and children cohort IV. *Journal of the American Academy of Dermatology*, 71(3), 493-8.
- Grol-Prokopczyk, H. (2014). Age and sex differences in anchoring vignette studies: Methodological and empirical contributions. *Survey Research Methods*; 8(1): 1-17.
- Gustafsson, J. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*; 28(4), 407-434.
- Hare, C., Armstrong II, D.A., Bakker, R., Carroll, R., & Poole, K.T. (2015). Using Bayesian Aldrich-McKelvey scaling to study citizens' ideological preferences and perceptions. *American Journal of Political Science*; 59(3), 759-774.
- Hayes, R.D., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity not a separate dimension. *Quality of Life Research*, 1(1), 73-5.
- Hays, R.D., Brodsky, M., Johnston, M.F., Spritzer, K.L., & Hui, K. (2005). Evaluating the statistical significance of health-related quality of life change in individual patients. *Evaluation & the Health Professions*, 28(2), 160-70.
- Heiland, F. & Yin, N. (2015). Anchoring vignettes in the health and retirement study: How do medical professionals and disability recipients characterize the severity of work limitations? *PLoS ONE*; 10(5): e0126218.
- Herrick, A.L., Ennis, H., Bhushan, M., Silman, A.J., & Baildam, E.M. (2011). Clinical features of childhood localized scleroderma in an incidence cohort. *Rheumatology*, 50, 1865-1868.
- Hirve, S., Gomez-Olive, X., Oti, S., Debpuur, C., Juvekar, S., Tollman, S., Blomstedt, Y., Wall,

- S., & Ng, N. (2013). Use of anchoring vignettes to evaluate health reporting behavior amongst adults aged 50 years and above in Africa and Asia – testing assumptions. *Global Health Action*; 6: 210-54.
- Hoffman, M. L. (1994). The contribution of empathy to justice and moral judgment. *Reaching Out: Caring, Altruism, and Prosocial Behavior*, 7, 161-194.
- Hopkins, D.J., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*; 74(2); 201-222.
- Huitt, W., & Hummel, J. (2003). Piaget's theory of cognitive development. *Educational psychology interactive*, 3(2).
- IBM Corp. Released 2015. IBM SPSS Statistics for Macintosh, Version 23.0. Armonk, NY: IBM Corp.
- ISOQOL. "What is health related quality of life?". Retrieved on April 11, 2016 from: <http://www.isoqol.org/about-isoqol/what-is-health-related-quality-of-life-research>
- Joreskog, K.G. (1999). How large can a standardized coefficient be? Retrieved on July 27, 2017 from: <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf>
- Jurges, H., & Winter, J. (2013). Are anchoring vignettes ratings sensitive to vignette age and sex? *Health Economics*; 22: 1-13.
- Kane, M.T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4). 319-342.
- Kapteyn, A., Smith, J.P., & van Soest, A., (2007). Vignettes and self-reports of work disability in the United States and Netherlands. *The American Economic Review*; 97(1): 461-73.
- Kapteyn, A., Smith, J.P., van Soest, A., & Vonkova, H. (2011). Anchoring vignettes and response consistency. *RAND Labor and Population Working Paper Series*; WR-840.
- Kelsey, C., & Torok, K. (2013). The localized scleroderma assessment tool (LoSCAT): Responsiveness to change in a pediatric clinical population. *Journal of the American Academy of Dermatology*, 69(2), 214-220.
- Klepper, S.E. (2003). Measures of pediatric function: Childhood Health Assessment Questionnaire (CHAQ). *Arthritis & Rheumatism*, 49(5S), S5-S14.
- King, G., Murray, C.J.L., Salomon, J.A., & Tandon, A., (2004). Enhancing the validity and

- cross-cultural comparability of measurement in survey research. *American Political Science Review*; 98(1): 191-207.
- King, G. & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis*, 15(1), 46-66.
- Klepper, S. (2003). Measures of pediatric function: The Childhood Health Assessment Questionnaire. *Arthritis & Rheumatism*, 49(5S), S5-S14.
- Klimas, N.K., Shedd, A.D., Berstein, I.H., & Jacobe, H. (2014). Health-related quality of life in morphea. *British Journal of Dermatology*, 172, 1329-37.
- Korfage, I.J., de Koning, H.J., & Essink-Bot, M. (2007). Response shift due to diagnosis and primary treatment of localized prostate cancer: a then-test and a vignette study. *Quality of Life Research*; 16: 1627-1634.
- Kreuter, A. (2012). Localized scleroderma. *Dermatologic Therapy*, 25, 135-147.
- Kreuter, A., Hyan, J., Stucker, M., Sommer, A., Altmeyer, P. & Gambichler, T. (2006). A randomized controlled study of low-dose UVA1, medium-dose UVA1, and narrowband UVB phototherapy in the treatment of localized scleroderma. *Journal of the American Academy of Dermatology*, 54(3), 440-447.
- Kristensen, N., & Johansson, E. (2006). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics*; 15(1): 96-117.
- Kroft, E.B., Creemers, M.C., van den Hoogen, F.H., Boezeman, J.B., & de Jong, E.M. (2009). Effectiveness, side-effects, and period of remission after treatment with methotrexate in localized scleroderma and related sclerotic skin diseases: an inception cohort study. *British Journal of Dermatology*, 160(5), 1075-82.
- Kroft, E.B., de Jong, E.M., & Evers, A.W. (2009). Psychological distress in patients with morphea and eosinophilic fasciitis. *Archives of Dermatology*, 145(8), 1017-22.
- Kvannli, L., Finlay, V., Edgar, D.Q., Wu, A., Wood, F.M. (2011). Using the burn specific health scale-brief as a measure of quality of life after a burn – What score should clinicians expect? *Burns*, 41(4), 721-6.
- Laitakari, E., Koljonen, V., PYorala, S., Rintala, R., Roine, R.P., & Sintonen, H. (2015). The long-term health-related quality of life in children treated for burns as infants 5-9 years earlier. *Burns*, 41(6), 1186-92.
- Laxer, R. M., & Zulian, F. (2006). Localized scleroderma. *Current Opinion in Rheumatology*, 18(6), 606-613.

- Leitenberger, J.J., Cayce, R.L., Haley, R.W., Adams-Huet, B., Bergstresser, P.R., & Jacobe, H.T. (2009). Morphea subtypes are distinct autoimmune syndromes: A review of 245 adult and pediatric cases. *Archives of Dermatology*, 145(5), 545-550.
- Lewis-Jones, M.S. & Finlay, A.Y. (1995). The children's dermatology life quality index (CDLQI): Initial validation and practical use. *British Journal of Dermatology*, 132(6), 942-49.
- Li, S. C., Torok, K. S., Pope, E., Dedeoglu, F., Hong, S., Jacobe, H. T., Rabinovich, C.E., Laxer, R.M., Higgins, G.C., Ferguson, P.J., Lasky, A., Baszis, K., Becker, M., Campillo, S., Cartwright, V., Cidon, M., Inman, C.J., Jerath, R., O'Neil, K.M., Vora, S., Zeft, A., Wallace, C.A., Illowite, N.T., Fuhlbrigge, R.C. & the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Localized Scleroderma Workgroup. (2012). Development of consensus treatment plans for juvenile localized scleroderma: A roadmap toward comparative effectiveness studies in juvenile localized scleroderma. *Arthritis Care & Research*, 64(8), 1175-1185.
- Likert Scale. (2008). In W. A. Darity, Jr. (Ed.), *International Encyclopedia of the Social Sciences* (2nd ed., Vol. 4, pp. 447-448). Detroit: Macmillan Reference USA.
- MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-89.
- Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-49.
- Maskell, J., Newcombe, P., Martin, G., & Kimble, R. (2013). Psychosocial functioning differences in pediatric burn survivors compared with health norms. *Journal of Burn Care Research*, 34(4), 465-76.
- McGlothlin, A.E., & Lewis, R.J. (2013). Minimal clinically important difference: Defining what really matters to patients. *The Journal of the American Medical Association*, 312(13), 1342-43.
- Mertens, J.S., Seyger, M.M.B., Kievit, W., Hoppenreijns, E.P.A.H., Jansen, T.L.Th.A., van de Kerkhof, P.C.M., Radstake, T.R.D., & de Jong, E.M.G.J. (2014). Disease recurrence in localized scleroderma: A retrospective analysis of 344 patients with paediatric- or adult-onset disease. *British Journal of Dermatology*, 172, 722-28.
- Mojtabai, R. (2015). Depressed mood in middle-aged and older adults in Europe and the United States: A comparative study using anchoring vignettes. *Journal of Aging and Health*; 1-23.
- Moorthy, L.N., Peterson, M.G.E., Baratelli, M., Harrison, M.J., Onel, K.B., Chalom, E.C.,

- Haines, K., Haskhes, P.J., & Lehman, T.J.A. (2007). Multicenter validation of a new quality of life measure in pediatric lupus. *Arthritis & Rheumatism*, 57(7), 1165-73.
- Morren, M. Gelissen, J., & Vermunt, J. (2012). The impact of controlling extreme responding on measurement equivalence in cross-cultural research. *Methodology*; 8(4), 159-170.
- Mulligan, K., Kassoumeri, L., Etheridge, A., Moncrieffe, H., Wedderburn, L.R., & Newman, S. (2013). Mothers' reports of the difficulties that their children experience in taking methotrexate for Juvenile Idiopathic Arthritis and how these impact on quality of life. *Pediatric Rheumatology Online Journal*, 11(1), 11-23.
- Mulligan, K., Wedderburn, L.R., & Newman, S. (2015). The experience of taking methotrexate for juvenile idiopathic arthritis: Results of a cross-sectional survey with children and young people. *Pediatric Rheumatology*, 13(1), 58.
- Murphy, M.E., Holzer, C.E., Richardson, L.M., Epperson, K., Ojeda, S., Martinez, E.M., Suman, O.E., Herndon, D.N., & Meyer, W.J. (2015). Quality of life of young adult survivors of pediatric burns using world health organization disability assessment scale II and burn specific health scale-brief: A comparison. *Journal of Burn Care Research*, 36(5), 521-33.
- Murray, K.J. & Laxer, R.M. (2002). Scleroderma in children and adolescents. *Rheumatic Disease Clinics of North America*, 28, 603-624.
- Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Nelson, E.C., Eftimovska, E., Lind, C., Hager, A., Wasson, J.H., & Lindblad, S.(2015). Patient reported outcome measures in practice. *The BMJ*, 350, 7817.
- Nelson, E., Conger, B., Douglass, R., Gephart, D., Kirk, J., Page, R., Clark, A., Johnson, K., Stone, K., Wasson, J., & Zubkoff, M. (1983). Functional health status of primary care patients. *JAMA*, 249(24), 3331-8.
- Orzechowski, N. M., Davis, D. M., Mason, T. G., Crowson, C. S., & Reed, A. M. (2009). Health-related quality of life in children and adolescents with juvenile localized scleroderma. *Rheumatology*, 48(6), 670-672.
- Osborne, J.W. What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evaluation*, 20(2), 1-7.
- Paccagnella, O. (2011). Anchoring vignettes with sample selection due to non-response. *Journal of the Royal Statistical Society*; 174(3), 665-687.
- Paccagnella, O. (2013). Modeling individual heterogeneity in ordered choice models: Anchoring

- vignettes and the CHOPIT model. *Journal of Methodological and Applied Statistics*; 15, 69-89.
- Palmero, M.L., Uziel, Y., Laxer, R.M., Forrest, C.R., & Pope, E. (2010). En coup de sabre scleroderma and Parry-Romberg syndrome in adolescents: surgical options and patient-related outcomes. *Journal of Rheumatology*, 37(10), 2174-9.
- Patrick, D.L. & Deyo, R.A. (1989). Generic and disease-specific measures in assessing health-related quality of life. *Medical Care*, 27(3 Suppl), S217-32.
- Peracchi, F., & Rossetti, C. (2012). Heterogeneity in health responses and anchoring vignettes. *Empirical Economics*; 42: 513-538.
- Pouchot, J. Ruperto, N., Lemelle, I., Sommelet, D., Grouteau, E., David, L., Duquesne, A., Deslandre, C.J., Paut, I.K., Pillet, P., Goumy, L., Barbier, C., Guyot, M.H., Mazingue, F., Laloum, S.G., Fischbach, M., Quartier, P., Guyot, C., Jean, S., Le Gall, E., Plouvier, E., Bost, M., de Lumley, L., LePlege, A., Larbre, J.P., Guillemin, F., Coste, J., Landgraf, J.M., Prieur, A.M. Necker, Paris, and Nantes Mère Enfant. (2001). The French version of the childhood health assessment questionnaire (CHAQ) and the child health questionnaire (CHQ). *Clinical and Experimental Rheumatology*, 19(23) S60-S65.
- PROMIS Instrument Development and Scientific Standards Version 2.0. (2013). Available from: http://www.nihpromis.org/Documents/PROMISStandards_Vers2.0_Final.pdf
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rae, G. (2007). A note on using stratified alpha to estimate the composite reliability of a test composed of interrelated nonhomogeneous items. *Psychological Methods*, 12(2), 177-84.
- Raj, V., & Bell, M.A. (2010). Cognitive processes supporting episodic memory formation in childhood: The role of source memory, binding, and executive functioning. *Developmental Review*, 30(4), 384-402.
- Reiffe, C., Ketelaar, L., & Wiefferink, C.H. (2010). Assessing empathy in young children: Construction and validation of an empathy questionnaire (EmQue). *Personality and Individual Differences*, 49(5), 363-367.
- Rice, N., Robone, S., & Smith, P.C. (2010). International comparison of public sector performance: The use of anchoring vignettes to adjust self-reported data. *Evaluation*; 16(1): 81-101.
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*;

47(3), 361-372.

- Riley, A.W. (2004). Evidence that school-aged children can self-report on their health. *Ambulatory Pediatrics, 4*(4), 371-376.
- Rumsey, N., & Harcourt, D. (2004). Body image and disfigurement: issues and interventions. *Body image, 1*(1), 83-97.
- Salek, M.S, Jung, S., Brincat-Ruffini, L.A., MacFarlane, L., Lewis-Jones, M.S., Basra, M.K.A., & Finaly, A.Y. (2013). Clinical experience and psychometric properties of the Dermatology Life Quality Index (CDLQI), 1995-2012. *The British Journal of Dermatology, 169*(4), 734-59.
- Salomon, J.A., Tandon, A., & Murray, C.J.L. (2004). Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *BMJ*; doi:10.1136/bmj.37963.691632.44.
- Saxton-Daniels, S. & Jacobe, H.T. (2010). An evaluation of long-term outcomes in adults with pediatric-onset morphea. *Archives of Dermatology, 146*(9), 1044-5.
- Schmid, L. & Leiman, J.M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53-61.
- Schwartz, C.E. (2010). Application of response shift theory and methods to participation measurement: A brief history of a young field. *Archives of Physical Medicine and Rehabilitation; 91*(1), S38-S43.
- Singh, G., Athreya, B. H., Fries, J. F., & Goldsmith, D. P. (1994). Measurement of health status in children with juvenile rheumatoid arthritis. *Arthritis & Rheumatism, 37*(12), 1761-1769.
- Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and response consistency. *Journal of Personality Assessment, 80*(1), 99-103.
- Szkultecka-Debek, M., Drozd, M., Bem, M., Kiepuska, N., & Mazur, J. (2015). Is the vignette method used to assess quality of life in practice? *Current Issues in Pharmacy and Medical Sciences; 28*(1): 8-12.
- Szramka-Pawlak, B., Danczak-Pazdrowska, A., Rzepa, T., Szewczyk, A., Sadowska-Przytocka, A., & Zaba, R. (2013). Health-related quality of life, optimism, and coping strategies in persons suffering from localized scleroderma. *Psychological Health Medicine, 18*(6), 654-63.
- Thissen, D., Liu, Y., Magnus, B., Quinn, H., Gipson, D.S., Dampier, C., Huang, I., Hinds, P.S.,

- Selewski, D.T., Reeve, B.B., Gross, H.E., & DeWalt, D.A. (2015). Estimating minimally important differences (MID) in PROMIS pediatric measures using the scale-judgment method. *Quality of Life Research*; online advanced publication. DOI 10.1007/s11136-015-1058-8
- Torok, K.S., Kurzinski, K., Kelsey, C., Yabes, J., Magee, K., Vallejo, A.N., Medsger Jr., T., & Feghali-Bostwick, C.A. (2015). Peripheral blood cytokine and chemokine profiles in juvenile localized scleroderma: T-helper cell-associated cytokine profile. *Seminars in Arthritis and Rheumatism*; 45(3), 284-293.
- U.S FDA. (2009). *U.S Department of Health and Human Services Food and Drug Administration Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. [<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>]
- Uziel, Y., Laxer, R.M., Krafchik, B.R., Yeung, R.S., & Feldman, B.M. (2000). Children with morphea have normal self-perception. *Journal of Pediatrics*, 137(5), 727-30.
- van der Laerhove, H.J., van der Zaag-Loonen, H.J., & Derkx, B.H.F. (2004). A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatrica*, 93(6), 830-835.
- van der Meer, A., Wulffraat, N.M., Prakken, B.J., Gijsbers, B., Rademaker, C.M., & Sinnema, G. (2007). Psychological side effects of MTX treatment in juvenile idiopathic arthritis: A pilot study. *Clinical and Experimental Rheumatology*, 25, 480-485.
- van Geel, J.M., Maatkamp, M., Oostveen, A.M., de Jong, E.M., Finlay, A.Y., van de Kerkhof, P.C., & Seyger, M.M. (2016). Comparison of the Dermatology Life Quality Index and the Children's Dermatology Life Quality Index in assessment of quality of life in patients with psoriasis aged 16-17 years. *British Journal of Dermatology*, 174(1), 152-7.
- van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J.P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*; 174(3): 575-595.
- Varni, J.W., Limbers, C.A., & Burwinkle, T.M. (2007). How young can children reliably and validity self-report their health-related quality of life? An analysis of 8,591 children across age subgroups with the PedsQLTM 4.0 Generic Core Scales. *Health and Quality of Life Outcomes*, 5(1), 1-13.
- Varni, J.W., Seid, M., Knight, T.S., Burwinkle, T., Brown, J. & Szer, I.S. (2002). The PedsQL in

- pediatric rheumatology: Reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory generic core scales and rheumatology module. *Arthritis and Rheumatism*, 46(3), 714-25.
- Varni, J.W., Sherman, S.A., Burwinkle, T.M., Dickinson, P.E., & Dixon, P. (2004). The PedsQL Family Impact Module: Preliminary reliability and validity. *Health and Quality of Life Outcomes*, 2(55), 1-6.
- Vol., H., Flank, J., Lavoratore, S.R., Nathan, P.C., Taylor, T., Zelunka, E., Maloney, A.M., & Dupuis, L. (2016). Poor chemotherapy-induced nausea and vomiting control in children receiving intermediate or high dose methotrexate. *Support Care Cancer*, 24(3), 1365-71.
- Wand, J. (2013). Credible comparisons using interpersonally incomparable data: Nonparametric scales with anchoring vignettes. *American Journal of Political Science*; 57(1): 249-262
- Wand, J., Gary K., & Lau, O. (2014). anchors: Statistical analysis of surveys with anchoring vignettes. R package version 3.0-8.
<https://CRAN.Rproject.org/package=anchors>
- Watkins, M.W. & Beaujean, A.A. (2014). Bifactor structure of the Wechsler preschool and primary scale of intelligence – Forth Edition. *School Psychology Quarterly*; 29(1), 52-63.
- Weibel, L., Sampaio, M.C., Visentin, M.T., Howell, K.J., Woo, P., & Harper, J.I. (2006). Evaluation of methotrexate and corticosteroids for the treatment of localized scleroderma (morphea) in children. *Paediatric Dermatology*, 155, 1013-20.
- Yung, Y., Thissen, D., & McLeod, L.D. (1999). On the relationship between the higher-ordered factor model and the hierarchical factor model. *Psychometrika*; 64(2), 113-128.
- Zulian, F., Athreya, B.H., Laxer, R., Nelson, A.M., Feitosa de Oliveira, S.K., Punaro, M.G., Cuttica, R., Higgins, G.C., Van Suijlekom-Smit, L.W.A.i, Moore, T.L.j, Lindsley, C., Garcia-Consuegra, J., Esteves Hilário, M.O., Lepore, L.n Silva, C.A., Machado, C., Garay, S.M., Uziel, Y., Martini, G., Foeldvari, I., Peserico, A., Woo, P., & Harper, J. (2006). Juvenile localized scleroderma: Clinical and epidemiological features in 750 children. An international study. *Rheumatology*, 45, 614-620.
- Zulian, F., Vallongo, C., Woo, P., Russo, R., Ruperto, N., Harper, J., Espada, G., Corona, F., Mukamel, M., Vesely, R., Musiej-Nowakowska, E., Chaitow, J., Ros, J., Apaz, M.T., Gerloni, V., Wouters, C., Martini, G., Cimaz, R., Laxer, R., & Athreya, B.H. (2005). Localized scleroderma in childhood is not just a skin disease. *Arthritis & Rheumatism*; 52(9), 2873-2881.
- Zulian, F., Vallongo, C., Patrizi, A., Belloni-Fortina, A., Cutrone, M., Alessio, M., Martino, S.,

- Gerloni, V., Vittadello, F., Martini, G. (2012). A long-term follow-up study of methotrexate in juvenile localized scleroderma (morphea). *Journal of the American Academy of Dermatology*, 67(6), 1151-6.
- Zumbo, B.D., Gaderman, A.M., & Zeisser, C. (2007). Ordinal versions of coefficient alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21-29.