

CLEFT-Q: A PATIENT-REPORTED OUTCOME MEASURE

CLEFT-Q: DEVELOPMENT OF A PATIENT-REPORTED OUTCOME
MEASURE TO PROVIDE CLINICALLY MEANINGFUL OUTCOMES IN
PATIENTS WITH CLEFT LIP AND/OR PALATE

By KAREN W.Y. WONG, M.D., M.Sc., B.Sc.Hon., FRCSC

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2017)

Hamilton, Ontario (Health Research Methodology)

TITLE: CLEFT-Q: Developing a Patient-Reported Outcome Measure to Provide Clinically Meaningful Outcomes in Patients with Cleft Lip and/or Palate

AUTHOR: Karen W.Y. Wong, M.D. (University of Toronto), M.Sc. (University of Toronto), B.Sc.Hon. (University of Toronto), FRCSC

SUPERVISOR: Professor Anne F. Klassen, DPhil

COMMITTEE: Professor Peter Rosenbaum, M.D., B.Sc., FRCPC; and Professor Achilles Thoma, M.D., B.Sc., FRCSC, FACS

NUMBER OF PAGES: xix, 190

LAY ABSTRACT

Measuring outcomes of treatment for cleft lip and/or palate (CL/P) should include the patient perspective. The objective of this thesis is to show that through rigorous methods of development, a patient-reported outcome (PRO) measure, the CLEFT-Q, can provide clinically meaningful evaluation of outcomes. First, 136 patients with CL/P from six countries were interviewed to learn what concepts related to having a cleft or its treatment are important to them. A conceptual framework was developed that informed the CLEFT-Q scales. Describing the methodology behind developing the CLEFT-Q then served to inform and engage members of the community. A field-test of the CLEFT-Q scales showed that in a sample of 2,434 patients with CL/P from twelve countries, CLEFT-Q outcomes varied in patients with different types of CL/P. The CLEFT-Q can be used to provide rigorous measurement of PROs in patients with CL/P in the future.

ABSTRACT

Background: The management of cleft lip and/or palate (CL/P) includes multidisciplinary care beginning in infancy and continuing through to adulthood. Outcomes of cleft care have been difficult to measure because of the subjective nature of evaluating concepts such as appearance and speech. Including the patient perspective in outcome evaluation through the use of a patient-reported outcome (PRO) measure would provide a more accurate reflection of a patient's status. The overall objective of this thesis is to show that through adherence to rigorous methods of development, a PRO measure can provide clinically meaningful outcome evaluation in cleft care.

Methods: The first paper uses the qualitative method of interpretive description to define a conceptual framework to guide the development of a PRO measure for patients with CL/P, the CLEFT-Q. The second paper describes the protocol for the entire development of the CLEFT-Q. The third paper analyzes the results of the cross-sectional field-test of the CLEFT-Q scales to determine whether or not the CLEFT-Q is able to detect differences between specific cleft types.

Results: The qualitative study included 138 patients with CL/P from six countries. The final conceptual framework contained thirteen concepts within the domains of appearance, facial function, and health-related quality of life. The second paper details the process of designing the CLEFT-Q scales. The field-test included

2,434 patients from thirty sites in twelve countries, and CLEFT-Q scores were found to vary with cleft type for all scales.

Conclusions: PRO measures need to be rigorously designed in order to provide scientifically sound, clinically meaningful measurement. The CLEFT-Q is able to detect differences between patients with various cleft types, and will be a useful tool to provide the patient perspective in future outcome evaluation in cleft care.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Anne Klassen for believing in the story from the beginning, for believing in me, and for showing me what it takes to be a successful researcher and a successful mother with poise and without pretense.

I would like to thank my committee members Dr. Peter Rosenbaum and Dr. Achilles Thoma for supporting me throughout this process and for their invaluable insight into the research and clinical worlds.

I would like to thank my mentor Dr. Christopher Forrest for inspiring my interest in pediatric plastic surgery, for nurturing my interest in research, and for the encouragement to continually think and dream about how to make things better.

I would like to thank my mentor Dr. David Fisher for taking me along on the first trip to India that changed my career path, for continuing to be a source of wisdom ever since, and for showing me the artistry of surgery.

I would like to thank the patients who participated in this study for sharing their experiences with us.

Collaboration with clinicians from around the world was perhaps the greatest strength of this study. Thank you to all of the collaborators for being part of our team.

Of the many things I have learned over the course of this degree that are not meant to be written into academic reports, the most poignant is that family members will go to the greatest lengths to take care of each other, and this love serves to cushion against the inevitable challenges that are part of life. I am beyond fortunate to have had the love and support of my family. Thank you for the motivation, encouragement, and strength.

TABLE OF CONTENTS

CHAPTER 1	1
Cleft Lip and/or Palate	2
Setting the Context.....	3
Patient-Reported Outcome Measures	7
How Do PRO Measures Work?	8
Reliability	9
1. Internal Consistency.....	9
2. Reproducibility	10
3. Measurement Error	10
Validity	10
1. Content Validity	11
2. Criterion Validity	12
3. Construct Validity.....	13
Responsiveness	13
Psychometric Methods.....	14
Traditional Psychometric Methods	14
Modern Psychometric Methods	15
Knowledge Gaps in the Measurement of Outcomes in Patients with CL/P	16
1. The Need for a Patient-Reported Outcome Measure for Patients with CL/P	16

2. The Need to Understand the Science Behind Patient-Reported Outcomes	19
3. The Need to Make PRO Measurement Clinically Meaningful	21
The CLEFT-Q Project	23
Thesis Objectives and Structure	24
Chapter 2: What Should We Measure?	24
Chapter 3: How was the CLEFT-Q developed?	25
Chapter 4: How do CLEFT-Q outcomes vary by cleft type?	26
Chapter 5: Discussion	26
References	27
CHAPTER TWO	48
Preface	48
Abstract	52
Introduction	54
Methods	56
Sampling	57
Data Collection	58
Data Analysis	59
Reflexivity	60
Results	60
Appearance	62
Health-Related Quality of Life	65

Psychological	65
Social and School.....	66
Speech-Related Distress.....	67
Facial Function.....	68
Speech	68
Eating/Drinking.....	69
Other Facial Functions	70
Discussion.....	70
Conclusion	74
References	80
CHAPTER THREE	85
Preface.....	85
Abstract.....	88
Introduction.....	91
Methods and Analysis.....	93
Phase 1: What Should We Measure?	94
Conceptual Framework.....	94
Qualitative Study	95
Participants, Setting, and Recruitment.....	95
Sampling	96
Data Collection	96
Data Analysis	97

Rigor	97
Item Generation	98
Refining the Preliminary Scales.....	98
Cognitive Debriefing Interviews.....	99
Expert Clinician Input.....	99
Translation	100
Phase 2: What questions are effective in measuring the concepts identified in Phase 1?	101
Pilot Field-Test.....	102
Study Participants	102
Data Collection	103
Data Analysis	103
International Field-Test and RMT Analysis	103
Study Participants	104
Data Collection	104
Data Analysis	104
Differential Item Functioning (DIF)	106
Item Reduction.....	106
Normative Data and Construct Validity	107
Phase 3: How does the instrument work?	108
Reliability.....	108
Validity	108

Responsiveness	109
Study Participants	110
Data Collection	110
Data Analysis	110
Test-Retest Reliability	110
Criterion Validity	111
Responsiveness	111
Ethics and Dissemination	111
References	116
CHAPTER FOUR.....	124
Preface.....	124
Abstract.....	128
Introduction.....	130
Methods.....	132
Study Participants	132
CLEFT-Q Scales	133
Data Collection	133
Data Analysis	134
Visualization of CLEFT-Q Scores.....	134
Results	135
Appearance Scores.....	136
Speech Function and Speech-Related Distress Scores	136

Health-Related Quality of Life Scores.....	137
Visualizing CLEFT-Q Scores.....	138
Discussion.....	138
Conclusions.....	142
References.....	166
CHAPTER FIVE.....	173
Discussion.....	174
Strengths and Limitations.....	179
Knowledge Translation.....	181
Conclusions.....	182
References.....	183

LIST OF FIGURES AND TABLES

Chapter 2

Figure 1: Protocol for a multiphase mixed methods study to design the CLEFT-Q.....	75
Figure 2: The preliminary and final conceptual frameworks of concepts of interest to patients with CL/P.....	76
Table 1: Interview guide for qualitative study.....	77
Table 2: Characteristics of the participants interviewed.....	78
Table 3: An example of how the qualitative data were analyzed and categorized.....	79

Chapter 3

Figure 1: The phases of PRO development.....	113
Figure 2: Flow diagram of the multiphase mixed methods protocol for developing the CLEFT-Q.....	114
Table 1: Demographic characteristics collected in the international field-test.....	115

Chapter 4

Table 1: Demographic characteristics of CLEFT-Q field-test participants....	143
Table 2: Distribution of cleft types in the sample population.....	144
Table 2a: Distribution of overall cleft types.....	144

Table 2b: Distribution of patients with cleft palate using the Veau classification of cleft palate severity.....	144
Table 3: Mean CLEFT-Q scores by cleft subtype.....	145
Figure 1: Boxplots showing the distribution of scores for each CLEFT-Q scale.....	147
Figure 1a: Boxplot of CLEFT-Q Face scores by cleft type.....	147
Figure 1b: Boxplot of CLEFT-Q Nose scores by cleft type.....	148
Figure 1c: Boxplot of CLEFT-Q Nostrils scores by cleft type.....	149
Figure 1d: Boxplot of CLEFT-Q Lips scores by cleft type.....	150
Figure 1e: Boxplot of CLEFT-Q Cleft Lip Scar scores by cleft type.....	151
Figure 1f: Boxplot of CLEFT-Q Teeth scores by cleft type.....	152
Figure 1g: Boxplot of CLEFT-Q Jaws scores by cleft type.....	153
Figure 1h: Boxplot of CLEFT-Q Speech Function scores by cleft type.....	154
Figure 1i: Boxplot of CLEFT-Q Speech-Related Distress scores by cleft type.....	155
Figure 1j: Boxplot of CLEFT-Q Social scores by cleft type.....	156
Figure 1k: Boxplot of CLEFT-Q School scores by cleft type.....	157
Figure 1l: Boxplot of CLEFT-Q Psychological scores by cleft type.....	158
Table 4: Statistics for one-way ANOVA or Kruskal-Wallis H tests evaluating the differences in CLEFT-Q scores between groups with different cleft types.....	159
Figure 2: Speech function and speech-related distress in patients with cleft	160

palate by Veau type.....	
Figure 2a: Boxplot of CLEFT-Q Speech Function scores by Veau type.....	160
Figure 2b: Boxplot of CLEFT-Q Speech-Related Distress scores by Veau type.....	161
Figure 3: Scores for all CLEFT-Q scales by cleft subtypes, grouped in to broader cleft types.....	162
Figure 3a: Scores for groups with all CLP subtypes.....	162
Figure 3b: Scores for patients with CP.....	163
Figure 3c: Scores for groups with all CL subtypes.....	164
Figure 3d: An example radar chart with scores from one 14 year old male patient with CLP-UC, shown with the mean scores for this subtype.....	165

LIST OF ABBREVIATIONS

A	asymmetric
ACPA	American Cleft Palate-Craniofacial Association
ANOVA	analysis of variance
B	bilateral
C	complete
CHASQ	Cleft Hearing, Appearance and Speech Questionnaire
CL	cleft lip
CL/P	cleft lip and/or palate
CLA	cleft lip and alveolus
CLP	cleft lip and palate
COA	clinical outcome assessment
COHIP	Child Oral Health Impact Profile
COI	concept of interest
CONSORT	Consolidated Standards of Reporting Trials
COSMIN	Consensus-based Standards for the Selection of Health Status Measurement Instruments
CP	cleft palate
CSAG	Clinical Standards Advisory Group
CTT	Classical Test Theory
DALY	disability-associated life year
DIF	differential item functioning

FDA	Food and Drug Administration
HR-QOL	health-related quality of life
I	incomplete
IRT	Item Response Theory
ISOQOL	International Society of Quality of Life Research
LMIC	low- and middle-income country
MID	minimally important difference
NZ	New Zealand
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	patient-reported outcome
QUAL	qualitative
QUAN	quantitative
RMT	Rasch Measurement Theory
U	unilateral
UK	United Kingdom
USA	United States of America

DECLARATION OF ACADEMIC ACHIEVEMENT

I, Karen Wong, was the main contributor and primary author for all studies. The details of my and other authors' contributions are provided at the start of each chapter. The work presented in the thesis forms part of a large international study that I conceived of and designed together with the study lead, Dr. Anne Klassen.

CHAPTER 1

Introduction

Cleft Lip and/or Palate

Cleft lip and/or palate (CL/P) is a congenital malformation resulting in an abnormal connection between the mouth and the nose(1). CL/P is the most common craniofacial malformation, with an overall prevalence of 7.94 and 4.50 cases, per 10,000 live births, of cleft lip with or without cleft palate and isolated cleft palate, respectively(2, 3). While incidence may be a more useful indicator, previous studies have reported on prevalence. A population-based study evaluating birth defect registries from 30 countries between 2000 and 2005 reported an overall prevalence of CL/P of 9.9 per 10,000 live births with significant geographic variation(4). The etiology of CL/P is likely multifactorial, with maternal smoking and alcohol intake having been associated with an increased risk, and maternal multivitamin use around the time of conception associated with decreased risk(5). There is also a hereditary component, but the pattern of penetrance is variable(1). A birth cohort study is underway in the UK that will attempt to identify any genetic predispositions(6). Beyond these findings and ongoing studies, the large volume of literature investigating the causes of CL/P is perhaps indicative of our continued lack of in-depth understanding as to why CL/P occurs(1, 7).

Every cleft is different, with some presenting with just a mild notch in the lip and others with a complete gap in the lip, nose, and palate(1, 8). CL/P affects a patient's appearance, speech, feeding, dentition, hearing, and health-related quality of life. Treatment begins at birth and can continue through to adulthood.

Treatment protocols vary widely; in a study of European cleft centres, 194 different protocols were used in 201 centres for a single cleft subtype(9). For a patient with a complete cleft of the lip and palate, the bare minimum treatment would include surgery to repair the lip, another surgery to repair the palate, and a third surgery to fill the gap in the gums, or alveolus. Additional interventions can include feeding assessments at birth, surgery to drain fluid from the ear canals, dental work, orthodontic treatment, speech therapy, surgery to improve the appearance of the nose or lip, surgery to improve speech, and surgery to place the jaws in a better position once the patient has completed facial growth(10). The goal of treatment is to optimize appearance and function for patients, ideally to eliminate any stigmata of CL/P by the time treatment is complete.

Setting the Context

My first exposure to CL/P came during residency training in plastic and reconstructive surgery. I began my senior rotations at the Hospital for Sick Children in Toronto, Canada, and shortly after starting, I happened to sit down with one of the surgeons who sub-specialized in cleft care. I mentioned that I was scheduled to work with him in a few months, and he responded that he would be out of the country during that time. He was going to India on a surgical mission with Operation Smile. Disappointed that I would be missing out on learning from him while he was away doing exactly what I was supposed to be learning, I asked

if I could join him. A few months later, thanks to a fellowship from Operation Smile, I boarded a plane to India excited for a concentrated exposure to CL/P.

India is a country that triggers the senses much differently from the relatively quiet and underpopulated streets of Toronto. We landed in Kolkata in the early morning and set out to explore the city in an attempt to overcome jetlag. I was walking with a group of five surgeons: my mentor, two other surgeons with whom he regularly travelled on such missions, one of my mentor's former trainees, and another resident like myself. We walked in silence for most of the day, overwhelmed by the sounds of traffic unlike any snarled highways I'd ever seen, the bright colours of beautiful textiles and clothing, and the pungent smells of local markets with fresh meat hanging in the thirty degree heat. The day introduced the sensory overload that would continue throughout the trip; the experience was not just going to be a concentrated technical exposure to cleft surgery, but rather a sudden immersion into what it was like for people to grow up with a cleft. Almost none of what I learned was in any textbook I had read, and yet what I learned seemed to be the most important information for a cleft care provider to understand.

That night, the six of us settled into one of our hotel rooms and the experienced surgeons began telling stories of previous missions they had gone on together. They didn't speak about technical results or how they thought patients' appearances improved with surgery. Rather, they spoke of the teenaged girl who was beyond grateful after having her cleft lip repaired because she could now

possibly get married; the speechless mother who burst into tears upon seeing her child with a repaired cleft lip for the first time; and the young boy with a cleft palate who worked in a farmer's field because he couldn't speak properly and didn't go to school as a result. I was astounded hearing these stories; growing up in a country with universal health care, CL/P was a condition that was treatable from infancy. While obtaining an ideal result was still difficult, I had never seen a patient old enough to walk who still had an unrepaired cleft lip. The passion with which the surgeons described these cases explained why they kept returning to these missions. They understood firsthand that giving these patients access to treatment changed lives. When we finally started working a few days later in a small village in the countryside, hundreds of people were waiting in line at the makeshift clinic, a large tent in a field outside of the hospital. There was clearly an unmet need for cleft care. If there were this many people who had travelled to a small village to try and seek care, how many more were living with unrepaired clefts in the country? We worked into the night over ten days, operating on as many patients as possible, and I gained experiences and stories of my own to be told to a new trainee on the first day of a future mission.

After returning home with a deeper understanding of the impact of both having a cleft and its treatment, I expressed my newfound enthusiasm to a surgical fellow who finished her plastic surgery training in Boston. She told me that missions like the one I had just returned from were widely criticized by some in the surgical community(11). Our patients at home were followed closely for 18

years by a multidisciplinary team, but these patients overseas had one visit and were often then lost to follow-up. Critics felt that it wasn't worth doing operations when the patients didn't have access to the rest of the team(11). Was repairing a palate a waste of time when patients didn't undergo speech therapy afterwards? I went back to my mentor with these arguments, curious to know what his rebuttal would be. My mentor turned thoughtful as I was explaining the criticism of surgical missions. Rather than counter each point, he told a story.

An old man walked along a beach every morning. Hundreds of starfish would wash up and be stranded overnight. He would see a little boy pick up the starfish one by one, and throw them back into the ocean. One day, he asked the boy why he was bothering. There was no way that he would save them all, so what he was doing really didn't matter. The boy picked up a starfish, threw it into the ocean, and responded, "*It mattered to that one.*"(12)

What my mentor, his colleagues, and now I also recognized was that patients were grateful for their treatment and that it changed their lives. Having treatment mattered to them. In searching for literature describing the impact of treatment, several economic analyses of cleft surgery in low- and middle-income countries (LMICs) have been performed, but these studies depend on the measurement of disability-associated life years (DALYs)(13-19). While DALYs are useful for comparisons between conditions in population-level studies, this measurement does not provide an indication of the success, or quality, of treatment. The criticism of surgical missions was that the outcomes of surgery

may not be as successful as they are in a high-resource setting, but this was difficult to prove using DALYs. What was lacking in the literature was the measurement of the impact of surgery on patients; if we are operating to improve speech, we should be measuring an improvement in speech, and if we are operating to improve appearance, we should be measuring an improvement in appearance. My colleagues saw firsthand what the impact of improved speech and improved appearance was on a patient's quality of life, but this impact was not being measured and reported. After turning to all literature on outcomes of cleft care, it became clear that the measurement of success of treatment was problematic everywhere, not just in low- and middle-income countries.

“...I have been struck again and again by how important measurement is to improving the human condition.”

– Bill Gates, Annual Letter 2013(20)

Patient-Reported Outcome Measures

The U.S. Food and Drug Administration (FDA) describes four types of outcomes that can be measured: patient-reported outcomes (PROs), clinician-reported outcomes, observer-reported outcomes, and performance outcomes(21). PROs are “reports that come directly from patients about how they function or feel in relation to a health condition and its therapy, without interpretation by a physician or anyone else”(22) and are evaluated using PRO measures. Psychometrics, the science of measuring mental traits or processes, underpin the ability of PRO measures to reliably ‘turn feelings into numbers’. When this

science is not taken into consideration in the development of a PRO measure, as is often the case when ad hoc PRO measures are used, there is a risk that the resulting measurements are neither reliable nor valid. Unpublished rating scales have been shown to reflect a considerable source of bias; in a study of 300 randomized controlled trials in schizophrenia, the use of an unpublished rating scale resulted in a 37 per cent increased chance of reporting a treatment effect compared to studies that used published scales(23). The field of psychometrics, which initially evolved in the study of education and psychology, has become more relevant in health outcomes measurement as the focus of evaluation of outcomes has shifted to place more importance on the patients' perspectives(24).

How Do PRO Measures Work?

PRO measures can be thought of as rulers, where on one end of the scale, there is a small amount of the concept being measured, and on the other end of the scale, there is a large amount. The way that the ruler is designed determines how well it will measure the concept of interest. Quality criteria for health measurement scales have been proposed by several organizations(25-31). These criteria fall under the main categories of reliability, or the ability of the measure to perform the same way under different conditions; validity, or how well the measure evaluates what it is intended to evaluate; and responsiveness, or how well the measure is able to capture change if that is its intended purpose(31).

Reliability

Three main concepts determine the overall reliability of a measure: internal consistency, reproducibility, and measurement error(24, 30, 31).

1. Internal Consistency

Internal consistency is an indicator of how much the items in a scale are intercorrelated, or how much they measure the same concept. An important determinant of internal consistency is whether a PRO measure evaluates a single concept. If we conceptualize a PRO measure as a ruler meant to measure mobility, we can imagine that on one end of the scale, a person is less mobile, and at the other end, they are more mobile. However, if the ruler is designed to measure ‘recovery’ instead, with factors including mobility and pain, the scores become difficult to interpret. If a patient improves on this ruler, is it their mobility that is increasing or pain that is decreasing? If a patient becomes more mobile but also has more pain, they may show no change on the recovery ruler. Critical appraisal of PRO measures in the literature shows that scales often measure more than one concept. For example, the Pediatric Voice Outcome Survey includes four questions asking about being understood by others, social activities, swallowing, and straining(32). These four questions address very different concepts that may be related to voice outcomes overall, but are in fact distinct from each other. In clinical use, it would not be clear what aspect of outcome is changing as a patient moves up or down this ruler. Applying this reasoning to outcomes of treatment for CL/P, using a single score to show a ‘total’ result does not impart clinically useful

information. There are some treatments in cleft care that improve one aspect but may be detrimental to another. For example, there is ongoing debate as to whether early palate repair may improve speech outcomes but worsen appearance. In such cases, individual scores for each concept (for the example given, individual scores for speech and appearance) reflect a patient's status more accurately. When a scale measures a single concept, it is termed 'unidimensional'(33).

2. Reproducibility

Three tests evaluate the ability of a measure to provide reproducible results(31). Test-retest reliability assesses whether or not the same person will score similarly on two different occasions. Inter-rater reliability assesses the level of agreement of two or more observers who provide ratings on the scale. Intra-rater reliability assesses how closely the same observer will provide similar ratings on the scale on different occasions.

3. Measurement Error

Measurement error around each of the three tests for reliability listed above provides an indication of the precision of measurement. This kind of error is most commonly reported as standard error of measurement.

Validity

The ability of a measure to evaluate what it is meant to evaluate is classified into several categories of validity: three main ideas are content validity, criterion validity, and construct validity.

1. Content Validity

Perhaps the most important criterion in determining the appropriateness of a PRO measure, content validity refers to the extent to which the content of a PRO measure adequately represents the concept of interest(27, 28, 30). Eliciting patient input at various stages of development of a measure ensures that the content is comprehensive and valid, and that the questions are understood by the target population. A separate type of validity, face validity, is also categorized as being part of content validity. Face validity refers to a typically subjective assessment of the measure to determine if it appears to be measuring the concepts of interest(24).

It is essential to consider the clinical or research question when evaluating content validity. If the purpose of measurement is to compare two populations, the content of the measure will be different than if a study aims to discriminate between patients within a specific population. There are two main types of PRO measures: generic measures aim to compare populations, and condition-specific measures aim to discriminate within a population. The questions included in a condition-specific measure may not be relevant to patients in other populations, and likewise the questions in generic measures may not capture issues that are relevant to a specific population. If a PRO measure is planned for use in the context of evaluating treatment outcomes, a condition-specific measure is more likely to capture the concepts of interest to that population, and thus have stronger content validity.

Returning to the example of mobility, if a scale for mobility was initially designed for use in infants learning how to walk, it is easy to imagine that this scale would not be appropriate to use in athletes recovering from an injury. Differences in populations may not be as obvious as they are in this example, but it is important to consider that even within seemingly similar populations such as all patients with CL/P, there may be important differences. Patients from high-income and high-resource settings may not be the same as patients from low- and middle-income countries with less access to care. In order for a scale to be used across countries, it should ideally be designed and tested in a variety of countries in order to maximize content validity.

2. Criterion Validity

Criterion validity refers to the extent to which a measure correlates with a ‘gold standard’ that has been used previously(24, 30). While the literature often employs the term ‘gold standard’, it is perhaps more appropriate to refer to the evaluation of criterion validity using *existing* standards. Two types of criterion validity exist: concurrent validity, where a measure can be administered at the same time as another measure thought to assess similar issues in order to determine the correlation between the scores; and predictive validity, where a measure is administered first, and the extent to which the scores are predictive of a ‘gold standard’ outcome in the future is evaluated. Of course, criterion validity presupposes that there is a ‘gold standard’, whereas in fields like ours there rarely is such a tool.

3. *Construct Validity*

Construct validity assesses the extent to which a measure works the way it is ‘expected’ to, based on what we know of the phenomenon of interest(34). In the first description of construct validity, three steps are required. First, a theoretical framework of concepts of interest and how they relate to each other is required. This component is termed structural validity. Second, scales must be developed to measure the concepts. Third, the relationships between the concepts must be empirically tested based on *a priori* hunches, a process termed hypothesis-testing. Subtypes of validity related to hypothesis testing include convergent validity, whereby scores correlate to other measures that are hypothesized to be correlated because they assess somewhat similar issues, and divergent validity, whereby scores differ from other measures that are hypothesized to be different. Recently, an additional category of cross-cultural validity was proposed within construct validity to determine how similarly the measure performs in different cultural and linguistic contexts(31).

Responsiveness

Responsiveness refers to the ability of a measure to detect change over time, and is relevant only to measures specifically designed to be ‘evaluative’ measures. Testing responsiveness is similar to hypothesis-testing for construct validity, except the hypotheses relate to how patients are expected to change with treatments or over the natural course of the condition(30).

Psychometric Methods

Traditional Psychometric Methods

The predominant traditional psychometric method used in the design of rating scales is Classical Test Theory (CTT). CTT works on the assumption that the observed score is equal to the true score plus error, where the error is not correlated to the true score(24, 35).

Limitations of CTT include sample dependency, where the measures of reliability and validity apply only to the group of patients that filled out the scales(24, 35). If the scales are then used in a different group of patients, the norms and measures of reliability and validity also change, making the establishment of true population norms difficult, if in fact the original measure was meant to be referenced to norms.

CTT also assumes item equivalence, meaning that each item on the scale contributes equally to the total score regardless of how well the item correlates with the concept of interest(24, 35). Scales designed with CTT are often scored by simply adding the responses for each item to obtain a total score. This also assumes that the scale provides interval-level measurement when the items actually provide ordinal-level measurement. Returning once again to the example of mobility, we would assume that the notches on the ruler are evenly spaced apart, so an individual who moves up on the scale one notch at the lower end of the ruler is gaining the same amount of mobility as an individual who moves up the scale one notch at the higher end of the ruler. In ordinal-level measurement,

the notches are not evenly spaced, so the amount of mobility required to move up a notch on one part of the ruler does not necessarily equal the amount required to move up a notch on another part of the ruler. This creates an issue not only with monitoring change, but also in statistical analyses of scale scores that often assume interval-level measurement.

CTT assumes that the standard error of measurement is equal along the entire scale, and equal for all individuals that completed the scale(24, 35). Logically, these assumptions cannot hold true. The standard error of measurement also depends on the group of individuals tested rather than the individual alone. In this way, scales designed using CTT are not as well suited to placing an individual precisely at a point on the scale.

Modern Psychometric Methods

Modern psychometric methods have evolved to address the limitations of CTT. Rasch Measurement Theory (RMT)(36) and Item Response Theory (IRT)(37) were both developed in the 1960s to provide more focus at the item level rather than the scale level. Two main assumptions underpin modern psychometric methods(24, 33). First, it is assumed that the scale is unidimensional. Second, local independence is assumed, meaning that the probability that an individual responds to one item a certain way is unrelated to the probability of responding to another item a certain way. An individual's score is based on the probability that they responded to each of the items in a particular way(35). Items on scales form a hierarchy, in a sense creating notches on the

ruler, making it possible to place an individual on a precise point on the scale. The mathematical models underpinning modern psychometric methods also allow for scale invariance, where the scale performs the same way regardless of the population in which it was administered. In RMT, the mathematical model for scale performance is fixed, and the data must fit the model. If the data do not fit, the scale must be examined more closely in the development phase to determine if the items on the scale are in fact relevant and appropriate for evaluating the underlying construct. In this way, RMT provides rigorous standards for measurement.

Knowledge Gaps in the Measurement of Outcomes in Patients with CL/P

1. The Need for a Patient-Reported Outcome Measure for Patients with CL/P

What types of outcomes have been measured in cleft care? Just as protocols for the treatment of clefts abound, there are many different outcomes reported in the literature, making inter-centre comparisons difficult(38, 39). The largest multi-centred trials to date, the Eurocleft, CSAG, Americleft, Cleft Care UK, and Scandcleft studies(9, 40-73), have used primarily clinician-reported outcomes. The CSAG study led to the centralization of cleft care in the UK after low-volume centres were found to have poorer outcomes compared to high-volume centres(49-52). The Cleft Care UK study served as a follow-up study and was carried out using similar outcomes 15 years after the CSAG study(58-62). Panels of clinicians evaluated appearance and dental occlusion in the five studies.

Neither the Eurocleft nor Americleft studies included speech outcomes, and the CSAG/Cleft Care UK and Scandcleft studies used different measures to evaluate speech(51, 61, 74). The Eurocleft and CSAG studies employed ad hoc questionnaires to provide the patient perspective (47, 50), and the Cleft Care UK and Scandcleft studies assessed parent perspective only, likely because the studies included patients who were 5 years of age and considered too young to self-report (62, 71, 72).

A narrative review of studies assessing the psychological impact of CL/P published between 2004 and 2015 identified 148 studies that employed both quantitative and qualitative methods (75). This review was an update to a previous review (76), and both studies concluded that the variation in outcomes measured made it difficult to compare results between individual studies.

If the goal of cleft care is to improve appearance and function for the patient, ideally the patient's perspective would be an important component of evaluating outcome. The clinician-reported outcomes used thus far are not consistently agreed upon. Appearance in particular is a difficult concept to measure objectively. While efforts to develop better methods of evaluation continue, many of these methods are not fully validated(77). In addition, clinician-reported aesthetic outcome and self-reported satisfaction with appearance have been shown to differ(78). Asking the patient would not only provide perhaps the most important opinion of all, but would also avoid potential limitations with clinician-reported outcomes.

Existing literature has found that patients with CL/P are dissatisfied with the appearance of the nose, lip, teeth, and facial profile(44, 79-90). Studies have shown that, overall, patients with CL/P are less satisfied with their appearance compared to patients without CL/P(80, 81, 83, 85, 88, 91-97). In contrast, other studies have found that patients with CL/P are more satisfied with their overall appearance compared to control groups, and more satisfied with parts of the face not affected by CL/P such as the eyes, ears, and hair(86, 98). Studies have found that patients are overall reasonably satisfied with their appearance(88, 93, 99-102), that a majority of patients find their face beautiful(103), and that patients are more satisfied with their appearance compared with patients with clubfoot(104). The association of gender with satisfaction with appearance has been examined, with some studies showing that males are less satisfied(44, 83, 90, 105), some showing that females are less satisfied(82, 106), and others showing no association(81, 84, 107, 108). Likewise, some studies have shown an association between age and satisfaction with appearance(81, 109), whereas others have found none(84, 108).

A review of the methodology of these studies shows that all of them employed ad hoc measures or measures that have not been thoroughly validated to evaluate appearance, particularly in patients with CL/P. While this does not necessarily discount the findings, it is likely that using a rigorously designed measure to evaluate appearance would provide more accurate (and hence probably more valid) measurement, which may result in fewer conflicting results in the

future. In addition to being useful in clinical care, a PRO measure could serve as a useful tool in tandem with future economic evaluations such as cost-effectiveness analyses or studies of utility.

A systematic review aimed at identifying validated measures used in pediatric plastic surgery found none designed specifically for patients with CL/P(110), and other reviews have found that there is a need for a cleft-specific measure(111, 112). It is clear that the problems with measurement in cleft care extend beyond low- and middle-income countries; in order to begin to report accurately the impact of both having a cleft and its treatment that my colleagues recognized, we needed to develop a tool to measure this impact.

2. The Need to Understand the Science Behind Patient-Reported Outcomes

As health care moves towards becoming more patient-centered, PROs play a prominent role in determining a patient's status before and after treatments(113). Remuneration for health care is shifting towards a value-based model, where value is often defined from the patient perspective(114). As PROs are recognized as being key components of overall outcomes, we must ensure that the PRO measures that are used to evaluate these outcomes are both clinically meaningful and scientifically sound.

Clinicians who are unfamiliar with the field often ask if it is really possible to 'turn feelings into numbers', and are wary of using anything but 'hard' (i.e., objective) outcomes. In my short career to date as a pediatric plastic surgeon, I have found that PRO measurement is not well understood in the clinical realm.

Moving forward in this new era of patient-focused measurement with a lack of knowledge about how measurement works puts the system at risk for measuring improperly, a potentially great disservice to both patients and providers.

Misconceptions about PROs abound; for example, many clinicians believe that PROs encompass only psychosocial outcomes, rather than including all aspects of outcome that are important to patients. Others believe that selecting any measure that seems ‘close’ to an ideal measure is sufficient, or that a series of ad hoc questions can be turned into a scale and analyzed as such. These misconceptions stem from a lack of understanding of how measures work (including PROs), and the science behind the measurement, or psychometrics. Informing clinicians and other stakeholders on the basic methodology behind PRO measurement is important in order to engage the community in scientifically sound measurement.

As an example, a recent article in the plastic surgery literature questions the weight that has been placed on patient satisfaction in the evaluation of outcomes(115). The authors cite anecdotes of emergency room physicians prescribing antibiotics only to improve patient satisfaction, or family physicians avoiding conversations about smoking or exercise in order to minimize negative reviews. These are sequelae of poor or incomplete measurement; perhaps patient satisfaction with the process of care needs to be evaluated concurrently with patient satisfaction with their clinical result in order to achieve a more accurate reflection of overall outcome.

Similar to the need for research studies to have pointed, specific research questions that use equally specific outcomes to answer these questions, the evaluation of health care needs to use appropriate measures. High-level organizations have recommended the inclusion of PROs in evaluating outcomes of care(116, 117). Selecting the right PRO measures, and selecting PRO measures that are scientifically sound, should help to address some of the concerns with incorporating the patient perspective into outcome evaluation.

3. The Need to Make PRO Measurement Clinically Meaningful

A systematic review of the literature evaluating the impact of measuring PROs in clinical practice found that there may be an effect on health care, particularly process variables, but defining this impact was difficult due to methodological limitations(118). In a call for standardized measurement of patient-important outcomes, Porter et al. discuss the lack of true outcomes (7%), let alone patient-reported outcomes (<2%), in the National Quality Measures Clearinghouse, and argue that the current focus of measurement needs to shift beyond only compliance with evidence-based practice to include patient-important outcomes(119).

In cleft care, measurement of PROs has typically been done in the context of research studies rather than routine clinical care (112, 120). One of the difficulties in engaging stakeholders to implement routine measurement of PROs may be the fact that previous studies have shown conflicting results with respect to the impact of CL/P on quality of life. Patients with CL/P have been found to

have similar scores to control groups on measures evaluating health-related quality of life(84, 121-123). These studies employed generic measures such as the SF-36(121), the Rand 36(84), the KINDL(122), and a generic oral health-related quality of life measure, the COHIP(123). Having CL/P has also been shown to have a minimal effect on health-related quality of life using similar measures (100, 124-131). Other studies have identified an impact of having a cleft(132, 133). An extensive body of literature has also assessed self-concept, self-esteem, social functioning, and emotional functioning in the context of CL/P(75, 76). Similar to the evaluation of appearance in the past, there have been no condition-specific PRO measures used. For the purposes of research studies or the comparison of populations of patients, it is perhaps useful to employ generic measures; however, it is unlikely that any of these PRO measures would be incorporated into clinical practice as the results do not impart information that will help in treatment decision making or reflect change over time.

In order for an evaluative PRO measurement to be clinically meaningful, it must accurately impart information about a patient's health status in a way that is sensitive to change with treatment. There is some evidence that PRO measures are capable of evaluating change in patients with CL/P; for example, a recent study of patients with CL/P showed that oral health-related quality of life improved after surgery(134). The instrument used in this study, the Child Oral Health Impact Profile (COHIP), was not developed specifically for use in patients with CL/P and

asks only basic questions about the appearance of the face besides the teeth and speech.

Since CL/P presents with a spectrum of severity, it is also important for PRO measurements to account for variation due to cleft types. As efforts to compare outcomes between centres increase, reporting results in the appropriate context, or in a ‘risk-adjusted’ manner, is necessary to provide an accurate reflection of outcomes. Ideally, PRO measurement in CL/P would include scales tailored to outcomes of treatment (e.g., scales asking about the appearance of the lip in the case of lip surgery) that are sensitive to change and also reflect differences in the severity of presentation.

The CLEFT-Q Project

The project to develop a patient-reported outcome measure for patients with CL/P, the CLEFT-Q, was borne out of the knowledge gaps identified above. In planning the study, we knew that we wanted to adhere to several principles. First, we would include patients from low- and middle-income countries in order to develop an instrument that would help tell their stories in a measureable way. Second, we would engage clinician stakeholders during the entire process in order to form a network of collaborators who would be most likely to use the CLEFT-Q once it was completed. Finally, we would follow standard measurement development guidelines in order to achieve the best possible PRO measurement.

Thesis Objectives and Structure

The overall objective of this thesis is to show that through adherence to rigorous methods of development, a PRO measure can provide clinically meaningful outcome evaluation in cleft care. The addition of the patient perspective to outcome evaluation has the potential to transform cleft care; there is potential for advocacy for patients in low- and middle-income countries by providing an accurate reflection of their status, and potential for improved understanding of the patient's status and the impact of treatment for clinicians. The objective is achieved by describing the key phase of developing the conceptual framework for the CLEFT-Q, how the CLEFT-Q was developed, and how the CLEFT-Q can be useful in the evaluation of cleft care. The objectives are reached by addressing the knowledge gaps identified above in three publications that form the substance of this sandwich thesis.

Chapter 2: What Should We Measure?

The first stage of developing the CLEFT-Q included an extensive qualitative study to answer the question: “What concepts related to having a cleft and having it treated are important to patients with CL/P?” Many of the descriptions of quality criteria for PRO measures emphasize the importance of content validity(27, 28, 30), and patient input is the best way to ensure that the measure includes all concepts that are important. The aim of the study was to develop a conceptual framework of what should be measured in the CLEFT-Q. The qualitative methodology employed was Interpretive Description(135, 136).

This study, titled ‘What matters to patients with cleft lip and/or palate: an international qualitative study informing the development of the CLEFT-Q’, was published online in the Cleft Palate-Craniofacial Journal on December 14, 2017.

Chapter 3: How was the CLEFT-Q developed?

As the study continued, it became clear that the science behind PRO measurement was poorly understood in the clinical community. We found ourselves explaining the process on several occasions, including a 90-minute general session panel at the annual meeting of the American Cleft Palate-Craniofacial Association (ACPA) in 2014, an invited 50-minute talk at the ACPA Pre-Conference Symposium on outcomes measurement in 2015, and an invited 30-minute talk at the ACPA Pre-Conference Symposium on team care in 2016. We were asked numerous questions as the study progressed. *Why was the study taking so long? What is Rasch measurement? Why can't we just use an existing PRO measure?* We found that there was an appetite for understanding how PRO measurement worked and why it was important to use scientifically sound measures. As a result, we published a detailed protocol for the development of the CLEFT-Q. This second paper, titled ‘International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q)’, was published in BMJ Open in 2017.

Chapter 4: How do CLEFT-Q outcomes vary by cleft type?

The CLEFT-Q was incorporated into an international standardized set of outcome measures during development(39). Benchmarking and establishing the ability to compare outcomes between centres are becoming more of a focus in health care(114, 119). Incorporating measurements that are able to discriminate between mild and severe presentations of CL/P is important as these efforts move forward. The CLEFT-Q was developed in an international field-test that included 2,434 patients from 30 centres in 12 countries(137). The paper presented in this chapter, which is ready for submission at the end of 2017, analyzes these data in order to show how CLEFT-Q scores vary by cleft type. Visualizing multiple outcomes simultaneously is also an important component of evaluating overall outcomes. We present the data in this study using radar charts as a method to visualize multiple outcomes at once(138). The results of this study should help in future benchmarking efforts by providing an example of how the CLEFT-Q detects differences in outcomes based on cleft type, and also how the outcomes can be visualized in a clinically useful manner.

Chapter 5: Discussion

The closing chapter of the thesis discusses the overall conclusions of the research presented and the potential impact of the findings for future clinical care and research. The results are also considered in the context of the thesis objectives. This last chapter discusses strengths and limitations, ongoing and future work, and knowledge translation strategies.

References

1. Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., Shaw, W. C. Cleft lip and palate. *Lancet* 2009;374:1773-1785.
2. Tanaka, S. A., Mahabir, R. C., Jupiter, D. C., Menezes, J. M. Updating the epidemiology of cleft lip with or without cleft palate. *Plast Reconstr Surg* 2012;129:511e-518e.
3. Tanaka, S. A., Mahabir, R. C., Jupiter, D. C., Menezes, J. M. Updating the epidemiology of isolated cleft palate. *Plast Reconstr Surg* 2013;131:650e-652e.
4. Group, I. W. Prevalence at birth of cleft lip with or without cleft palate: data from the International Perinatal Database of Typical Oral Clefts (IPDTC). *Cleft Palate Craniofac J* 2011;48:66-81.
5. Watkins, S. E., Meyer, R. E., Strauss, R. P., Aylsworth, A. S. Classification, epidemiology, and genetics of orofacial clefts. *Clin Plast Surg* 2014;41:149-163.
6. Stock, N. M., Humphries, K., Pourcain, B. S., et al. Opportunities and Challenges in Establishing a Cohort Study: An Example From Cleft Lip/Palate Research in the United Kingdom. *Cleft Palate Craniofac J* 2016;53:317-325.
7. Aad, G., Abbott, B., Abdallah, J., et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys Rev Lett* 2015;114:191803.
8. Fisher, D. M., Sommerlad, B. C. Cleft lip, cleft palate, and velopharyngeal insufficiency. *Plast Reconstr Surg* 2011;128:342e-360e.

9. Shaw, W. C., Semb, G., Nelson, P., et al. The Eurocleft project 1996-2000: overview. *J Craniomaxillofac Surg* 2001;29:131-140; discussion 141-132.
10. Losee, J. E., Kirschner, R. E. *Comprehensive Cleft Care*, 2nd ed: CRC Press; 2015.
11. Mulliken, J. B. The changing faces of children with cleft lip and palate. *N Engl J Med* 2004;351:745-747.
12. Eiseley, L. The Star Thrower. *The Unexpected Universe*. San Diego, CA: Harcourt, Brace and World; 1969.
13. Alkire, B., Hughes, C. D., Nash, K., Vincent, J. R., Meara, J. G. Potential economic benefit of cleft lip and palate repair in sub-Saharan Africa. *World J Surg* 2011;35:1194-1201.
14. Corlew, D. S. Estimation of impact of surgical disease through economic modeling of cleft lip and palate care. *World J Surg* 2010;34:391-396.
15. Corlew, D. S., Alkire, B. C., Poenaru, D., Meara, J. G., Shrimpe, M. G. Economic valuation of the impact of a large surgical charity using the value of lost welfare approach. *BMJ Glob Health* 2016;1:e000059.
16. Poenaru, D., Lin, D., Corlew, S. Economic Valuation of the Global Burden of Cleft Disease Averted by a Large Cleft Charity. *World J Surg* 2016;40:1053-1059.
17. Grimes, C. E., Henry, J. A., Maraka, J., Mkandawire, N. C., Cotton, M. Cost-effectiveness of surgery in low- and middle-income countries: a systematic review. *World J Surg* 2014;38:252-263.

18. Magee, W. P., Vander Burg, R., Hatcher, K. W. Cleft lip and palate as a cost-effective health care treatment in the developing world. *World J Surg* 2010;34:420-427.
19. Poenaru, D. Getting the job done: analysis of the impact and effectiveness of the SmileTrain program in alleviating the global burden of cleft disease. *World J Surg* 2013;37:1562-1570.
20. Gates, B. Annual Letter 2013: Measuring Progress. Available at: <https://www.gatesfoundation.org/Who-We-Are/Resources-and-Media/Annual-Letters-List/Annual-Letter-2013>. Accessed November 8 2017.
21. Administration, U. S. F. a. D. Clinical Outcome Assessment Qualification Program. Available at: <https://www.fda.gov/drugs/developmentapprovalprocess/drugdevelopmenttoolsqualificationprogram/ucm284077.htm>. Accessed November 3 2017.
22. Valderas, J. M., Alonso, J., Guyatt, G. H. Measuring patient-reported outcomes: moving from clinical trials into clinical practice. *Med J Aust* 2008;189:93-94.
23. Marshall, M., Lockwood, A., Bradley, C., Adams, C., Joy, C., Fenton, M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249-252.
24. Streiner, D. L., Norman, G. R. *Health measurement scales : a practical guide to their development and use*, 4th ed. Oxford ; New York: Oxford University Press; 2008.

25. Administration, U. F. a. D. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. In F. a. D. A. Department of Health and Human Services, Center for Drug Evaluation and Research ed. Rockville, MD, USA, 2009.
26. Aaronson, N., Alonso, J., Burnam, A., et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193-205.
27. Patrick, D. L., Burke, L. B., Gwaltney, C. J., et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument. *Value Health* 2011;14:967-977.
28. Patrick, D. L., Burke, L. B., Gwaltney, C. J., et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. *Value Health* 2011;14:978-988.
29. Reeve, B. B., Wyrwich, K. W., Wu, A. W., et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889-1905.

30. Terwee, C. B., Bot, S. D., de Boer, M. R., et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
31. Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-745.
32. Hartnick, C. J. Validation of a pediatric voice quality-of-life instrument: the pediatric voice outcome survey. *Arch Otolaryngol Head Neck Surg* 2002;128:919-922.
33. Andrich, D. *Rasch models for measurement*. Newbury Park: Sage Publications; 1988.
34. CRONBACH, L. J., MEEHL, P. E. Construct validity in psychological tests. *Psychol Bull* 1955;52:281-302.
35. Hobart, J., Cano, S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13:iii, ix-x, 1-177.
36. Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen; 1960.
37. Lord, F. M., Novick, M. N. *Statistical theories of mental test development*. Reading, MA: Addison-Wesley; 1968.

38. Sitzman, T. J., Allori, A. C., Thorburn, G. Measuring outcomes in cleft lip and palate treatment. *Clin Plast Surg* 2014;41:311-319.
39. Allori, A. C., Kelley, T., Meara, J. G., et al. A Standard Set of Outcome Measures for the Comprehensive Appraisal of Cleft Care. *Cleft Palate Craniofac J* 2017;54:540-554.
40. Shaw, W. C., Asher-McDade, C., Brattström, V., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 1. Principles and study design. *Cleft Palate Craniofac J* 1992;29:393-397.
41. Mølsted, K., Asher-McDade, C., Brattström, V., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 2. Craniofacial form and soft tissue profile. *Cleft Palate Craniofac J* 1992;29:398-404.
42. Mars, M., Asher-McDade, C., Brattström, V., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 3. Dental arch relationships. *Cleft Palate Craniofac J* 1992;29:405-408.
43. Asher-McDade, C., Brattström, V., Dahl, E., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 4. Assessment of nasolabial appearance. *Cleft Palate Craniofac J* 1992;29:409-412.

44. Semb, G., Brattstrom, V., Molsted, K., Prah-Andersen, B., Shaw, W. C.
The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 1: introduction and treatment experience. *Cleft Palate Craniofac J* 2005;42:64-68.
45. Brattstrom, V., Molsted, K., Prah-Andersen, B., Semb, G., Shaw, W. C.
The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 2: craniofacial form and nasolabial appearance. *Cleft Palate Craniofac J* 2005;42:69-77.
46. Molsted, K., Brattstrom, V., Prah-Andersen, B., Shaw, W. C., Semb, G.
The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 3: dental arch relationships. *Cleft Palate Craniofac J* 2005;42:78-82.
47. Semb, G., Brattstrom, V., Molsted, K., et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 4: relationship among treatment outcome, patient/parent satisfaction, and the burden of care. *Cleft Palate Craniofac J* 2005;42:83-92.
48. Shaw, W. C., Brattstrom, V., Molsted, K., Prah-Andersen, B., Roberts, C. T., Semb, G. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 5: discussion and conclusions. *Cleft Palate Craniofac J* 2005;42:93-98.

49. Sandy, J. R., Williams, A. C., Bearn, D., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 1: background and methodology. *Cleft Palate Craniofac J* 2001;38:20-23.
50. Williams, A. C., Bearn, D., Mildinhall, S., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 2: dentofacial outcomes and patient satisfaction. *Cleft Palate Craniofac J* 2001;38:24-29.
51. Sell, D., Grunwell, P., Mildinhall, S., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 3: speech outcomes. *Cleft Palate Craniofac J* 2001;38:30-37.
52. Bearn, D., Mildinhall, S., Murphy, T., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 4: outcome comparisons, training, and conclusions. *Cleft Palate Craniofac J* 2001;38:38-43.
53. Long, R. E., Jr., Hathaway, R., Daskalogiannakis, J., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J* 2011;48:239-243.
54. Hathaway, R., Daskalogiannakis, J., Mercado, A., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 2. Dental arch relationships. *Cleft Palate Craniofac J* 2011;48:244-251.

55. Daskalogiannakis, J., Mercado, A., Russell, K., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 3. Analysis of craniofacial form. *Cleft Palate Craniofac J* 2011;48:252-258.
56. Mercado, A., Russell, K., Hathaway, R., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 4. Nasolabial aesthetics. *Cleft Palate Craniofac J* 2011;48:259-264.
57. Russell, K., Long, R. E., Jr., Hathaway, R., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 5. General discussion and conclusions. *Cleft Palate Craniofac J* 2011;48:265-270.
58. Persson, M., Sandy, J. R., Waylen, A., et al. A cross-sectional survey of 5-year-old children with non-syndromic unilateral cleft lip and palate: the Cleft Care UK study. Part 1: background and methodology. *Orthod Craniofac Res* 2015;18 Suppl 2:1-13.
59. Al-Ghatam, R., Jones, T. E., Ireland, A. J., et al. Structural outcomes in the Cleft Care UK study. Part 2: dento-facial outcomes. *Orthod Craniofac Res* 2015;18 Suppl 2:14-24.
60. Smallridge, J., Hall, A. J., Chorbachi, R., et al. Functional outcomes in the Cleft Care UK study--Part 3: oral health and audiology. *Orthod Craniofac Res* 2015;18 Suppl 2:25-35.

61. Sell, D., Mildinhall, S., Albery, L., Wills, A. K., Sandy, J. R., Ness, A. R. The Cleft Care UK study. Part 4: perceptual speech outcomes. *Orthod Craniofac Res* 2015;18 Suppl 2:36-46.
62. Waylen, A., Ness, A. R., Wills, A. K., Persson, M., Rumsey, N., Sandy, J. R. Cleft Care UK study. Part 5: child psychosocial outcomes and satisfaction with cleft services. *Orthod Craniofac Res* 2015;18 Suppl 2:47-55.
63. Semb, G., Enemark, H., Friede, H., et al. A Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 1. Planning and management. *J Plast Surg Hand Surg* 2017;51:2-13.
64. Rautio, J., Andersen, M., Bolund, S., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 2. Surgical results. *J Plast Surg Hand Surg* 2017;51:14-20.
65. Bannister, P., Lindberg, N., Jeppesen, K., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 3. Descriptive study of postoperative nursing care following first stage cleft closure. *J Plast Surg Hand Surg* 2017;51:21-26.
66. Lohmander, A., Persson, C., Willadsen, E., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 4. Speech outcomes in 5-year-olds - velopharyngeal competency and hypernasality. *J Plast Surg Hand Surg* 2017;51:27-37.
67. Willadsen, E., Lohmander, A., Persson, C., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 5. Speech outcomes in

5-year-olds - consonant proficiency and errors. *J Plast Surg Hand Surg* 2017;51:38-51.

68. Heliövaara, A., Küsseler, A., Skaare, P., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 6. Dental arch relationships in 5 year-olds. *J Plast Surg Hand Surg* 2017;51:52-57.

69. Karsten, A., Marcusson, A., Hurmerinta, K., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 7. Occlusion in 5 year-olds according to the Huddart and Bodenham index. *J Plast Surg Hand Surg* 2017;51:58-63.

70. Mølsted, K., Humerinta, K., Küsseler, A., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 8. Assessing naso-labial appearance in 5-year-olds - a preliminary study. *J Plast Surg Hand Surg* 2017;51:64-72.

71. Feragen, K. B., Rumsey, N., Heliövaara, A., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and Palate: 9. Parental report of social and emotional experiences related to their 5-year-old child's cleft diagnosis. *J Plast Surg Hand Surg* 2017;51:73-80.

72. Feragen, K. B., Semb, G., Heliövaara, A., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 10. Parental perceptions of appearance and treatment outcomes in their 5-year-old child. *J Plast Surg Hand Surg* 2017;51:81-87.

73. Shaw, W., Semb, G. The Scandleft randomised trials of primary surgery for unilateral cleft lip and palate: 11. What next? *J Plast Surg Hand Surg* 2017;51:88-93.
74. Lohmander, A., Willadsen, E., Persson, C., Henningsson, G., Bowden, M., Hutter, B. Methodology for speech assessment in the Scandleft project--an international randomized clinical trial on palatal surgery: experiences from a pilot study. *Cleft Palate Craniofac J* 2009;46:347-362.
75. Stock, N. M., Feragen, K. B. Psychological adjustment to cleft lip and/or palate: A narrative review of the literature. *Psychol Health* 2016;31:777-813.
76. Hunt, O., Burden, D., Hepper, P., Johnston, C. The psychosocial effects of cleft lip and palate: a systematic review. *Eur J Orthod* 2005;27:274-285.
77. Sharma, V. P., Bella, H., Cadier, M. M., Pigott, R. W., Goodacre, T. E., Richard, B. M. Outcomes in facial aesthetics in cleft lip and palate surgery: A systematic review. *J Plast Reconstr Aesthet Surg* 2012;65:1233-1245.
78. Mani, M. R., Semb, G., Andlin-Sobocki, A. Nasolabial appearance in adults with repaired unilateral cleft lip and palate: Relation between professional and lay rating and patients' satisfaction. *J Plast Surg Hand Surg* 2010;44:191-198.
79. Tyl, J., Dytrych, Z., Helclová, H., Schüller, V., Matějček, Z., Beránková, A. [Psychological and social stress in children with cleft lip and palate]. *Cesk Pediatr* 1990;45:532-536.

80. Marcusson, A., Paulin, G., Ostrup, L. Facial appearance in adults who had cleft lip and palate treated in childhood. *Scand J Plast Reconstr Surg Hand Surg* 2002;36:16-23.
81. Hunt, O., Burden, D., Hepper, P., Stevenson, M., Johnston, C. Self-reports of psychosocial functioning among children and young adults with cleft lip and palate. *Cleft Palate Craniofac J* 2006;43:598-605.
82. Sinko, K., Jagsch, R., Prechtel, V., Watzinger, F., Hollmann, K., Baumann, A. Evaluation of esthetic, functional, and quality-of-life outcome in adult cleft lip and palate patients. *Cleft Palate Craniofac J* 2005;42:355-361.
83. Hunt, O., Burden, D., Hepper, P., Stevenson, M., Johnston, C. Parent reports of the psychosocial functioning of children with cleft lip and/or palate. *Cleft Palate Craniofac J* 2007;44:304-311.
84. Oosterkamp, B. C., Dijkstra, P. U., Rimmelink, H. J., et al. Satisfaction with treatment outcome in bilateral cleft lip and palate patients. *Int J Oral Maxillofac Surg* 2007;36:890-895.
85. Chuo, C. B., Searle, Y., Jeremy, A., Richard, B. M., Sharp, I., Slator, R. The continuing multidisciplinary needs of adult patients with cleft lip and/or palate. *Cleft Palate Craniofac J* 2008;45:633-638.
86. Berger, Z. E., Dalton, L. J. Coping with a cleft: psychosocial adjustment of adolescents with a cleft lip and palate and their parents. *Cleft Palate Craniofac J* 2009;46:435-443.

87. Meyer-Marcotty, P., Stellzig-Eisenhauer, A. Dentofacial self-perception and social perception of adults with unilateral cleft lip and palate. *J Orofac Orthop* 2009;70:224-236.
88. Van Lierde, K. M., Dhaeseleer, E., Luyten, A., Van De Woestijne, K., Vermeersch, H., Roche, N. Parent and child ratings of satisfaction with speech and facial appearance in Flemish pre-pubescent boys and girls with unilateral cleft lip and palate. *Int J Oral Maxillofac Surg* 2012;41:192-199.
89. Nkenke, E., Stelzle, F., Vairaktaris, E., Knipfer, C. Do cleft lip and palate patients opt for secondary corrective surgery of upper lip and nose, frequently? *Head Face Med* 2013;9:38.
90. Byrne, M., Chan, J. C., O'Broin, E. Perceptions and satisfaction of aesthetic outcome following secondary cleft rhinoplasty: evaluation by patients versus health professionals. *J Craniomaxillofac Surg* 2014;42:1062-1070.
91. Slifer, K. J., Amari, A., Diver, T., et al. Social interaction patterns of children and adolescents with and without oral clefts during a videotaped analogue social encounter. *Cleft Palate Craniofac J* 2004;41:175-184.
92. Versnel, S. L., Duivendoorn, H. J., Passchier, J., Mathijssen, I. M. Satisfaction with facial appearance and its determinants in adults with severe congenital facial disfigurement: a case-referent study. *J Plast Reconstr Aesthet Surg* 2010;63:1642-1649.

93. Wehby, G. L., Tyler, M. C., Lindgren, S., Romitti, P., Robbins, J., Damiano, P. Oral clefts and behavioral health of young children. *Oral Dis* 2012;18:74-84.
94. Richman, L. C. Behavior and achievement of cleft palate children. *Cleft Palate J* 1976;13:4-10.
95. Kapp, K. Self concept of the cleft lip and or palate child. *Cleft Palate J* 1979;16:171-176.
96. Thomas, P. T., Turner, S. R., Rumsey, N., Dowell, T., Sandy, J. R. Satisfaction with facial appearance among subjects affected by a cleft. *Cleft Palate Craniofac J* 1997;34:226-231.
97. Marcusson, A., Akerlind, I., Paulin, G. Quality of life in adults with repaired complete cleft lip and palate. *Cleft Palate Craniofac J* 2001;38:379-385.
98. Gussy, M., Kilpatrick, N. The self-concept of adolescents with cleft lip and palate: a pilot study using a multidimensional/hierarchical measurement instrument. *Int J Paediatr Dent* 2006;16:335-341.
99. Clifford, E., Crocker, E. C., Pope, B. A. Psychological findings in the adulthood of 98 cleft lip-palate children. *Plast Reconstr Surg* 1972;50:234-237.
100. Bjornsson, A., Agustsdottir, S. A psychosocial study of Icelandic individuals with cleft lip or cleft lip and palate. *Cleft Palate J* 1987;24:152-157.
101. Slifer, K. J., Diver, T., Amari, A., et al. Assessment of facial emotion encoding and decoding skills in children with and without oral clefts. *J Craniomaxillofac Surg* 2003;31:304-315.

102. Tyler, M. C., Wehby, G. L., Robbins, J. M., Damiano, P. C. Separation anxiety in children ages 4 through 9 with oral clefts. *Cleft Palate Craniofac J* 2013;50:520-527.
103. Lorot-Marchand, A., Guerreschi, P., Pellerin, P., et al. Frequency and socio-psychological impact of taunting in school-age patients with cleft lip-palate surgical repair. *Int J Pediatr Otorhinolaryngol* 2015;79:1041-1048.
104. Damiano, P. C., Tyler, M. C., Romitti, P. A., et al. Type of oral cleft and mothers' perceptions of care, health status, and outcomes for preadolescent children. *Cleft Palate Craniofac J* 2006;43:715-721.
105. Landsberger, P., Proff, P., Dietze, S., et al. Evaluation of patient satisfaction after therapy of unilateral clefts of lip, alveolus and palate. *J Craniomaxillofac Surg* 2006;34 Suppl 2:31-33.
106. Mani, M., Reiser, E., Andlin-Sobocki, A., Skoog, V., Holmström, M. Factors related to quality of life and satisfaction with nasal appearance in patients treated for unilateral cleft lip and palate. *Cleft Palate Craniofac J* 2013;50:432-439.
107. Feragen, K. B., Stock, N. M., Kvaalem, I. L. Risk and Protective Factors at Age 16: Psychological Adjustment in Children With a Cleft Lip and/or Palate. *Cleft Palate Craniofac J* 2015;52:555-573.
108. Pitak-Arnop, P., Hemprich, A., Dhanuthai, K., Yildirim, V., Pausch, N. C. Panel and patient perceptions of nasal aesthetics after secondary cleft

rhinoplasty with versus without columellar grafting. *J Craniomaxillofac Surg* 2011;39:319-325.

109. Gkantidis, N., Papamanou, D. A., Karamolegkou, M., Dorotheou, D. Esthetic, Functional, and Everyday Life Assessment of Individuals with Cleft Lip and/or Palate. *Biomed Res Int* 2015;2015:510395.

110. Klassen, A. F., Stotland, M. A., Skarsgard, E. D., Pusic, A. L. Clinical research in pediatric plastic surgery and systematic review of quality-of-life questionnaires. *Clin Plast Surg* 2008;35:251-267.

111. Tsihklaki, A., O'Brien, K., Johal, A., Fleming, P. S. A scoping review of outcomes related to orthodontic treatment measured in cleft lip and palate. *Orthod Craniofac Res* 2017;20:55-64.

112. Eckstein, D. A., Wu, R. L., Akinbiyi, T., Silver, L., Taub, P. J. Measuring quality of life in cleft lip and palate patients: currently available patient-reported outcomes measures. *Plast Reconstr Surg* 2011;128:518e-526e.

113. Black, N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.

114. Porter, M. E. What is value in health care? *N Engl J Med* 2010;363:2477-2481.

115. Cohen, J. B., Myckatyn, T. M., Brandt, K. The Importance of Patient Satisfaction: A Blessing, a Curse, or Simply Irrelevant? *Plast Reconstr Surg* 2017;139:257-261.

116. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;4:79.
117. OECD. Recommendations to OECD Ministers of Health from the high level reflection group on the future of health statistics: Strengthening the international comparison of health system performance through patient-reported indicators. Paris: Organisation for Economic Co-operation and Development, 2017.
118. Valderas, J. M., Kotzeva, A., Espallargues, M., et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res* 2008;17:179-193.
119. Porter, M. E., Larsson, S., Lee, T. H. Standardizing Patient Outcomes Measurement. *N Engl J Med* 2016;374:504-506.
120. Klassen, A. F., Tsangaris, E., Forrest, C. R., et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. *J Plast Reconstr Aesthet Surg* 2012;65:547-557.
121. Mani, M., Carlsson, M., Marcusson, A. Quality of life varies with gender and age among adults treated for unilateral cleft lip and palate. *Cleft Palate Craniofac J* 2010;47:491-498.
122. Sagheri, D., Ravens-Sieberer, U., Braumann, B., von Mackensen, S. An Evaluation of Health-Related Quality of Life (HRQoL) in a group of 4-7 year-old children with cleft lip and palate. *J Orofac Orthop* 2009;70:274-284.

123. Aravena, P. C., Gonzalez, T., Oyarzún, T., Coronado, C. Oral Health-Related Quality of Life in Children in Chile Treated for Cleft Lip and Palate: A Case-Control Approach. *Cleft Palate Craniofac J* 2017;54:e15-e20.
124. Wirls, C. J., Plotkin, R. R. A comparison of children with cleft palate and their siblings on projective test personality factors. *Cleft Palate J* 1971;8:399-408.
125. Richman, L. C. Self-reported social, speech, and facial concerns and personality adjustment of adolescents with cleft lip and palate. *Cleft Palate J* 1983;20:108-112.
126. Heller, A., Rafman, S., Zvagulis, I., Pless, I. B. Birth defects and psychosocial adjustment. *Am J Dis Child* 1985;139:257-263.
127. Geier, R., Wittstock, C. [Psychosocial studies of surgically treated adults with unilateral total lip-maxilla-palate clefts]. *Psychiatr Neurol Med Psychol (Leipz)* 1986;38:656-661.
128. Bressmann, T., Sader, R., Ravens-Sieberer, U., Zeilhofer, H. F., Horch, H. H. [Quality of life research in patients with cleft lip and palate: preliminary results]. *Mund Kiefer Gesichtschir* 1999;3:134-139.
129. Kramer, F. J., Gruber, R., Fialka, F., Sinikovic, B., Schliephake, H. Quality of life and family functioning in children with nonsyndromic orofacial clefts at preschool ages. *J Craniofac Surg* 2008;19:580-587.
130. Augsornwan, D., Namedang, S., Pongpagatip, S., Surakunprapha, P. Quality of life in patients with cleft lip and palate after operation. *J Med Assoc Thai* 2011;94 Suppl 6:S124-128.

131. Smith, C. B., Walker, K., Badawi, N., Waters, K. A., MacLean, J. E. Impact of sleep and breathing in infancy on outcomes at three years of age for children with cleft lip and/or palate. *Sleep* 2014;37:919-925.
132. Ramstad, T., Ottem, E., Shaw, W. C. Psychosocial adjustment in Norwegian adults who had undergone standardised treatment of complete cleft lip and palate. II. Self-reported problems and concerns with appearance. *Scand J Plast Reconstr Surg Hand Surg* 1995;29:329-336.
133. Broder, H. L., Wilson-Genderson, M., Sischo, L. Examination of a theoretical model for oral health-related quality of life among youths with cleft. *Am J Public Health* 2014;104:865-871.
134. Broder, H. L., Wilson-Genderson, M., Sischo, L. Oral health-related quality of life in youth receiving cleft-related surgery: self-report and proxy ratings. *Qual Life Res* 2017;26:859-867.
135. Thorne, S., Kirkham, S. R., MacDonald-Emes, J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. *Res Nurs Health* 1997;20:169-177.
136. Thorne, S. E. *Interpretive description*. Walnut Creek, CA: Left Coast Press; 2008.
137. Klassen, A., Wong Riff, K., Longmire, N., et al. Psychometric findings and normative values for the CLEFT-Q based on 2,434 children and young adult patients with cleft lip and/or palate from 12 countries. *Canadian Medical Association Journal* 2017;in press.

138. Kaplan, R. S., Blackstone, R. P., Haas, D. A., Thaker, N. G. Measuring and communicating health care value with charts. *Harv Bus Rev*, 2015.

CHAPTER TWO

Preface

Chapter Two consists of the qualitative study performed to identify the concepts that patients with CL/P felt were important. The results of this study guided the development of the CLEFT-Q scales.

The study has been published in the Cleft Palate-Craniofacial Journal. Permission to reprint the article as part of this thesis was granted by Sage Publications (December 23, 2017) using the following citation:

Wong Riff KWY, Tsangaris E, Goodacre TEE, Forrest CR, Lawson J, Pusic AL, Klassen AF. What matters to patients with cleft lip and/or palate: an international qualitative study informing the development of the CLEFT-Q. December 14, 2017, DOI:10.1177/1055665617732854

My contribution to this study included study concept and design together with Anne Klassen, data acquisition in Canada and the USA, coordination of data acquisition in India and Kenya, analysis and interpretation of data together with Anne Klassen, and writing the manuscript.

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

**What Matters to Patients with Cleft Lip and/or Palate: An International
Qualitative Study Informing the Development of the CLEFT-Q**

Karen W.Y. Wong Riff, MD MSc FRCSC

Assistant Professor, Division of Plastic and Reconstructive Surgery, Hospital for
Sick Children, University of Toronto, Toronto, Canada

Elena Tsangaris, PhD

Department of Clinical Epidemiology and Biostatistics, McMaster University,
Hamilton, Canada

Timothy E. E. Goodacre, MBBS BSc FRCS

Consultant Plastic Surgeon, Spires Cleft Center, Oxford Radcliffe Children's
Hospital, Oxford, UK

Christopher R. Forrest, MD MSc FRCSC

Professor, Division of Plastic and Reconstructive Surgery, Hospital for Sick
Children, University of Toronto, Toronto, Canada

Jessica Lawson, MSc

Ph.D. Candidate, Ferkauf Graduate School of Psychology, Yeshiva University,
Bronx, USA

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

Andrea L. Pusic, MD MPH FACS

Associate Professor, Plastic and Reconstructive Surgery, Memorial Sloan-Kettering Cancer Center, New York, USA

Anne F. Klassen, DPhil BA

Associate Professor, Department of Pediatrics, McMaster University, Hamilton, Canada

Oral Presentations

Early findings of this study have been presented at the American Association of Plastic Surgeons Annual Meeting, 2012 and the American Cleft Palate-Craniofacial Association Meeting, 2012.

Study Support

The study was supported with the following grants:

Physicians' Services Incorporated Foundation Resident Research Grant

Canadian Society of Plastic Surgeons Educational Foundation Outcomes Research Grant

Acknowledgements

We would like to acknowledge our partners in cleft care centers for participating in this study, including Dr. Arun Gosain from Rainbow Babies and Children's

Hospital, Dr. Tristan De Chalain and Operation Restore Hope NZ, Drs. Alex Campbell, Hiteswar Sarma, and Gaurav Deshpande from the Guwahati Comprehensive Cleft Care Centre, Operation Smile India, Dr. Joseph Theuri and CURE International. We would also like to acknowledge Beryl Tan and Ananya Banerjee who performed qualitative interviews and participated in coding data.

Abstract

Objective: The goal of treatment for individuals with cleft lip and/or palate (CL/P) is to improve physical, psychological, and social health. Outcomes of treatment are rarely measured from the patient's perspective. The aim of the study was to develop a conceptual framework for a patient-reported outcome (PRO) instrument for individuals with clefts (CLEFT-Q) by developing an in-depth understanding of issues that individuals consider to be important.

Design: The qualitative methodology of *interpretive description* was used.

Setting, Participants and Intervention: We performed 136 individual in-depth interviews with participants with clefts of any age, presenting for cleft care, across six countries. Parents were involved if the child was more comfortable. Interviews were audio-recorded, transcribed verbatim, and coded using constant comparison. The data were used to develop a refined conceptual framework.

Results: Participants described concepts of interest in three top-level domains, each of which included sub-domains: appearance (face, nose, nostrils, teeth, lips, jaw, cleft lip scar), health-related quality of life (psychological, social, school, speech-related distress), and facial function (speech, eating/drinking). Participants were able to describe changes over time with regards to the three domains.

Conclusions: A conceptual framework of concepts of interest to individuals with CL/P formed the basis of the scales in the CLEFT-Q. Each sub-domain represents an independently functioning scale. Understanding what matters to patients is essential in guiding PRO measurement.

Key Words: cleft lip and/or palate, interpretive description, qualitative, patient-reported outcomes

Introduction

Cleft lip and/or palate (CL/P) is prevalent in children worldwide and affects appearance, dentition, hearing, speech, and psychosocial functioning. The goal of treatment of CL/P is to improve the patient's physical, psychological, and social health. Measuring these outcomes requires the inclusion of the patient perspective, but a comprehensive, condition-specific patient-reported outcome (PRO) instrument including all aspects of cleft care for patients with CL/P has not been available to date (Eckstein et al., 2011).

PROs are 'reports that come directly from patients about how they function or feel in relation to a health condition and its therapy, without interpretation by a physician or anyone else' (Valderas et al., 2008; p. 93). Measuring the patient perspective in a scientifically sound, clinically meaningful manner is crucial as health care increasingly focuses on value for the patient as the most important outcome (Porter, 2010). In order to develop a PRO instrument for this purpose, an in-depth understanding of what the patients feel to be important in their treatment outcomes is required. Thorough qualitative input into a PRO instrument is essential to establishing content validity (Patrick et al., 2011a, 2011b).

Previous literature has identified the need to include self-perception in the evaluation of treatment outcomes (Alansari et al., 2014). Stock et al. (2016) have reported that background, external, and internal factors contribute to psychological adjustment in adults with CL/P. Individuals with CL/P have been

identified as having behavioral and emotional difficulties (Endriga and Kapp-Simon, 1999; Stock et al., 2016) as well as conditions resulting from facial differences such as depression and anxiety (Thompson and Kent, 2001). Satisfaction with overall appearance has been correlated with psychosocial functioning (Thomas et al., 1997; Stock et al., 2016). Feragen and Stock (2016) reported an association between teasing and satisfaction with appearance as well as depressive symptoms in adolescent females. Several reviews have highlighted methodological concerns with the existing literature, particularly in the lack of uniformity or consistency (Hunt et al., 2005; Stock et al., 2016; Thompson and Kent, 2001; Turner et al., 1997). Although there have been numerous quantitative studies using various instruments to evaluate patients with CL/P (Stock et al., 2016), a systematic review of qualitative studies describing the experience of children and young patients found two studies focused on patients of this age (Sharif et al., 2013). Hall et al. (2013) found that children with CL/P develop gradual awareness of the cleft and the treatment pathway, and that the frequency and complexity of treatment are important in patients' accounts of having CL/P. Cleft-related surgery has been shown to improve oral health-related quality of life in patients with CL/P (Broder et al., 2017). Children with lower health-related quality of life were also found to seek revision surgery more frequently (Ranganathan et al., 2016). With the variety of treatment options available to a patient with CL/P, it is important to be able to evaluate and understand the patient perspective before and after interventions. While the studies listed above have

identified associations between treatment and health-related quality of life, using a specific, discriminative instrument to assess the different impacts treatments have on patients would be beneficial.

This study represents a key qualitative component of a multiphase mixed methods study to develop a PRO measure for patients with CL/P, the CLEFT-Q. Figure 1 depicts the flow of the different phases of the study. The protocol for the overall study has been published elsewhere (Wong Riff et al., 2017). The CLEFT-Q is being designed to be a clinically meaningful, discriminative instrument to assess outcomes of treatment at the individual level. The methodology follows the gold standard for developing PRO measures set forth by the Scientific Advisory Committee of the Medical Outcomes Trust (Aaronson et al., 2002), the USA Food and Drug Administration (2009), and the International Society for Pharmacoeconomics and Outcomes Research (Patrick et al., 2011a; Patrick, 2011b).

The aim of the current study was to use qualitative methodology to identify concepts that are important to patients with CL/P from their own perspective in order to build a corresponding comprehensive framework to build the CLEFT-Q scales.

Methods

The study followed the qualitative methodology of Interpretive Description, which presumes that theoretical knowledge, clinical knowledge, and a scientific basis inform the study (Thorne, 2008; Thorne et al., 1997). Qualitative

studies are underpinned by philosophical assumptions. In this study, a pragmatic approach was taken, meaning that while there may be academic explanations of a certain concept, the way in which the concept is understood by the individual is of greatest importance (Welford et al., 2011). A preliminary conceptual framework of what concepts had been measured in individuals with CL/P in the past, developed from a systematic review of the literature (Klassen et al., 2012), was used to guide the study (Figure 2).

Institutional review board approval was obtained for all participating sites in the study, and written informed assent and/or consent was obtained from participants.

Sampling

Participants in six different countries (Canada, Kenya, India, Philippines, England, and USA) were eligible for inclusion in the study if they had a diagnosis of CL/P. Centers were contacted for participation in the study by the primary authors, and those that showed interest were included. In the high-income countries (Canada, England, USA), participants aged 6 to 22 years were included. Participants were recruited from pediatric cleft care centers. In the low- and middle-income countries (Kenya, India, and the Philippines), individuals with CL/P of any age were included if they were presenting for clinical care, resulting in participants aged from birth to 70 years. Parents of children with CL/P in low- and middle-income countries were invited to participate if the children were more comfortable; this variation was due to our desire to gather as much information as

possible as well as cultural differences in how research interviews were received at these sites. We included parents of children under the age of 6 in low- and middle-income countries and used their perceptions in the analysis in order to maximize data collection at these sites. We excluded participants in high-income countries that did not speak English or participants in any country that were unable to undergo an individual interview due to a cognitive delay. Participants were purposively sampled to include a maximum variety of age, gender, and cleft type. Sampling and recruitment continued until the point of saturation, or when no further new concepts were elicited from subsequent interviews (Sandelowski, 2008). In Canada, England, and USA, participants were recruited through posters in clinics and contacted by telephone. In Kenya, India, and the Philippines, a researcher recruited participants face-to-face in the clinical setting.

Data Collection

The interviewer obtained written assent and/or consent as appropriate and then carried out an audio-recorded semi-structured individual interview. An interview guide based on the preliminary conceptual framework directed the interview (Table 1). In order to obtain the participants' own perspectives, no clinician input was sought to create this interview guide and only the previous concepts evaluated in the literature were included. As interviews progressed and participants discussed additional concepts, the interview guide was adapted in order to include any concepts requiring further probing, following qualitative methodology. Age, gender, and cleft type were recorded. Six trained qualitative

interviewers (1-2 per site) performed the interviews in English whenever possible. The same team member trained all interviewers for consistency using previous interview transcripts and practice interviews. Participants were interviewed alone in Canada, England, and the USA. In Kenya, India, and the Philippines, interviews were carried out with a translator when necessary, and with parents present if the participants preferred.

Data Analysis

All interviews were transcribed verbatim, and a bilingual native speaker (English and target language) transcribed and translated the interviews that were carried out in a different language in order to confirm the translation done in person. Data collection and analysis occurred concurrently and iteratively in order to build on the knowledge gained from each interview, according to the qualitative methodology. The transcripts were analyzed, meaning that the content of each line was categorized, using line-by-line coding in NVivo (QSR International Pty Ltd., 2012), allowing for the identification of recurrent themes and patterns. All codes were then categorized into conceptual top-level domains and sub-domains using constant comparison (Pope et al., 2000). These top-level domains and sub-domains were then used to refine and expand the preliminary conceptual framework.

Reflexivity

Coding was performed by one team member (quality of life researcher or quality of life researcher/plastic surgeon) and confirmed by a second team member (quality of life researcher or quality of life researcher/plastic surgeon). A third team member (quality of life researcher) resolved any conflicts. Member-checking, where concepts that are identified are confirmed in the target population, was performed on an iterative basis by adapting the interview guide to include themes that arose during the interviews in subsequent interviews (Cohen and Crabtree, 2008). Peer debriefing, where the concepts were discussed amongst the research team, was also carried out on a regular basis to verify the findings (Cohen and Crabtree, 2008).

Results

We carried out a total of 136 interviews with 138 participants across the six countries (two sets of twins were interviewed together). Interview locations and demographic characteristics of the participants are shown in Table 2.

Three main domains emerged from the interviews and form the basis of the refined conceptual framework, with sub-domains within each of these top-level domains (Figure 2). Examples of how the data were categorized are shown in Table 3. This process was repeated with all data from the interview transcripts in order to develop the final conceptual framework. Participants also discussed the

role of the clinical team and issues related to process of care, which will be reported separately.

Importantly, the words with which participants described concepts and themes were noted. The CLEFT-Q scales were developed using words from the transcripts as much as possible in order to maximize understanding and familiarity with the concepts of interest.

An overarching pattern was that individuals became more aware of their facial and/or speech differences in early adolescence. At this age, teasing and bullying were referred to in the present, as opposed to older individuals who expressed experiencing teasing and bullying in the past. Participants also expressed satisfaction with changes with treatments over time.

While the study was not designed to identify differences between groups of participants, we observed overall patterns in the qualitative data between groups. Participants in high-income countries tended to discuss appearance more than facial function, whereas participants in low- and middle-income countries had more concerns about facial function than about appearance. Many participants from low- and middle-income countries received treatment late, often beyond school age, and consequently experienced the stigmatization of having a cleft for a longer time. As would be expected developmentally, younger patients were able to discuss their appearance, but in less detail than older patients, and younger patients also did not express as much detail with regards to the psychological, social, and school sub-domains. Quantifying this difference is not appropriate

based on qualitative data, so these differences will be assessed in subsequent stages of the study through cognitive debriefing interviews (Wong Riff et al., 2017).

Appearance

This top-level domain relates to the individuals' appraisals of their own appearance. The sub-domains within this top-level domain were categorized into overall facial appearance and then specific anatomic areas. Participants discussed appearance more than any other concept; in particular, the nose was the most frequently mentioned. Importantly, participants did not use clinical terminology to describe their appearance or the changes sought with treatment. General concepts about appearance included feeling different:

“Since I was a kid, ‘What is that on your face? What is that?’ Like, it gets – it’s just ingrained into you that you look different.” (23 year old female with CLP, Canada)

Alongside the concept of feeling different, looking ‘normal’ was frequently a goal of treatment. Participants who were uncomfortable with their appearance did not like having photos taken:

“I only take photos when necessary... Social gatherings I don’t take photos, or the other places I don’t take photos.” (29 year old female with CL, Kenya)

Participants often noted the appearance of their smile or laugh. The smile represented a manifestation of participants' perceptions of their lips, teeth, jaws, and nose. For example:

“...I just started smiling in photos...in the photos you tend to like refrain from smiling, and as I got older, I just like... as my teeth got very, very perfect, and... I saw that in photos and everything... I just started smiling... I'm more open to laughing with my teeth.” (13 year old male with CLP, Canada)

Participants also discussed their profile:

“I wasn't really too happy with my profile 'cause it just sort of went in to about the forehead and the chin, it all went in behind... I remember whenever I used to go, uh, to the shops and you go into the changing rooms, and they always had the mirrors. They're angled so you can see it from the side and I hated that, to be honest.” (21 year old male with CLP, England)

Participants were able to describe specific findings on anatomic areas of the face. For example, the lip was described using general positive comments such as looking 'good', 'pretty', 'nice', and 'normal', and neutral or negative comments such as 'different', 'weird', or 'ugly'. Participants did not use formal clinical terminology such as 'Cupid's bow'. Instead, they used terms such as the 'points' of the lip, or described the lip as looking 'flat' or 'straight'. The thickness of the lip was described as 'puffy' or 'full'. Participants described their cleft lip

scars using simple terms such as ‘faint’, ‘fading’, ‘bumpy’, or ‘wide’. The nose was described using a large variety of terms. Symmetry, the most commonly mentioned concept, was expressed as ‘uneven’, ‘crooked’, size differences between sides, ‘drooping’ on one side, ‘tilted’, or ‘off-centered’.

With respect to the jaws, a hypoplastic maxilla was described as the top jaw being ‘so far back’ or the lower jaw ‘sticking out’. The relationship between the top and bottom jaws was described as an ‘underbite’, ‘not aligned’, or ‘not symmetrical’.

Finally, participants described wanting their teeth to be ‘straight’, ‘aligned’, or ‘even’, as opposed to being ‘wonky’, ‘mangled’, or ‘all over the place’. Smiling and laughing while showing teeth was a common indicator of how satisfied the participants were with their teeth.

The psychosocial impact of appearance was a common theme:

“...I remember specifically I was in a movie theatre in a not very nice area with my friends because it’s the only time the movie was showing. And there’s girls behind me who were obviously either just trying to be mean, or I don’t know, but I don’t even know them and they were saying stuff about my face and I didn’t turn around. I didn’t say anything. I acted like it didn’t bother me, but of course when I got home it made me really angry.” (12 year old female with CLP, USA)

While health-related quality of life concepts were often related to appearance, the frequency and role of appearance was a distinct top-level concept for participants.

Health-Related Quality of Life

Within the top-level domain of health-related quality of life, the sub-domains of psychological, social, school, and speech-related distress were identified.

Psychological

Participants expressed both positive and negative concepts related to their psychological function. Confidence improved with positive treatment outcomes. Participants felt anger, sadness, fear, hurt, worry, and embarrassment as a result of the cleft.

“I feel bad that they are good looking, that they are all good and I’m like this.” (18 year old male with CL, India)

Self-consciousness, impact on self-esteem, and feeling shy were frequently mentioned. Participants used concealing or hiding as a method of coping:

“Sometimes, he’s aware of [meeting new people] and hides his mouth with his hands.” (9 year old male with CL, India)

Participants also described being different or not feeling ‘normal’:

“I feel like I am left out, I feel like I am not loud, I feel that I am the only one having the problem because I am not like other people.” (9 year old female with CP, Kenya)

“I know I would have been more happy being born normal.” (22 year old male with CLP, England)

Participants also described benefit-finding and other positive psychological outcomes of having a cleft.

“...as I’ve grown up, I’ve realized that if I didn’t, if I wasn’t born with it, I would not be who I am right now. I wouldn’t have learned to get such a thick skin... I don’t need other people to help me or take care of me, whereas I think if I had been born different, and I didn’t have to go through everything I did, I probably would be able a bit more dependent. I probably, I wouldn’t be as accepting as I am now of other people. Because I know the people who were born, I hate to say it, but normally, they are not always as accepting as I am. I am extremely accepting of other people. I do judge people by their looks sometimes, but then I remember how I felt. I stick up for people who are being teased.” (20 year old female with CLP, Canada)

Social and School

The social settings that the participants discussed included general social settings, family, school, dating, and social media. In the general setting, participants expressed difficulty with meeting new people, speaking in public, and feeling like they fit in with others. Being asked about their face or their speech was a common source of frustration and embarrassment. Family tended to be a source of comfort, where family members mostly understood participants if their speech was different or would protect them from teasing, although some participants experienced teasing from siblings or parents. Overall, social isolation

in the form of bullying or teasing was often experienced, but changed over time with treatment and/or age.

Since the majority of patients were school-aged, many of the concepts related to social function at school. Relationships with teachers and other students were important to participants. Feeling accepted and not being bullied were key concepts for participants to be comfortable at school.

“They disturb him a bit... He can’t focus seriously at school. He thinks they see him differently from other students... like less.” (14 year old male with CLP, India, through a translator)

“There was also another guy in grade eight... on the very first day, he kind of like, told me to go get a nose job, and that got me... I started to cry, like not in public... I went to the bathroom and cried.” (17 year old female with CLP, Canada)

Changes with treatment were often manifested in the school setting:

“Like my voice, I used to not raise my hand, like, I’m starting gradually to raise my hand more instead of less.” (13 year old female with CLP, USA)

Speech-Related Distress

Speech-related distress related to the psychosocial impact of the speech difficulties. Participants described feeling embarrassed or nervous with speech, frustrated at having to repeat themselves, and being teased because of their speech:

“Sometimes she doesn’t speak because she gets ashamed that she might not be understood and the teacher might get angry with her.” (7 year old female with CLP, Philippines, through a translator)

Participants also described the impact of their speech function on their confidence, being around other people, the difference in their comfort with their speech around family and friends, meeting new people, participation in sports and singing, and worries about dating and work.

Facial Function

Facial function formed a third top-level domain, including the sub-domains of speech and eating/drinking. While participants were asked about breathing and hearing, we did not obtain a significant amount of data within these sub-domains and did not include them within the final conceptual framework.

Speech

Within the theme of speech function, participants identified the fact that they had a speech problem and then described the problem using concepts such as the ability to make certain sounds, say certain letters, or say certain words. Rather than using clinical terms, participants described hypernasality as sounding ‘nasal’ or ‘nasally’ and described their overall speech function with concepts such as needing to concentrate in order to speak well, the need to slow down their speech, speaking less or avoiding certain words, and the need to repeat themselves in

order for others to understand them. Participants often framed their difficulties with speech by how well others could understand them:

“My speech is not perfect, but it’s as good as it’s going to be, and other people can understand and not comment on it, which is what I need especially for work, for getting a job.” (18 year old male with CP, England)

As with the other sub-domains, participants also expressed if they either wanted their speech to change, or if their speech had changed after treatments.

Eating/Drinking

Eating/drinking was an important concept for participants, particularly for those from low- and middle-income countries. The theme of eating/drinking was mentioned in the context of food or drinks going up the nose, or coming out of the nose:

“...the rice gets into the hole... it gets into the nose.” (23 year old male with CLP, Philippines)

Participants also mentioned difficulty blowing or whistling, and the inability to drink through a straw. The etiology of problems with eating/drinking included symptoms consistent with a palatal fistula as well as suboptimal orthodontic alignment. Participants noted having to avoid certain foods, being unable to chew certain foods, and needing to take small bites when eating.

Other Facial Functions

Breathing, hearing, and pain were concepts that were raised, but not frequently discussed. With respect to breathing, participants described difficulty breathing through their noses, or breathing through their mouths. Only one participant recognized that she was snoring. Hearing was not frequently mentioned, but those that discussed hearing mentioned the need to have tubes in their ears, the risks of ear infections, and occasional difficulty with hearing their teachers or speaking loudly without knowing. Finally, some participants mentioned pain, particularly in the context of treatments such as having braces or immediately after surgery.

Discussion

This study presents a clear framework of concepts important to individuals with CL/P with respect to their outcomes from cleft care. Importantly, participants articulated what changes they were seeking with treatment, and were often specific in their descriptions of anatomic differences associated with CL/P. However, participants rarely used clinical terminology to describe their outcomes.

Appearance and speech were the most frequently discussed in the interviews. Individuals with CL/P may struggle with the psychological and social sequelae of having differences in their appearance and speech. The preliminary conceptual framework, based on a review of the literature, addressed these issues from a general perspective. The conceptual framework we present is more

specific and targeted, breaking appearance into seven parts of the face that are affected by CL/P and speech into speech function as well as speech-related distress. We hypothesize that the thorough qualitative inquiry described will guide the development of specific, clinically meaningful PRO measurement in the future. For example, if a patient presents for orthognathic surgery, it would be of greatest clinical meaning to use scales measuring appraisal of appearance of the jaws and face, speech function and speech-related distress, and psychological, social, and school well-being before and after surgery.

Participants described finding benefit from having a cleft in the interviews. When this concept was probed in subsequent interviews, participants were not enthusiastic when directly asked if they felt that positive attributes resulted from having a cleft. This concept was left off of the framework as a result. Benefit finding is an important concept worthy of much more discussion, but it is perhaps neither necessary nor appropriate to measure clinically over time with treatments.

The primary goal of the study was to create a list of concepts that should be measured from the patient perspective. The study has generalizability due to the large volume and scope of interviews performed. First, including individuals of different ages gave an indication of how concepts of interest may change over time. Second, by including individuals with CL/P from high-, low-, and middle-income countries, we were able to develop a conceptual framework based on themes identified across subgroups. While the frequency and manifestation of concepts such as appearance or facial function varied between countries, the

concepts themselves were similar amongst participants from different countries. For example, nasal symmetry was described differently between patients, but the *concept* of nasal symmetry was consistently raised. This was a valuable finding in creating a PRO measure that will be applicable internationally. Older adults with unrepaired clefts were included in low- and middle-income countries since they were seeking care; this population provided a unique and important perspective on the impact of having a cleft in their adult lives. While the goal of the study was not to describe differences between participants based on country of origin or cleft type, we found that the overall content of interviews with participants receiving cleft care later on in life focused more on concerns with function than appearance. Participants who had bilateral clefts also tended to discuss concepts related to symmetry less than those with unilateral clefts. All of these differences will be able to be measured in greater detail using the CLEFT-Q in the future.

One limitation of this study is that we were not able to sample participants from a broader range of countries. This will be addressed in the future field-test stage through cognitive interviews with patients in any new participating countries. Another limitation of this study was the unexpected need to interview children along with their parents in the low- and middle-income countries. We included the parents in cases where the child was more comfortable with them present. Since the interviews in languages other than English were carried out through a translator, this may have impacted the children's willingness to be interviewed alone. The concepts of hearing and breathing were raised, but not

endorsed consistently in subsequent interviews when questions probing these concepts were included. This may have been due to sampling bias. Other sources of potential sampling bias include the different methods of recruitment of participants high-income countries compared to low- and middle-income countries, and having fluency in English as an exclusion criterion in high-income countries.

We found that participants discussed appearance and speech in more detail than previous studies have described, which may reflect the content of the interview questions used. Stock and Feragen (2016) performed an extensive narrative review of both quantitative and qualitative studies examining psychological adjustment to CL/P. The domains described in the review included developmental trajectory, behavior, emotional well-being, social experiences, and satisfaction with appearance and treatment. Our findings complement these domains with additional detail with regards to appearance and speech. This study provides a detailed framework to guide ongoing outcome measurement from a patient's perspective over time and with treatment, such as improved oral health-related quality of life after completing cleft-related surgeries (Broder et al., 2017).

The conceptual framework derived from this study served as a guideline for developing preliminary scales for the CLEFT-Q. We have developed independent scales for each of the concepts of interest in the framework using the patients' own words as much as possible, and maintaining the lowest possible grade reading level. Although participants described many of the concepts in

negative terms, the scales were developed using either positive or neutral language in order to minimize any potential negative impact of completing the scales themselves. These scales are now being field-tested and we expect to find strong content validity as a result of this extensive qualitative study.

Conclusion

The appraisal of CL/P outcomes has typically focused on objective measurement such as observer- or clinician-reported outcome assessments (Heliovaara et al., 2016; Long et al., 2011; Semb et al., 2005). PRO instruments are a critically important adjunct to these objective measurements and help to create a more complete assessment of treatment outcomes by including the patient perspective. Findings from this qualitative study reveal that concepts important to individuals with CL/P with regards to their outcomes from cleft care fall into the top-level domains of appearance, health-related quality of life, and facial function. This conceptual framework has guided the development of a new PRO instrument called the CLEFT-Q.

Figure 1: Protocol for a multiphase mixed methods study to design the CLEFT-Q. The bolded components and products are the focus of this study.

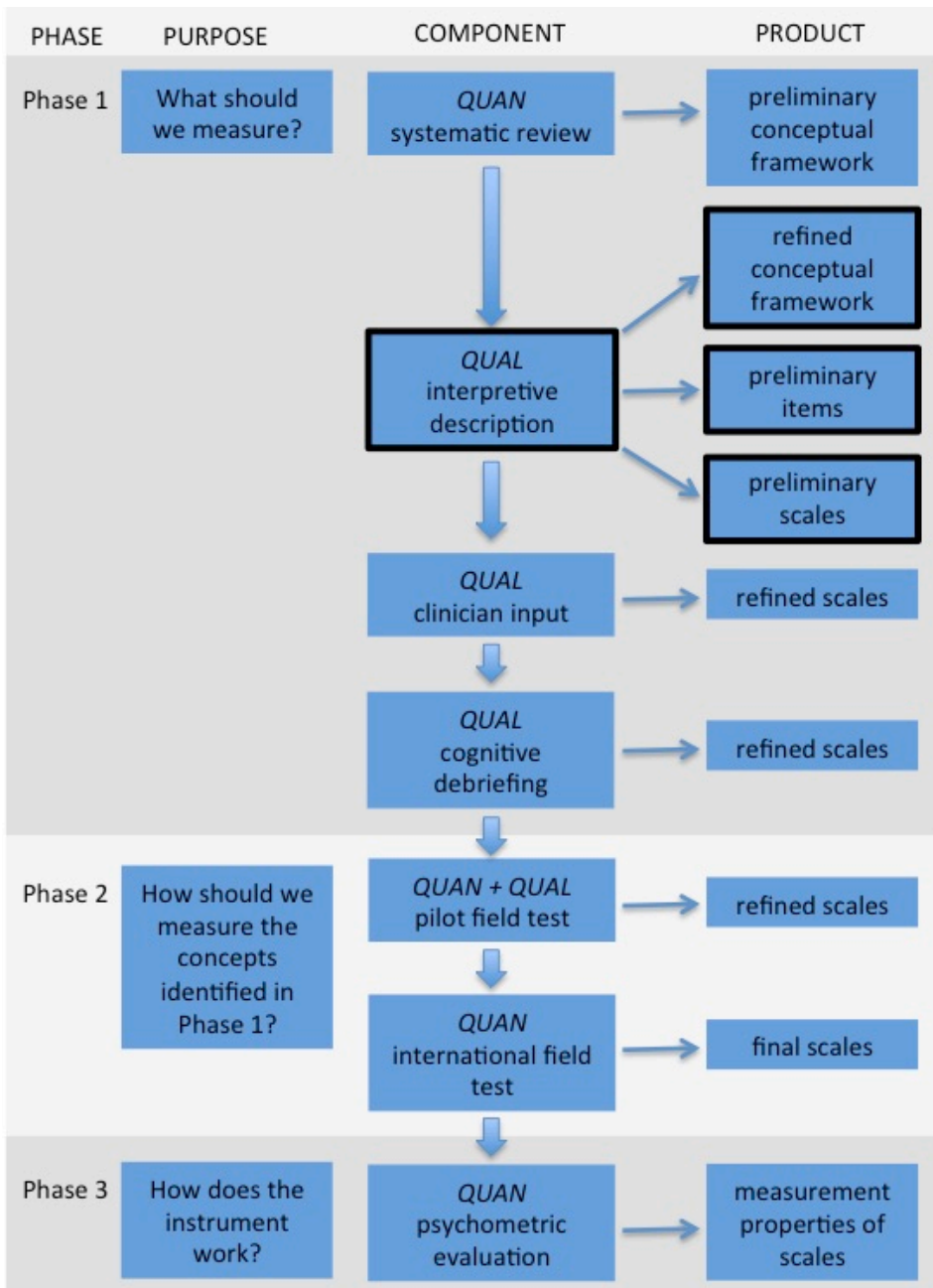


Figure 2: The preliminary conceptual framework was developed in a systematic review of the literature that identified concepts previously measured using PRO instruments in individuals with CL/P. This preliminary framework was revised through the study to generate the final conceptual framework based on participant input. The final conceptual framework includes three top-level domains with more specific sub-domains related to the treatment of CL/P. The sub-domains have since been developed into independently functioning scales in the CLEFT-Q.

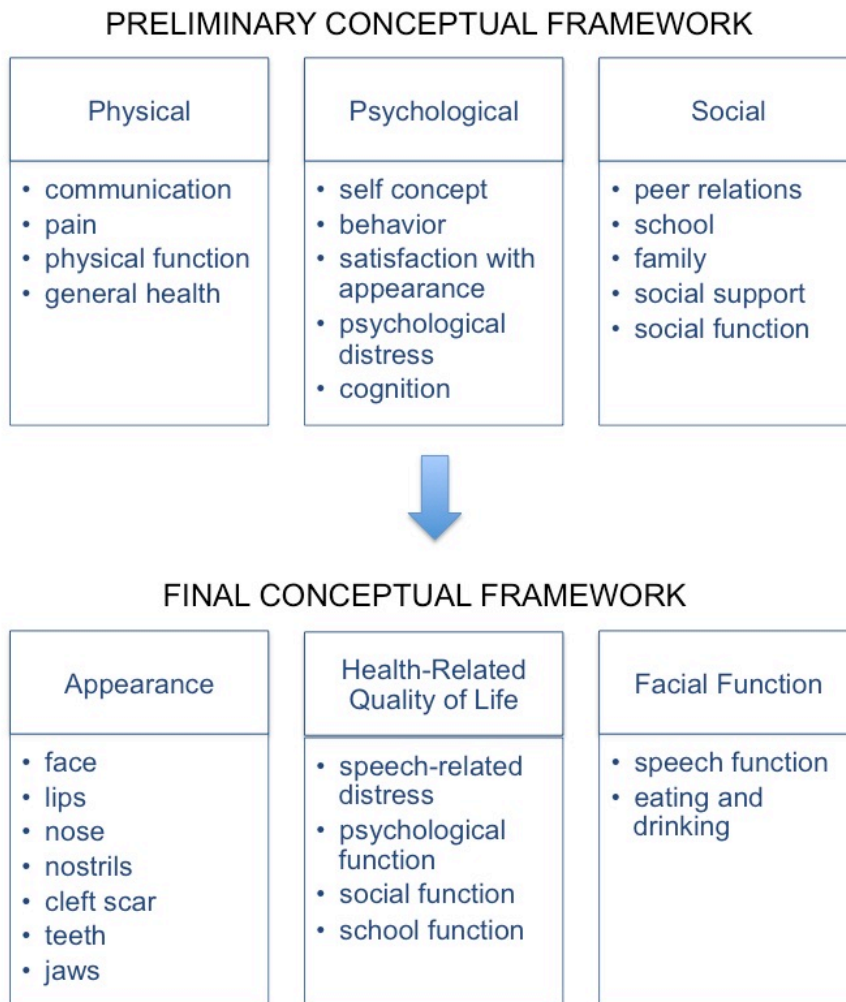


Table 1: The interview guide was developed from the preliminary conceptual framework. The guide was revised as interviews progressed to incorporate concepts that arose in previous interviews.

CLEFT-Q Qualitative Study Interview Guide

Preamble:

I would like you to tell me stories that will help me to understand what it's been like for you to have had cleft lip and/or palate (CLP) as a child or teenager. In particular, I am interested in how you feel about your appearance and whether or not it has had an effect on your life. I'm also interested in how you feel about the various surgeries and treatments you've had for CLP. We'd like to know how CLP has affected you and how we can make your treatment better.

Questions:

How old are you?

Tell me about yourself. What sorts of activities do you like to do?

Do you like going to school? What do you like or dislike about it?

How would you describe yourself?

Can you tell me about your friends?

How would your friends describe you?

Can you tell me about the details of your cleft lip and/or palate?

What treatments have you had for your cleft lip and/or palate?

Can you tell me about what having a cleft lip and/or palate means to you?

Can you tell me about how having a cleft lip and/or palate makes you feel?

How do you feel about your appearance?

How do you think other people feel about your appearance?

Do you have or have you had any problems with your speech? How does that affect you?

How often do you see the cleft team?

Can you tell me about each of your surgeries? Did you like the results?

How did the surgeries change your feelings about your appearance or speech?

Is there a member of the cleft team that you feel closely connected to?

Are you satisfied with the treatment you've had for your cleft lip and/or palate?

How would you change the treatment you've had for your cleft lip and/or palate?

Table 2: Characteristics of the participants interviewed.

	<i>Canada¹</i>	<i>UK²</i>	<i>USA³</i>	<i>Philippines⁴</i>	<i>India⁵</i>	<i>Kenya⁶</i>	<i>Total (%)</i>
Patients	25	19	9	30	23	32	138
Interviews	25	19	9	30	23	30	136
Age range in years (mean±SD)	10-22 (17±4)	7-21 (15±4)	8-18 (14±3)	5-23 (11±6)	<1-70 (13±15)	<1-56 (10±15)	<1-70 (13±10)
Gender							
Female	13	10	6	16	4	17	66 (48)
Male	12	9	3	14	19	15	72 (52)
Cleft Type							
CLP	19	14	9	23	11	20	96 (70)
CL/CLA	5	1	0	5	8	10	29 (21)
CP	1	4	0	2	4	2	13 (9)
Interview in English							
Yes	25	19	9	14	1	5	74 (54)
No	0	0	0	16	21	27	64 (46)

¹Hospital for Sick Children

²Spires Cleft Centre

³Rainbow Babies & Children's Hospital

⁴Operation Restore Hope NZ

⁵Operation Smile Guwahati Comprehensive Cleft Care Centre

⁶CURE International Children's Hospital

Table 3: An example of how the qualitative data were analyzed and categorized into domains and concepts. Some codes mapped to multiple domains as shown.

Country	Gender	Age	Cleft Type	Code from Transcript	Domain	Concepts	
UK	F	15	CLP	I used to like be <i>quite shy</i>	<u>Speech Function</u>	• having a problem	
				cause I <u>didn't</u> want to talk as		• speaking less	
				<u>much</u> cause I was worried that		• needing to repeat myself	
				<i>people would point it out</i> or		<u>Speech Distress</u>	• feeling shy
				something. And then <i>it'd</i> be			• feeling upset
				<i>upsetting</i> when <i>people ask me</i>			• around other people
<i>to repeat things.</i>	(people point it out)						
Philippines	F	21	CLP	(translator) ...she gets <i>really</i>	<u>Speech Function</u>	• having a problem	
				<i>nervous</i> , so sometimes like		• speaking less	
				she <u>just doesn't</u> talk because		• being understood	
				she's just <i>worried</i> that <i>people</i>		• feeling nervous	
				<u>won't understand her</u>		• feeling worried	
						• around other people	

References

- Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, Stein RE. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002; 11(3):193-205.
- Alansari R, Bedos C, Allison P. Living with cleft lip and palate: the treatment journey. *Cleft Palate Craniofac J* 2014; 51(2):222-9.
- Broder HL, Wilson-Genderson M, Sischo L. Oral health-related quality of life in youth receiving cleft-related surgery: self-report and proxy ratings. *Qual Life Res* 2017; 26(4):859-867.
- Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. *Ann Fam Med* 2008; 6(4):331-339.
- Eckstein DA, Wu RL, Akinbiyi T, Silver L, Taub PJ. Measuring quality of life in cleft lip and palate patients: currently available patient-reported outcomes measures. *Plast Reconstr Surg* 2011; 128(5):518e-526e.
- Endriga MC, Kapp-Simon KA. Psychological issues in craniofacial care: state of the art. *Cleft Palate Craniofac J* 1999; 36(1):3-11.
- Feragen KB, Stock NM. A longitudinal study of 340 young people with or without a visible difference: the impact of teasing on self-perceptions of appearance and depressive symptoms. *Body Image* 2016; 16:133-42.
- Hall MJ, Gibson BJ, James A, Rodd HD. Children's and adolescent's perspectives on cleft lip and/or palate. *Cleft Palate Craniofac J* 2013; 50(2):e18-26.

- Heliövaara A, Kuksela A, Skaare P, Shaw W, Molsted K, Karsten A, Brinck E, Rizell S, Marcusson A, Saele P, Hurmerinta K, Ronning E, Chalien MN, Mooney J, Eyres P, Semb G. SCANDCLEFT randomized trials of primary surgery for unilateral cleft lip and palate: 6. Dental arch relationships at 5 years. *J Plast Surg Hand Surg* 2016;25:1–6.
- Hunt O, Burden D, Hepper P, Johnston C. The psychosocial effects of cleft lip and palate: a systematic review. *Eur J Orthod* 2005; 27(3):274-285.
- Klassen AF, Tsangaris E, Forrest CR, Wong KW, Pusic AL, Cano SJ, Syed I, Dua M, Kainth S, Johnson J, et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. *J Plast Reconstr Aesthet Surg* 2012; 65(5):547-557.
- Long RE Jr, Hathaway R, Daskalogiannakis J, Mercado A, Russell K, Cohen M, Semb G, Shaw W. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. principles and study design. *Cleft Palate Craniofac J* 2011;48(3):239-43.
- Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. *Value Health* 2011;14(8):978-88.

Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L.

Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 1—eliciting concepts for a new PRO instrument. *Value Health* 2011;14(8):967-77.

Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ* 2000; 320(7227):114-116.

Porter ME. What is value in health care? *N Engl J Med* 2010; 363(26):2477-2481.

Ranganathan K, Shapiro D, Aliu O, Vercler CJ, Baker M, Kasten

SJ, Warschausky SA, Buchman SR, Waljee JF. Health-related quality of life and the desire for revision surgery among children with cleft lip and palate. *J Craniofac Surg* 2016; 27(7):1689-1693.

Sandelowski M. Theoretical saturation. In: Given LM, ed. *The Sage encyclopedia of qualitative methods*, Vol. 1. Thousand Oaks: Sage; 2008:pp. 875-876.

Semb G, Brattström V, Mølsted K, Prahll-Andersen B, Shaw WC. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. part 1: introduction and treatment experience. *Cleft Palate Craniofac J* 2005;42(1):64-8.

Sharif MO, Callery P, Tierney S. The perspectives of children and young people living with cleft lip and palate: a review of qualitative literature. *Cleft Palate Craniofac J* 2013; 50(3):297-304.

- Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology
- Stock NM, Feragen KB. Psychological adjustment to cleft lip and/or palate: a narrative review of the literature. *Psychol Health* 2016; 31(7):777-813.
- Stock NM, Feragen KB, Rumsey N. Adults' narratives of growing up with a cleft lip and/or palate: factors associated with psychological adjustment. *Cleft Palate Craniofac J* 2016; 53(2):222-39.
- Thompson A, Kent G. Adjusting to disfigurement: processes involved in dealing with being visibly different. *Clin Psychol Rev* 2001; 21(5):663-682.
- Thomas PT, Turner SR, Rumsey N, Dowell T, Sandy JR. Satisfaction with facial appearance among subjects affected by a cleft. *Cleft Palate Craniofac J* 1997; 34(3):226-231.
- Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. *Res Nurs Health* 1997; 20(2):169-177.
- Thorne SE. Interpretive description. Walnut Creek, CA: Left Coast Press, 2008.
- Turner SR, Thomas PW, Dowell T, Rumsey N, Sandy JR. Psychological outcomes amongst cleft patients and their families. *Br J Plast Surg* 1997; 50(1):1-9.
- US Food and Drug Administration Guidance for Industry. Patient-Reported Outcome Measures: use in medical product development to support labeling claims. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 2009.

Valderas JM, Alonso J, Guyatt GH. Measuring patient-reported outcomes:

moving from clinical trials into clinical practice. *Med J Aust* 2008;

189(2):93-94.

Welford C, Murphy K, Casey D. Demystifying nursing research terminology. Part

1. *Nurse Res* 2011; 18(4):38-43.

Wong Riff KW, Tsangaris E, Goodacre T, Forrest CR, Pusic AL, Cano SJ,

Klassen AF. International multiphase mixed methods study protocol to

develop a cross-cultural patient-reported outcome instrument for children

and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open* 2017;

7(1):e015467.

CHAPTER THREE

Preface

Chapter Three consists of a publication detailing the protocol for the entire development of the CLEFT-Q. This was published due to a recognized need for more understanding of the methods behind developing PRO measures.

The article has been published in BMJ Open. The Open Access article is distributed and licensed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license using the following citation:

Wong Riff KWY, Tsangaris E, Goodacre T, *et al* International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open* 2017;7:e015467. doi: 10.1136/bmjopen-2016-015467

My contribution to this study included study concept together with Anne Klassen, study design together with all co-authors, and writing the manuscript.

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

**An International Multiphase Mixed Methods Study Protocol to Develop a
Cross-Cultural Patient-Reported Outcome Instrument for Children and
Young Adults with Cleft Lip and/or Palate (CLEFT-Q)**

Karen W.Y. Wong Riff, MD MSc FRCSC

Division of Plastic and Reconstructive Surgery, Department of Surgery, Hospital
for Sick Children, University of Toronto, Toronto, Canada

Elena Tsangaris, MSc

Department of Clinical Epidemiology and Biostatistics, McMaster University,
Hamilton, Canada

Tim Goodacre, FRCS BSc MBBS

Spires Cleft Center, Oxford Radcliffe Children's Hospital, Oxford, UK

Christopher R. Forrest

Division of Plastic and Reconstructive Surgery, Department of Surgery, Hospital
for Sick Children, University of Toronto, Toronto, Canada

Andrea L. Pusic, MD MSc FRCSC

Department of Plastic and Reconstructive Surgery, Memorial Sloan-Kettering
Cancer Center, New York, USA

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

Stefan J. Cano, PhD CPsychol AFBPsS

Modus Outcomes, Letchworth Garden City, UK

Anne F. Klassen

Department of Pediatrics, McMaster University, Hamilton, Canada

Keywords: patient-reported outcomes, child health, cleft lip and/or palate,
psychometrics, quality of life, speech

Abstract

Introduction: Patient-reported outcome (PRO) instruments should be developed according to rigorous guidelines in order to provide clinically meaningful, scientifically sound measurement. Understanding the methodology behind instrument development informs the selection of the most appropriate tool. This mixed methods protocol describes the development of an internationally applicable PRO instrument, the CLEFT-Q, for evaluating outcomes of treatment for cleft lip and/or palate (CL/P).

Methods and analysis: The study includes 3 main phases that occur iteratively and interactively. In Phase 1, we determine what concepts are important to patients regarding their outcome. A conceptual framework for the CLEFT-Q is formed through a systematic review and an extensive international qualitative study. The systematic review ascertains what concepts have previously been measured in patients with CL/P. The qualitative study employs interpretive description, and involves in-depth interviews with patients in high- and lower-middle income countries. Preliminary items are generated from the qualitative data. Preliminary scales are then created for each theme in the framework. Cognitive debriefing interviews and expert clinician input are used to refine the scales in an iterative process. In Phase 2, the preliminary scales are administered to a large international group of patients with CL/P. The modern psychometric method of Rasch Measurement Theory (RMT) analysis is employed to define the measurement characteristics. The preliminary scales are shortened based on these

results. In Phase 3, further tests assess reliability, validity, and responsiveness of the instrument.

Ethics and dissemination: The study is approved by Research Ethics Boards for each participating site. Findings from this study will be published in open access peer-reviewed journals and presented at national and international conferences. Integrated knowledge translation is employed to engage stakeholders from the outset of the study. Successful execution of the CLEFT-Q will result in an internationally applicable PRO instrument for children and young adults with CL/P.

Strengths of this study:

- Multicenter, international study that includes patients in high- and lower-middle income countries will ensure the CLEFT-Q internationally applicable
- Extensive qualitative component of the study will ensure content validity of the CLEFT-Q
- Adherence to rigorous guidelines of instrument development and use of modern psychometric methods will make the CLEFT-Q as scientifically sound and clinically relevant as possible
- Multiple translations of the CLEFT-Q will be developed and tested

Limitations of this study:

- The scope of the study, which includes participants from high- and lower-middle income countries, necessitates a long time frame to completion

- Comparisons between patients with CL/P and those without should not be made without further research to establish if the CLEFT-Q has content validity in different populations
- The CLEFT-Q field-test will not include children with CL/P aged under 8 years

Introduction

Patient-reported outcomes (PROs) are increasingly important in the assessment of treatment effectiveness [1, 2]. If PRO data is to be used to drive quality improvement and treatment decisions, PROs should be evaluated in a scientifically sound manner using a PRO instrument developed according to rigorous guidelines [3]. The methodology behind the development or ‘validation’ of an instrument can be complex. A clear description and understanding of the methods can help to inform researchers selecting an appropriate PRO instrument for their target patient population.

Cleft lip and/or palate (CL/P) is the most common congenital craniofacial anomaly, with 7.94 cases per 10 000 live births annually [3]. The condition affects individuals worldwide and impacts an individual’s appearance, dentition, hearing, and speech. Treatment protocols vary widely, both within and between countries [4,5]. Observer-reported or clinician-reported outcomes form the majority of clinical outcome assessments (COAs) to date [6-8]. However, the goal of treatment of CL/P is to improve the patient’s physical, psychological, and social health, all of which are difficult to evaluate accurately with observer- or clinician-reported outcomes. Measuring these outcomes requires the patient perspective, but there is currently no comprehensive, specific PRO instrument for patients with CL/P available [9].

Beyond the scope of CL/P, few scales exist that measure appraisal of appearance from the patient perspective [10]. Congenital anomalies, trauma, and

other benign and malignant conditions can cause facial or other differences that are stigmatizing and may lead to social isolation. The treatment of these conditions addresses both form and function, yet the outcomes of treatment cannot be measured without appropriate PRO instruments that evaluate these concerns specifically and directly. The current study begins to fill this gap in measurement of appraisal of appearance from the patient perspective.

Many clinical conditions are prevalent around the world in high-income as well as low- and middle-income countries. Multinational studies are increasingly common, and PROs are frequently used as primary or secondary endpoints. The Consolidated Standards of Reporting Trials (CONSORT) recommendations for reporting randomized controlled trials have included a PRO extension to guide PRO reporting [2]. However, PRO instruments have typically been developed in a single language, and often in a single country [11]. Few PRO instruments have been designed for use in low- and middle-income countries [4]. While clinical outcome assessments such as clinician-reported or observer-reported outcomes are more easily compared between countries, it is difficult to compare PROs globally in the absence of instruments designed for global use. While guidelines exist for translation and cross-cultural adaptation of PRO instruments [11], the optimal design would be to develop the instrument in a cross-cultural manner from the outset.

Establishing scientifically sound, cross-cultural measurement tools involves a rigorous process. The following protocol describes the methodology

for an international study to develop a cross-cultural PRO instrument for children and young adults with CL/P, called the CLEFT-Q. To our knowledge, the CLEFT-Q will be the first international PRO measure that evaluates appraisal of appearance in addition to quality of life and function.

Methods and Analysis

Development of the CLEFT-Q follows the guidelines set forth by the Scientific Advisory Committee of the Medical Outcomes Trust [12], the United States Food and Drug Administration [13], and the International Society for Pharmacoeconomics and Outcomes Research [14, 15]. The aim is to develop a self-report instrument for patients 8 to 29 years of age that is internationally applicable, multidimensional (e.g., measures a number of different concepts of interest), and useful in clinical practice as well as in clinical audits and research.

The study employs a multiphase mixed methods approach, with an iterative combination of qualitative and quantitative inquiries [16]. Measurement properties of instruments fall into the 3 categories of (1) reliability, (2) validity, and (3) responsiveness. The Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist was designed to both ensure and evaluate validity and reliability in measuring health-related PROs [17, 18]. Similarly, a minimum standard for PRO instruments was outlined by members of the International Society for Quality of Life Research (ISOQOL) [19]. There are 3 main phases to developing a PRO instrument, including item

generation, item reduction, and psychometric evaluation, and these phases are carried out in an iterative and interactive manner as opposed to a linear progression (Figure 1). These 3 phases ensure that the resulting instrument fulfills the minimum standards outlined by ISOQOL as well as the COSMIN criteria for reliability and validity. The components of each phase are shown in Figure 2.

Phase 1: What Should We Measure?

The aims of Phase 1 are to establish content validity of the CLEFT-Q and to generate preliminary scales. First, a systematic review of the literature was performed to ensure that there was indeed no existing instrument available, and to define what PRO instruments have been validated and used in patients with CL/P in the past [20]. A comprehensive search following PRISMA guidelines yielded 4595 citations, of which 26 studies met inclusion criteria [20]. The studies were carried out in 9 high-income countries, confirming the lack of PRO measurement in low- and middle-income countries. Twenty-nine different PRO instruments were used in the 26 studies, and 20 measures were used only once. Based on these findings, a need for a comprehensive PRO instrument for CL/P exists, and we proceeded with the current study.

Conceptual Framework

The first step in Phase 1 is to develop a conceptual framework, or ‘a rationale for and description of the concepts and the populations that a measure is intended to assess and the relationship between those concepts’ [12]. From the systematic review performed at the outset of the study [20], concepts of interest

(COIs) that were previously measured are mapped to create a preliminary conceptual framework.

Qualitative Study

Next, a comprehensive qualitative study is carried out with participants with CL/P in high- and lower-middle income countries. The qualitative methodology employed is Interpretive Description, which seeks to generate relevant knowledge for a clinical context presuming that there is theoretical and clinical knowledge informing the study [21, 22]. For this study, the theoretical knowledge is derived from the systematic review, and clinical knowledge is derived from the team members carrying out the study. The philosophical underpinning of the qualitative study is pragmatism, meaning that the individual's understanding of a concept is of greatest importance, regardless of clinical explanations [23].

Participants, Setting, and Recruitment

Eliciting knowledge in high- and lower-middle income countries allows for cultural differences to be identified from the outset, facilitating accurate targeting of the scales in subsequent phases. The participating centers in this phase of the study are in 6 countries (Canada, Kenya, India, Philippines, UK, and USA). Recruitment takes place at cleft care centers. In the high-income countries (Canada, UK, USA), participants are recruited either through posters in clinics (and contacted by telephone to arrange an interview), or face-to-face in the clinical setting. In the lower-middle income countries (Kenya, India, Philippines),

a study team member recruits participants face-to-face in the clinical setting. Participants are eligible for inclusion if they have a diagnosis of CL/P. In the high-income countries, participants between 8 and 29 years of age are included. In the lower-middle income countries, participants of any age are included if they are presenting for clinical care to maximize the information gathered at these sites. In addition, parents of children with CL/P in the lower-middle income countries are invited to participate if the child prefers. This difference is important since a study team member, foreign to these countries, is present and working with a translator, which may make the child feel less comfortable if they are alone. Exclusion criteria include the inability to speak the language of the interviewer or translator in each country or a cognitive delay such that the individual cannot participate in a semi-structured interview.

Sampling

Participants are purposively sampled to gain a heterogeneous sample based on age, gender, and cleft type. Sampling continues until the point of saturation, when no further new concepts arise in subsequent interviews [24].

Data Collection

After obtaining written assent and/or consent as appropriate, a study team member trained in qualitative interviewing technique carries out individual, semi-structured interviews that are audio-recorded, using a translator in the lower-middle income countries as needed [24]. Participant age, gender, and cleft type is documented. An interview guide is developed based on the preliminary

conceptual framework, providing a list of open-ended questions for the interview. The interviewer probes new concepts as they arise. As standard qualitative methods dictate, data from interviews are analyzed on an ongoing basis, allowing for changes to be made to the interview guide for subsequent interviews to include new concepts that warrant further probing.

Data Analysis

Interviews are transcribed verbatim. Interviews performed through a translator, which would have language in both English and the target language, are again translated to English by a bilingual individual to confirm the translation. The interview data is then analyzed within NVivo8 software (QSR International Pty Ltd., 2012) using the line-by-line approach to coding data, with constant comparison used to identify and classify the concepts of interest identified. These concepts are then categorized into overarching domains with themes within the domains to refine the preliminary conceptual framework. Concurrent and iterative data collection and analysis are performed, allowing for changes to be made to the interview guide as new concepts arise. When no further new concepts are elicited from interviews, data collection ends and the conceptual framework is finalized. This conceptual framework represents all the concepts of interest to patients in 6 different countries with CL/P.

Rigor

Rigor in the qualitative study is ensured using several strategies. One team member performs data coding and a second team member then confirms the

analysis. By performing interviews in an iterative fashion, member-checking is employed to confirm that concepts identified are indeed valuable and important to participants with CL/P. Finally, peer debriefing is used to verify data analysis between members of the study team.

Item Generation

Coding of the qualitative data creates an exhaustive list of potential items to include in scales. A list of scales to be created is derived from the conceptual framework arising from the qualitative study. Each theme within the domains of the conceptual framework is turned into an individual preliminary scale. In this way, the entire suite of scales should cover all the concepts of interest to patients with CL/P. Individual scales are populated with items generated from the patients' own language whenever possible with the lowest feasible grade reading level (Fleisch-Kincaid level). Positive or neutral wording is adopted for the items in the scales as much as possible to limit any negative effects of filling out the CLEFT-Q in the future.

Refining the Preliminary Scales

The final stage of Phase 1 aims to refine the preliminary scales through an iterative process of returning to the patients to perform cognitive debriefing interviews [14] and obtaining expert multidisciplinary clinician input.

Cognitive Debriefing Interviews

Once the preliminary scales are formed, further semi-structured individual interviews are carried out to ensure that patients with CL/P understand the items on the scales and to confirm that no concepts are missing. Recruitment is carried out in a similar fashion to the qualitative study with participants from multiple countries to ensure cross-cultural input. Participants go through all the items on the preliminary scales with the interviewer using the ‘think aloud’ technique. The interviewer records items that are problematic and the reasons why these items are problematic. Cognitive debriefing interviews are carried out iteratively alongside obtaining expert clinician input as described below. Data from both sources are analyzed concurrently, again allowing for progressive improvements to the scales. Cognitive debriefing interviews follow a similar strategy as the qualitative study in that interviews continue until no further issues with the items on the scales arise.

Expert Clinician Input

Expert clinician input is sought to ensure that no further concepts should be included in the scales. Clinicians involved in cleft care from different disciplines (nursing, orthodontics, otolaryngology, pediatrics, psychology, social work, speech-language pathology, surgery) are purposively sampled from multiple countries through the networks of the study team. Focus groups with groups of clinicians are performed in a similar fashion to the cognitive debriefing interviews. In cases where focus groups cannot be performed, email input is

sought. The interviewer goes through all the items on the scales, looking for input on any missing items or on the wording of items. Again, data is analyzed concurrently with the cognitive debriefing interviews to refine the scales.

Translation

In the next phase of the study, the scales are field-tested in a large population of patients from multiple countries. The preliminary scales are translated into the necessary target languages according to guidelines set forth by the International Society for Pharmacoeconomics and Outcomes Research [25] and Mapi Research Trust [26]. Briefly, each translation is performed using 2 translators whose mother tongue is in the target language, and are fluent English. The 2 translators perform independent translations of the CLEFT-Q from English to the target language. Resulting translations are then reconciled to create a single translated version. A third individual whose mother tongue is English and is fluent in the target language then translates this version back into English, and this English version is compared to the original. The group then resolves the discrepancies together. The translated versions are then taken back to the patient population in further cognitive debriefing interviews to ensure that the meaning of the items, response options and instructions are the same, and that the wording is appropriate. At the end of this phase, a complete set of CLEFT-Q scales is ready to be tested in a population of patients.

Phase 2: What questions are effective in measuring the concepts identified in Phase 1?

The next phase of developing the CLEFT-Q involves field-testing the scales in a large population of patients with CL/P to determine which items on the scales are the most effective in measuring the concepts of interest. We employ the modern psychometric method of Rasch Measurement Theory (RMT) analysis to identify which items perform well on scales and to determine the measurement properties of the scales [27]. In order to provide rigorous measurement, the data must fit the requirements of a mathematical model, i.e., the Rasch model. Briefly, RMT creates a scale where an individual is placed along the scale based on the probability that he/she answered the questions or items in a certain way. This method contrasts with classical test theory, where scores are designed for group level analyses. This difference in mathematical modeling allows RMT analysis to provide an accurate individual person estimate. A RMT scale can be conceptualized as a ruler, with an ordered arrangement or hierarchy of items from a low to high ‘amount’ of the construct. RMT analysis creates interval-level measurement, or a scale where the notches on the scale are evenly spaced, as opposed to ordinal-level measurement, or a scale where the notches are not necessarily evenly spaced. Interval-level measurement allows for accurate tracking of change over time [28]. In addition, RMT analysis results in a scale that provides person estimates that are independent of the sampling distribution of the items. In other words, the scale functions the same way regardless of the people

that it is measuring, meaning that the same scale can be used accurately in different subsets of the target population (participants in different countries, or of different ages, for example).

Through the RMT analysis, the psychometric properties of the scale are defined. Items that are effective in measurement within the preliminary scales are then kept, and items that do not function as well in measurement or items that are identified as being redundant can be dropped. The final scales are created through this process of item reduction as described below.

Pilot Field-Test

A large-scale field-test of the CLEFT-Q is planned to take place in multiple countries. Since a multi-centered field-test is a resource-intensive endeavor, a pilot field-test is carried out at 2 sites in Ontario, Canada to identify any logistic obstacles and to perform an early preliminary Rasch Measurement Theory (RMT) analysis to troubleshoot any early issues with scale performance.

Study Participants

Patients with CL/P who are 8 to 29 years of age and who do not have a cognitive delay resulting in an inability to fill out the scales are recruited from 2 clinical settings in Canada. A minimum of 200 patients is required to perform the preliminary RMT analysis. Since this pilot study is meant to optimize the scales prior to the large-scale field-test, the preliminary RMT analysis may trigger further data collection and recruitment prior to finalizing the field-test versions of the scales.

Data Collection

Participants are asked to fill out the CLEFT-Q scales on paper and to give qualitative feedback in written format upon completion. Demographic characteristics including age, gender, cleft type, and stage of treatment are collected. Participants are also asked if they feel that the length of the entire CLEFT-Q is ‘about right’, ‘too long’, or ‘too short’. The time to complete the scales is recorded.

Data Analysis

Qualitative feedback is analyzed in a similar fashion to the cognitive debriefing interviews. Details of the RMT analysis are described in further detail below. The results from the qualitative and RMT analyses are used to further refine the scales. This iterative nature to scale development optimizes the likelihood that the scales will function well with minimal logistical obstacles in the ensuing large-scale field-test.

International Field-Test and RMT Analysis

The goal of the international field-test is to gather CLEFT-Q data from a large population of patients with CL/P internationally to define which items should be included in the final scales and to examine the measurement properties of the scales.

Study Participants

The international field test includes participants from 12 countries (Australia, Canada, Chile, Colombia, England, Ireland, India, Netherlands, Spain, Sweden, Turkey, USA). Centers are included based on interest and feasibility of recruiting the sample size required in a reasonable time frame. Participants with CL/P between the ages of 8 to 29 years are recruited to fill out the CLEFT-Q scales. Exclusion criteria include a cognitive delay resulting in the inability to complete the scales. Recruitment takes place either face-to-face or by mail depending on each center's preferences. The goal is to recruit a minimum of 108 from each country; a sample size from 108 to 200 results in item calibrations that are stable within 0.5 logits (person location estimates) with a 99% confidence interval [29].

Data Collection

The demographic characteristics collected are listed in Table 1. Participants will fill out the CLEFT-Q scales either on paper or on tablets in Research Electronic Data Capture (REDCap), a secure, web-based application for electronic data capture [30].

Data Analysis

Field-test data is entered into REDCap if participants filled out the scales on paper. Completed data files are then downloaded into IBM SPSS 22.0 [31]. The SPSS file is then imported into RUMM2030, the Rasch analysis software [32]. Each scale is analyzed independently. The psychometric function of each

scale is examined using a number of tests and various criteria. First, the thresholds for the item response options must be ordered, meaning that a '1' on a 4-point scale must sit lower in the continuum than a '2', and so on. The RMT analysis then defines the hierarchy of items on the scale, from the 'easiest' question for a patient to endorse to the 'hardest' question. Second, 3 item fit statistics are used to evaluate whether the items in a scale work together as a set: (1) log residuals, which represent item-person interaction; (2) chi-square values, which represent item-trait interaction; and (3) item characteristic curves. Items that are not functioning well with respect to these 3 statistics will be dropped from the scales unless they represent clinically important concepts. Third, the scale must be targeted to the population. The range of the construct measured by the scale is compared to the range of the construct experienced by the population, and maximal overlap is preferable to ensure that the scale can measure the construct in the population of interest.

The next component of the analysis ensures internal consistency, which refers to the interrelatedness amongst items on a scale. First, the scale is tested for unidimensionality, or whether the items on the scale all measure a single construct [33]. Second, the scale is evaluated using the Person Separation Index, a measure of the precision of a person estimate, which is a corollary of reliability (Cronbach's α) in classical test theory [34]. At any stage of the analysis, scales that are not functioning appropriately can be analyzed with poorly functioning

items dropped. This process continues until all the above statistics are within the acceptable range.

Differential Item Functioning (DIF)

Since the Rasch model creates a fixed ruler that is independent of the individual person estimates, differences between subgroups can be identified. DIF occurs when 1 subset of the target population answers a question differently than another subset [34]. In creating an international PRO instrument for children and young adults, differences based on country and age are an important consideration in creating scientifically sound instruments. In the field-test, DIF can be identified in RUMM2030 and items that show DIF can be dropped in the item reduction phase, or kept in with adjustments made to the scoring to account for the differences.

Item Reduction

From the item and threshold locations, the location of each question within the field-test scale on the overall ruler can be determined. Poorly functioning items can be dropped as described above, and extra items that measure in a similar fashion (showing residual correlations in the RMT analysis) can be dropped to develop a scale with the optimal number of items. It should be noted that at some point, further dropping of items will result in less precise measurement. The final decision regarding the optimal number of items depends on the distribution of the item locations as well as some clinical indication of a requirement for a certain degree of precision. Once item reduction is complete, the scales are finalized. The

RMT analysis then provides a scoring table for each scale, since calculating the score on each scale is more complex than simply summing the responses to each of the individual items.

Normative Data and Construct Validity

Once the scale scoring has been determined, scores are calculated for the field-test participants. Normative data and basic associations between scores and demographic characteristics can then be calculated using analysis of variance (ANOVA) in SPSS.

Construct validity includes the aspects of structural validity, which assesses internal relationships, hypotheses testing, and cross-cultural validity. Both structural validity and cross-cultural validity are addressed in the RMT analysis with unidimensionality and DIF, respectively. Hypotheses testing is used to establish whether the responses either correlate or differ in different patient groups in a way that would be expected [35]. In the CLEFT-Q, we test the following hypotheses: (1) that patients with a visible difference, i.e., CL and CL/P, will have lower scores on appraisal of appearance compared to those with an invisible difference (i.e., CP only); (2) that patients undergoing speech therapy or speech surgery will have lower scores on the speech scales than those not requiring any further intervention; (3) that patients requiring further treatment to the nose, lip, or jaw will have lower scores on the appearance scales as well as the quality of life scales compared to those not needing any further treatment; (4) that patients who rank their appraisal of their overall appearance or speech to be

higher ('like' their appearance more) on a 4-point scale will have higher scores on the appearance or speech and quality of life scales, respectively; and (5) that patients who are receiving psychological counseling or therapy will have lower scores on the quality of life scales. Analysis of variance in SPSS will be used to test these hypotheses.

Phase 3: How does the instrument work?

Several components of the COSMIN checklist are addressed in Phases 1 and 2 of development. Additional tests to ensure reliability, validity and responsiveness comprise Phase 3. All tests of the CLEFT-Q employ the finalized scales in this phase.

Reliability

Reliability includes 2 measurement concepts: (1) internal consistency, which is evaluated in Phase 2; and (2) test-retest reliability, which is evaluated in Phase III. To establish test-retest reliability, a smaller group of patients complete the CLEFT-Q scales, and then complete the scales again 1 week after the first administration. Scales that are reliable will have a minimum test-retest reliability of 0.70 in studies including at least 50 patients [36].

Validity

In the COSMIN checklist, the domain of validity includes 3 measurement properties, i.e., content validity, construct validity, and criterion validity [17, 18].

Content validity is addressed in Phase 1 of the study, and construct validity is addressed in Phase 2.

The final component of validity is criterion validity, or the degree to which the instrument reflects the findings on a ‘gold standard’ instrument [17]. When an instrument is comparable to similar instruments, concurrent validity is established. While we did not identify any single instrument as comprehensive as the CLEFT-Q, the aim of this sub study is to compare the results on the CLEFT-Q to 2 other instruments used in the past in patients with CL/P: (1) the Child Oral Health Impact Profile (COHIP) [37, 38], and (2) the CHASQ [39]. We hypothesize that CLEFT-Q scores for similar constructs will moderately correlate with the scores on these 3 instruments.

Responsiveness

Responsiveness evaluates the instrument’s ability to detect clinically meaningful change over time. The 2 main methods of evaluating responsiveness include an anchor-based and a distribution-based approach. In the anchor-based approach, patient-rated, clinician-rated, or condition-specific variables are used to estimate a minimally important difference (MID) for a scale [40]. The distribution-based approach estimates the MID based on the distribution of scores from a target population [40]. Techniques to evaluate responsiveness are debated in the literature [17]. RMT analysis has been shown to allow for increased detection of responsiveness [41]. We employ a variety of methods to best define responsiveness.

Study Participants

Participants for the test-retest reliability and criterion validity testing are recruited simultaneously. Again, participants from 8 to 29 years of age are recruited from the clinical setting with the same exclusion criteria as the field-test. Since this phase requires fewer numbers of participants (50), the number of participating centers is lower than the field-test (Canada, UK, USA). To study responsiveness, participants who are undergoing either (1) orthognathic surgery, (2) rhinoplasty, or (3) lip revision are recruited.

Data Collection

Participants fill out the CLEFT-Q scales in addition to the COHIP and the CHASQ on tablets through REDCap. Contact information is collected and participants are sent a link to complete the CLEFT-Q scales online 1 week later. Similar demographic data to the field-test is collected. For the responsiveness sub study, participants fill out the CLEFT-Q scales pre-operatively. Contact information is collected and participants are sent a link to the complete the CLEFT-Q scales again at least 6 months later.

Data Analysis

Test-Retest Reliability

CLEFT-Q scores are calculated from the 2 separate administrations of the scales for each participant. Test-retest reliability is then calculated in SPSS.

Criterion Validity

CLEFT-Q, COHIP, and CHASQ scores are calculated for each participant. Scores on each of the scales are then compared using a Pearson's r correlation in SPSS.

Responsiveness

Anchor-based techniques are used to calculate the MID from the transformed Rasch scores. To support the anchor-based methods, a distribution-based approach is used. The transformed Rasch scores are compared using paired t-tests, and then an effect size and standardized response means, 2 indicators of change, can be calculated [40, 41]. One of the strengths of RMT analysis is the ability to perform individual person level analyses for responsiveness. The tests listed above provide group-level comparisons. In individual person level comparison, the significance of a person's own change can be calculated using the individual person estimates, which are associated with bespoke standard errors [41]. Using both group level and individual level comparisons, the responsiveness of the CLEFT-Q can be defined as clearly as possible.

Ethics and Dissemination

Institutional ethics review board approval has been obtained for every participating center. Throughout the study, participants may be asked to discuss or answer questions about issues that are sensitive and may experience distress as a result. To address this concern, study team members explain during the consent

process that should this occur, an option to follow up with a clinical team member will be provided. Participants are also assured that all information is kept confidential; in the qualitative phase, interviews are transcribed with no identifying data, and in the qualitative phase, identifying data is kept in a separate file at each institution.

The intention of the study is not to directly compare different centers with respect to their outcomes. Any publications or presentations arising from this study will not identify specific centers.

An integrated knowledge translation approach is taken in this study. Collaborations with multiple sites internationally will hopefully result in increased uptake and use of the CLEFT-Q in the future. All Phase 2 and Phase 3 results for participants from each site will be sent back to the individual sites for their own use.

Finally, results of the study will be published in open access journals as required by the granting agency. Study team members will present the results at international and national conferences.

Figure 1: The phases of PRO instrument development. It is important to note that the phases can occur iteratively and interactively rather than in a linear progression.

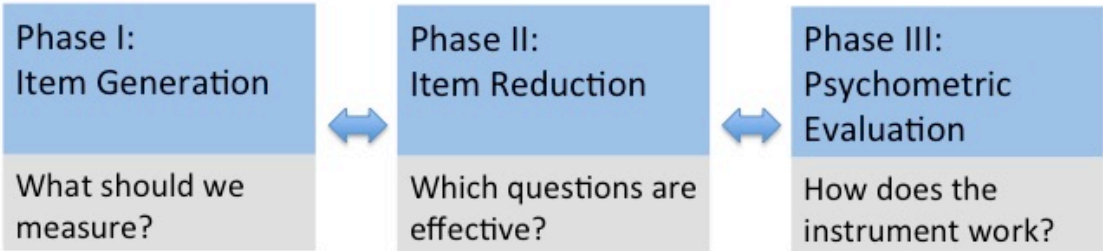


Figure 2: Flow diagram showing the multiphase mixed methods protocol for developing the CLEFT-Q. *QUAN* represents a quantitative study component, and *QUAL* represents a qualitative study component. It is important to note that the process can be iterative and interactive as opposed to strictly linear.

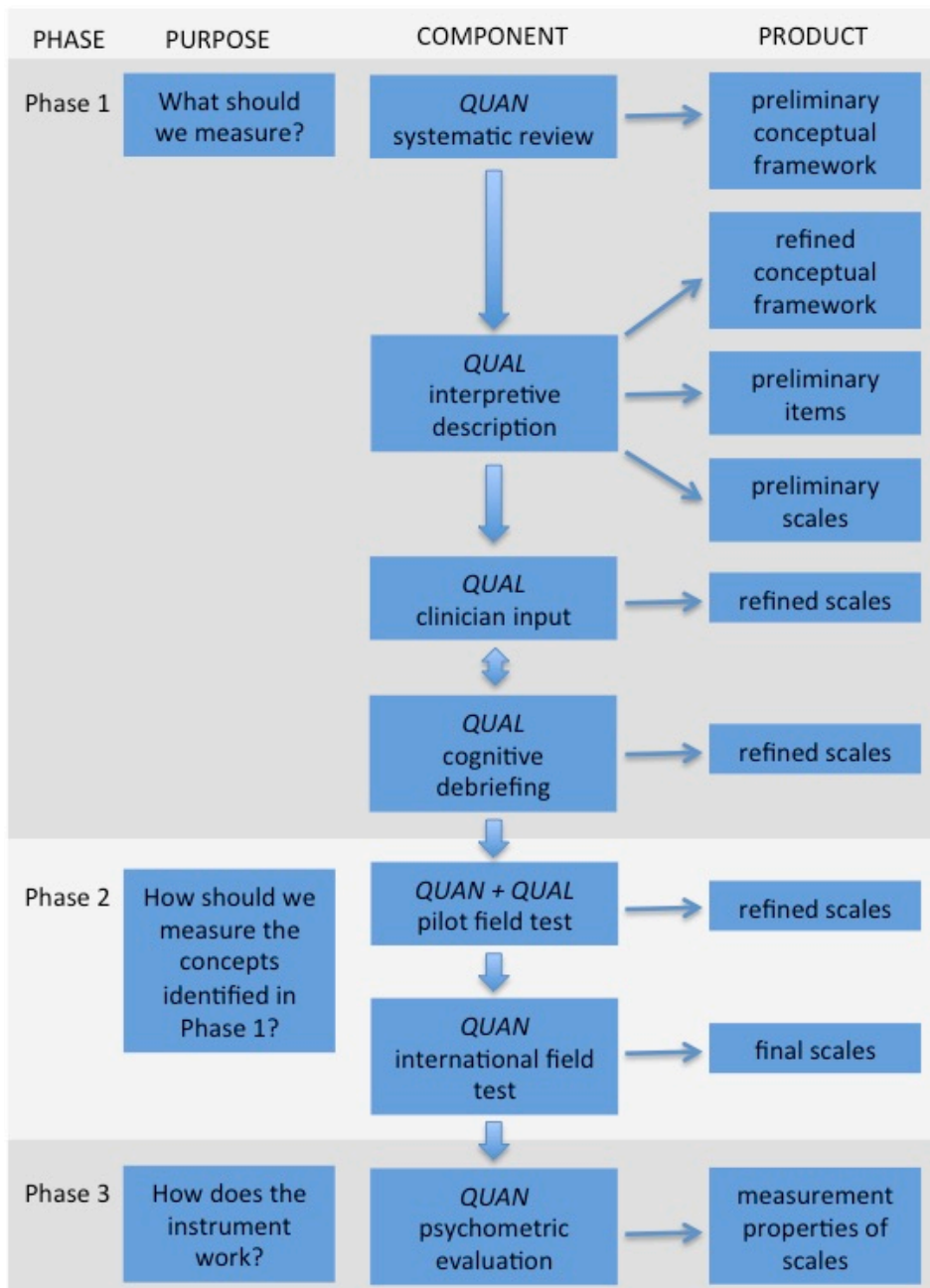


Table 1: Demographic characteristics collected for participants in the international field-test.

Demographic Characteristics

Age	Syndromes
Gender	Other Craniofacial Anomalies
Cleft Type	Developmental Disabilities
Country	Past Treatments
Student Status	Current Treatments
Language spoken at home	Future Treatments
Adopted	

References

1. Basch E. The missing voice of patients in drug-safety reporting. *The New England journal of medicine*. 2010;362(10):865-9.
doi:10.1056/NEJMp0911494.
2. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA: the journal of the American Medical Association*. 2013;309(8):814-22. doi:10.1001/jama.2013.879.
3. Tanaka SA, Mahabir RC, Jupiter DC, Menezes JM. Updating the epidemiology of cleft lip with or without cleft palate. *Plast Reconstr Surg*. 2012;129(3):511e-8e. doi:10.1097/PRS.0b013e3182402dd1.
4. Paltzer J, Barker E, Witt WP. Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: a review of existing measures. *Qual Life Res*. 2013;22(6):1177-87. doi:10.1007/s11136-012-0260-1.
5. Shaw WC, Semb G, Nelson P, Brattstrom V, Molsted K, Prah-Andersen B et al. The Eurocleft project 1996-2000: overview. *Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery*. 2001;29(3):131-40; discussion 41-2.
doi:10.1054/jcms.2001.0217.
6. Semb G, Brattstrom V, Molsted K, et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part

- 1: introduction and treatment experience. *Cleft Palate Craniofac J* 2005; 42:64-8.
7. Long RE, Jr., Hathaway R, Daskalogiannakis J, Mercado A, Russell K, Cohen M et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J*. 2011;48(3):239-43. doi:10.1597/09-180.1.
 8. Semb G, Enemark H, Friede H, et al. Scandcleft randomised trials of primary surgery for patients with unilateral complete cleft lip and palate: 1. Planning and management. *J Plast Surg and Hand Surg* 2017.
 9. Eckstein DA, Wu RL, Akinbiyi T, Silver L, Taub PJ. Measuring quality of life in cleft lip and palate patients: currently available patient-reported outcomes measures. *Plast Reconstr Surg*. 2011;128(5):518e-26e. doi:10.1097/PRS.0b013e31822b6a67.
 10. Wickert NM, Wong Riff K WY, Mansour M, Forrest CR, Goodacre TEE, Pusic AL, Klassen AF. Content Validity of Patient Reported Outcome Instruments Used with Pediatric Patients with Facial Differences: A Systematic Review. *Cleft Palate Craniofac J*. In Press.
 11. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value*

- in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2009;12(4):430-40. doi:10.1111/j.1524-4733.2008.00471.x.
12. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
13. Administration UFaD. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. In: Department of Health and Human Services FaDA, Center for Drug Evaluation and Research, editor. Rockville, MD, USA2009.
14. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011;14(8):978-88. doi:10.1016/j.jval.2011.06.013.
15. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument. *Value in health: the journal of the*

- International Society for Pharmacoeconomics and Outcomes Research.
2011;14(8):967-77. doi:10.1016/j.jval.2011.06.014.
16. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 2nd ed. Los Angeles: SAGE Publications; 2011.
17. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22. doi:10.1186/1471-2288-10-22.
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539-49. doi:10.1007/s11136-010-9606-8.
19. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res.* 2013;22(8):1889-905. doi:10.1007/s11136-012-0344-y.
20. Klassen AF, Tsangaris E, Forrest CR, Wong KW, Pusic AL, Cano SJ et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. *Journal of plastic, reconstructive & aesthetic surgery: JPRAS.* 2012;65(5):547-57. doi:10.1016/j.bjps.2011.11.004.

21. Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. *Research in nursing & health*. 1997;20(2):169-77.
22. Thorne SE. Interpretive description. *Developing qualitative inquiry*, vol 2. Walnut Creek, CA: Left Coast Press; 2008.
23. Welford C, Murphy K, Casey D. Demystifying nursing research terminology. Part 1. *Nurse Researcher*. 2011;18(4):38-43.
24. Sandelowski M. Theoretical saturation. In: Given LM, editor. *The Sage encyclopedia of qualitative methods*. Thousand Oaks: Sage; 2008. p. 875-6.
25. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2005;8(2):94-104. doi:10.1111/j.1524-4733.2005.04054.x.
26. Acquadro C, Conway K, Giroudet C, Mear I. *Linguistic validation manual for health outcome assessments*. Mapi Institute, Lyon, 2012.
27. Rasch G. Probabilistic models for some intelligence and attainment tests. *Studies in mathematical psychology*, vol 1. Copenhagen: 1960.
28. Cano SJ, Hobart JC. The problem with health measurement. *Patient preference and adherence*. 2011;5:279-90. doi:10.2147/PPA.S14399.

29. Linacre JM. Sample Size and Item Calibration (or Person Measure) Stability. *Rasch Measurement Transactions*. 1994;7(4):328.
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-81. doi:10.1016/j.jbi.2008.08.010.
31. IBM Corp. IBM SPSS Statistics for Windows. 22.0 ed. Armonk, NY: IBM Corp.; Released 2013.
32. RUMM Lab. <http://www.rummlab.com.au/>. Accessed October 10 2014.
33. Fayers PM, Hand DJ, Bjordal K, Groenvold M. Causal indicators in quality of life research. *Qual Life Res*. 1997;6(5):393-406.
34. Andrich D. Rasch models for measurement. Sage University papers series Quantitative applications in the social sciences, vol no 07-068. Newbury Park: Sage Publications; 1988.
35. Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Annu Rev Clin Psychol*. 2009;5:1-25. doi:10.1146/annurev.clinpsy.032408.153639.
36. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med*. 2001;20(21):3205-14.

37. Broder HL, Wilson-Genderson M. Reliability and convergent and discriminant validity of the Child Oral Health Impact Profile (COHIP Child's version). *Community dentistry and oral epidemiology*. 2007;35 Suppl 1:20-31. doi:10.1111/j.1600-0528.2007.0002.x.
38. Broder HL, Wilson-Genderson M, Sischo L. Reliability and validity testing for the Child Oral Health Impact Profile-Reduced (COHIP-SF 19). *J Public Health Dent*. 2012;72(4):302-12. doi:10.1111/j.1752-7325.2012.00338.x.
39. Emerson M, Spencer-Bowdage S, Bates A. Relationships between self-esteem, social experiences and satisfaction with appearance: standardisation and construct validation of two cleft audit measures. Presented at the Annual Scientific Conference of the Craniofacial Society of Great Britain and Ireland, April 2004; Bath, U.K.
40. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-9. doi:10.1016/j.jclinepi.2007.03.012.
41. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? *J Neurol Neurosurg Psychiatry*. 2010;81(9):1044-8. doi:10.1136/jnnp.2009.201392.

Authors' contributions

KWR and AK conceived of the study. KWR, AK, ET, TG, CF, AP, and SC participated in the design of the study. All authors read and approved the final manuscript.

Funding Statement

This work is supported by the Canadian Institutes of Health Research Open Operating Grant number 299133 to AK, KWR, AP, and CF, the Physicians' Services Incorporated Foundation Resident Research Grant to KWR, AK, and CF, and the Canadian Society of Plastic Surgeons Educational Foundation Grant to KWR, AK, and CF.

Acknowledgements

We would like to acknowledge our partners in cleft care centres around the world for participating in data collection in various phases of the study.

Competing Interests

KWR, ET, TG, CF, AP, and AK declare that they have no competing interests. SC started Modus Outcomes, a company that designs PRO instruments, after the inception of this study.

Data Sharing Statement

No additional data available.

CHAPTER FOUR

Preface

Chapter 4 consists of a manuscript describing how CLEFT-Q outcomes vary with cleft type, with an example of how the results can be visualized for clinical use. This article was written to demonstrate the clinical usefulness of the CLEFT-Q.

The manuscript is being submitted for publication in Plastic and Reconstructive Surgery.

My contribution to this study included study concept and design together with Anne Klassen, coordination of data acquisition at the Hospital for Sick Children, analysis and interpretation of data, and writing the manuscript.

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

**CLEFT-Q: Detecting Differences in Outcomes Between Patients with
Varying Cleft Types**

Karen W.Y. Wong Riff, MD MSc FRCSC

Division of Plastic and Reconstructive Surgery, Department of Surgery, Hospital
for Sick Children, University of Toronto, Toronto, Canada

Elena Tsangaris, PhD

Department of Clinical Epidemiology and Biostatistics, McMaster University,
Hamilton, Canada

Christopher R. Forrest

Division of Plastic and Reconstructive Surgery, Department of Surgery, Hospital
for Sick Children, University of Toronto, Toronto, Canada

Tim Goodacre, FRCS BSc MBBS

Spires Cleft Center, Oxford University Hospitals NHS Foundation Trust, Oxford,
United Kingdom

Natasha Longmire, MSc

Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

Gregory Allen, MD FAAP FACS

Associate Professor, Department of Otolaryngology – Head & Neck Surgery,
University of Colorado School of Medicine, Aurora, Colorado, USA

Doug Courtemanche, MD MS FRCSC

Division of Plastic Surgery, Department of Surgery, University of British
Columbia, BC Children's Hospital, Vancouver, British Columbia, Canada

Jesse Goldstein, MD

Department of Plastic Surgery, University of Pittsburgh Medical Center,
Children's Hospital of Pittsburgh of UPMC, Pittsburgh, Pennsylvania, USA

Aisling O'Mahony, MA BDentSc FDS MS(Oral Biology) DDS

Consultant in Restorative Dentistry, National Maxillofacial Unit, St. James'
Hospital, Dublin, Ireland

Andrea L. Pusic, MD MSc FRCSC

Department of Plastic and Reconstructive Surgery, Memorial Sloan-Kettering
Cancer Center, New York, USA

Rona Slator, FRCS FRCS(Plast) DPhil

Ph.D. Thesis – K. Wong; McMaster University – Health Research Methodology

West Midlands Cleft Service, Birmingham Children’s Hospital, Birmingham,
United Kingdom

Marc Swan, DPhil FRCS(Plast)

Spires Cleft Centre, Oxford University Hospitals NHS Foundation Trust, Oxford,
United Kingdom

Achilleas Thoma, BSc MD FRCSC FACS

Division of Plastic Surgery, Department of Surgery, McMaster University,
Hamilton, Ontario, Canada

Federico Vargas, MD

Fundación Operación Sonrisa Colombia; Division of Plastic Surgery, Universidad
el Bosque, Bogotá, Colombia

Anne F. Klassen, DPhil

Department of Pediatrics, McMaster University, Hamilton, Canada

Acknowledgements

The authors acknowledge all study collaborators for their participation in the study, and Dr. Peter Rosenbaum for critical review of the manuscript.

Abstract

Background: Measuring the patient perspective is important in evaluating effectiveness of cleft care. Understanding how treatment outcomes vary depending on cleft type may allow for better planning of treatments, setting of expectations, and more accurate benchmarking efforts. The CLEFT-Q is a patient-reported outcome measure for patients with cleft lip and/or palate (CL/P).

Methods: The twelve CLEFT-Q scales measuring appearance (face, nose, nostrils, lips, cleft lip scar, teeth, jaws), function (speech), and health-related quality of life (speech-related distress, social, school, psychological) were field-tested in thirty centres in twelve countries. Patients with CL/P aged 8-29 years were recruited from clinical settings. Differences in CLEFT-Q scores by cleft subtypes were evaluated using one-way ANOVA or Kruskal-Wallis H tests, with Tukey or Dunn's procedure with Bonferroni corrections post-hoc analyses, respectively. Scores are presented using radar charts in order to visualize all outcomes at once.

Results: The field-test included 2,434 patients. Scores on all twelve CLEFT-Q scales varied significantly with cleft subtype. Patients with unilateral or bilateral cleft lip and palate (CLP) scored lower on all appearance scales compared to patients with cleft palate or unilateral incomplete cleft lip. Scores on the speech function and speech-related distress scales decreased with each progressive group in the Veau classification. Patients with complete bilateral CLP scored the lowest on the social, school, and psychological scales.

Conclusions: Patient-reported outcomes measured with the CLEFT-Q vary significantly with cleft type. Visualizing multiple outcomes at once with radar charts allows for an understanding of a patient's overall status in a single graph.

Introduction

Cleft lip and/or palate (CL/P) affects appearance, speech, dentition, and hearing. Since the cleft affects different facial functions as well as appearance, many corresponding outcomes must be evaluated in assessing both the impact of the condition and the impact of its treatment(1). Previous reviews have highlighted the difficulty with developing evidence in cleft care due to the variation in outcomes evaluated(2, 3).

Patient-reported outcomes (PROs) have not been studied as often as so-called objective outcomes, partly due to the lack of a cleft-specific outcome instrument(4-7). Surgical intervention has been shown to improve oral health-related quality of life (HR-QOL) in children with CL/P(8). Many treatments in cleft care are targeted at improving one aspect of appearance or facial function (i.e., rhinoplasty targets appearance of the nose, or secondary speech surgery targets speech), and it is important to incorporate PRO instruments capable of measuring those specific aspects of having CL/P. Some treatments may also have a detrimental effect on other aspects of appearance or facial function (i.e., worsening speech following a LeFort I maxillary advancement aimed at improving appearance and dentition(9)). Assessing opposing effects of interventions from the patient perspective requires the use of PRO instruments that evaluate each of these aspects separately, in a clinically meaningful manner.

Interpreting and using PROs to aid in decision-making requires the results to be placed in context with other relevant PROs and objective outcomes. Ideally,

PROs would be collected routinely at regular clinical visits to allow for better understanding of a patient's status(10). However, incorporating PROs into clinical practice is challenging when there are multiple concepts that are important, as is the case in cleft care(11). The way in which all relevant outcomes are visualized by both patients and clinicians is an important consideration when creating PRO instruments to optimize future uptake.

Benchmarking using patient-reported indicators is becoming an important part of evaluating quality of health care(12, 13). There have been several efforts to standardize outcome evaluation between centres to allow for comparison between sites(1, 14-16). A standardized set of outcomes has been proposed to help with benchmarking efforts in the future(17). Since CL/P presents with varying severity, it is important to understand whether and how outcomes differ between cleft subtypes. Previous studies have evaluated variation in outcomes according to the main subtypes of cleft palate, cleft lip, cleft lip and alveolus, and cleft lip and palate(18-26). However, the clinical presentation varies within each of these broad subtypes, including whether or not the cleft is complete or incomplete, and unilateral or bilateral. If quality of care is to be compared between centres or populations, accounting for these differences should provide a more accurate representation of outcomes.

The CLEFT-Q is a patient-reported outcome (PRO) instrument developed internationally for children and young adults with CL/P using a mixed methods approach(11, 27, 28). Twelve independently functioning scales assess subdomains

within the top-level domains of appearance, facial function (speech), and HR-QOL. Initial analyses of the scales showed that scores varied with age, gender, and cleft type (classified as cleft palate, cleft lip, cleft lip and alveolus, and cleft lip and palate)(28). The primary objective of this study was to determine whether CLEFT-Q outcomes vary depending on specific cleft types as determined from the CLEFT-Q field-test. The secondary objective was to present a method of visualizing multiple outcomes at once in order to facilitate a better overall understanding of a patient's status.

Methods

The methodology behind the development of the CLEFT-Q is described in detail elsewhere(27, 29). The field-test was a phase in the overall study used to shorten the scales to their final form and to determine the psychometric properties of the scales. Once the scales were finalized, scores on each scale could be calculated for each participant in the field-test. A prospective, cross-sectional study was carried out in thirty centres in twelve countries. Research ethics board approval was obtained for each site.

Study Participants

Participants with CL/P aged 8 to 29 years were included in the study. Participants who were unable to complete the CLEFT-Q scales independently were excluded.

CLEFT-Q Scales

The CLEFT-Q includes 12 scales assessing appearance (face, nose, nostrils, lip, cleft lip scar, teeth, jaws), facial function (speech), and HR-QOL (psychological function, school function, social function, speech-related distress)(28). The scales were translated into Spanish (with different versions as appropriate for South American Spanish, Spanish, and Catalan), Hindi, Dutch, Swedish and Turkish(30, 31).

Data Collection

Participants were recruited from participating sites between October 2014 and November 2016, either in person or by mail. A site researcher completed a clinical form including identifying the cleft type. Clefts were classified as clefts of the secondary palate (CP), clefts of the lip and primary palate only (CL), and clefts of the lip, primary, and secondary palate (CLP). CL and CLP cleft types were also classified as unilateral or bilateral (U or B), and complete, incomplete, or asymmetric (C, I, or A). Cleft palates were also recorded using the Veau classification, indicating the severity of the cleft(32). Participants filled out all scales with the exception of the jaw scale (aged 12 to 29 years only), the cleft lip scar scale (those with a cleft lip only), the speech function and speech-related distress scales (those with a cleft alveolus and/or cleft palate only), and the school scale (aged 8 to 18 years and attending school only). Data were collected either on a tablet or on paper depending on the site preference and entered into REDCap (Research Electronic Data Capture)(33).

Data Analysis

Once the CLEFT-Q scales were finalized, CLEFT-Q scores for each participant were calculated by converting the logits to scores on a 0 to 100 scale, according to Klassen et al(34). Cleft type was determined from the clinical form. For the appearance, social, school, and psychological function scales, mean CLEFT-Q scores were calculated for the eleven subgroups of different cleft types. For the speech function and speech distress scales, the bilateral asymmetric cleft types were excluded since the clinical presentation and thus the impact of the cleft on speech were variable and not recorded. Normality and homogeneity of variances were checked using skewness and kurtosis values(35), and Levene's test, respectively. One-way ANOVA was used to identify whether there were significant differences in each of the CLEFT-Q scores between cleft types when the distributions of scores were found to be normal with homogeneous variance. Tukey post-hoc tests were used to further define differences. In cases where the distributions were not normally distributed, a non-parametric Kruskal-Wallis H test was performed with Dunn's procedure with Bonferroni corrections post hoc analyses. Groups containing less participants than the total number of groups plus one were excluded to comply with minimum sample size for ANOVA.

Visualization of CLEFT-Q Scores

In order to visualize multiple scale results in a clinically useful manner, we present the data using radar charts with CLEFT-Q outcomes represented as axes

or ‘spokes’. Radar charts have been used to depict clinical outcomes as well as cost simultaneously(36). Charts were created using Microsoft Excel 2011.

Results

The field-test included 2,434 patients from 12 countries. Demographic characteristics of the participants are shown in Table 1. The majority of patients completed the CLEFT-Q in English. Patients were evenly distributed across age groups except for the adult cohort (>21 years of age), where there were fewer patients. Fifty-five percent of participants were male. The distribution of cleft types is shown in Table 2. The most common cleft types were unilateral complete cleft lip and palate (34.3%), followed by cleft palate only (23.3%) and bilateral complete cleft lip and palate (14.6%).

The following groups were excluded from analyses due to there being fewer patients in the group than the total number of groups: patients with bilateral asymmetric cleft lip from all analyses, patients with bilateral asymmetric cleft lip and palate from analysis of the jaw score, and patients with bilateral incomplete cleft lip and palate from analyses of the speech function and speech-related distress scales.

Mean scores and standard deviations are shown in Table 3 and Figure 1. Test statistics and p-values are shown in Table 4. The Kruskal-Wallis H test was used for the school and psychological function scales as the distributions of groups were not found to be normal.

Appearance Scores

Scores on all appearance scales were significantly different between groups of patients with different cleft types (Table 4). Tukey's post-hoc analyses showed that patients with CLP-UC and CLP-BC scored the lowest on the appearance scales, significantly lower than patients with CP only and patients with CL-UI ($p < 0.05$ for all comparisons). Lip, nose, and nostril scores were the most markedly different between patients with visible and non-visible clefts. Patients with CLP-BC did not score significantly differently compared to patients with CLP-UC on any appearance scales except on the lip scale, where those with bilateral clefts scored lower. On both the teeth and jaw scales, patients with CLP-UC and CLP-BC scored lower than those with CP, CL-UC, and CL-UI.

Speech Function and Speech-Related Distress Scores

Speech function and speech-related distress scores were significantly different between groups of different cleft types as shown by one-way ANOVA (Table 4). Tukey post-hoc analysis showed that the patterns of differences were largely the same for both the speech function and speech-related distress scales. Patients with CL-UC or CL-UI had significantly better scores than patients with CP, CLP-UC, CLP-UI, and CLP-BC for both scales ($p < 0.05$ for all comparisons). Patients with CP had higher scores than those with CLP-BC ($p < 0.05$ for both scales). Speech function scores were lower in patients with CLP-BI compared to those with CL-UI ($p = 0.023$), but speech-related distress scores were not different between these two groups.

When speech function and speech-related distress scores were analyzed using the Veau classification of cleft palate types, one-way ANOVA showed that both speech function and speech-related distress varied significantly depending on Veau type (Table 4). Scores for both scales decreased with each Veau type (Figure 2). Tukey post-hoc analyses showed that patients with a Veau I or Veau II CP had significantly better speech function scores than those with a Veau IV CP. Patients with a Veau I CP had significantly better speech-related distress scores than those with Veau III and Veau IV clefts, and patients with Veau II CP also scored significantly higher than those with Veau IV clefts ($p < 0.05$ for all comparisons).

Health-Related Quality of Life Scores

Social function scores were statistically significantly different between groups on one-way ANOVA (Table 4). Tukey post-hoc analyses showed that patients with CLP-BC had lower scores than those with CP, CL-UC, and CL-UI ($p < 0.05$). Patients with CLP-UC had lower scores than those with CL-UC ($p = 0.024$). School and psychological function scores were statistically significantly different between groups as shown using the Kruskal-Wallis H test (Table 4). Dunn's post-hoc analyses of school function showed that patients with CLP-BC had lower scores than those with CL-UC and CL-UI ($p < 0.05$). For psychological function, patients with CLP-BC had lower scores than those with CP, CL-UI, and CLP-UI ($p < 0.05$). Patients with CLP-UC scored lower than those with CP ($p = 0.007$).

Visualizing CLEFT-Q Scores

Mean scores for each of the CLEFT-Q scales are shown in the radar charts in Figure 3.

Discussion

The primary aim of the study was to determine whether and how CLEFT-Q scores varied by cleft type. The clinical presentation of cleft lip and/or palate varies widely, and accounting for this variation would improve the evaluation of outcomes of cleft care. Our study showed that scores on all twelve CLEFT-Q scales varied significantly depending on cleft type in a large international sample of patients, with the overall tendency of patients with more severe clefts to score lower. This study reports statistical differences in CLEFT-Q scores with cleft types, and future psychometric evaluation will define what constitutes a meaningful clinically important difference. Previous studies assessing outcomes of cleft care have found differences by cleft type(18-21), whereas others have found no difference(22-26). This study uses a PRO instrument designed specifically for patients with cleft lip and/or palate and includes a larger international population of patients. A more detailed classification of cleft types was also employed. These details made it possible to explore outcomes in a manner not previously undertaken.

Evaluating aesthetic outcome objectively has been difficult due to the lack of consensus over a method of evaluation(37). While tools such as the Asher-

McDade scale have been employed in multi-centred studies such as Eurocleft and Americleft(1, 38, 39), a specific PRO instrument would provide valuable insight in conjunction with an objective measure. In the absence of an ideal objective measure, using a PRO instrument is of even greater value. Our study showed that patients with CLP-UC and CLP-BC had the lowest scores on the appearance scales. These results suggest that patients are reliable judges of their appearance since the pattern of differences is in keeping with the clinical severity of each type of cleft. The CLEFT-Q appearance scales provide a detailed evaluation of the patient's appraisal for the relevant anatomic components addressed in cleft care, thus maximizing clinical utility. The results of a lip revision could be evaluated using the CLEFT-Q lip and cleft scar scales pre- and post-operatively, together with photographs, in order to better understand the impact and success of treatment. Understanding the average scores for each cleft type may further help with planning and timing future interventions.

Objective speech outcomes have been correlated with the severity of cleft palate(40-42). Scores on the CLEFT-Q speech function and speech-related distress scales follow a similar pattern in our study, with patients with a Veau IV cleft scoring the lowest. These findings suggest that patients are able to self-report their own functional speech limitations, and that speech-related distress correlates to these findings.

Scores on the HR-QOL scales (social, school, and psychological function) were significantly different between groups with different cleft types overall. Post

hoc analyses did not reveal patterns of differences as clear as those for the appearance or speech scales, but the differences identified showed that patients with a more severe cleft type scored lower. A previous study using an ad hoc questionnaire to assess the social implications of cleft lip and/or palate by cleft type found that children with CLP and CP were found to differ from children without clefts to a similar extent(18). In our study, patients with CLP-BC scored the lowest on all three scales. Scores for all patients likely represent the combined effect of differences in appearance and speech, but it is also likely that other determinants are of clinical importance, such as socioeconomic status or family support.

The second aim of the study was to present a method of visualizing multiple different outcomes simultaneously in order to better understand a patient's status. Radar charts can provide a straightforward representation of all relevant outcomes. As an example, a patient who is undergoing orthognathic surgery may be at risk for developing velopharyngeal insufficiency. Without individual measures for appearance and speech, the overall impact of surgery would be difficult to define. Using a radar chart with PROs in addition to any relevant objective outcomes would facilitate the identification of patients at higher risk pre-operatively, as well as those who develop worsening speech post-operatively. Radar charts can be used in clinical visits for any condition with multiple important outcomes to give care providers an initial sense of a patient's

status, perhaps facilitating better communication and improved decision making (see Figure 3d).

One limitation of this study is that patients were recruited to form a convenience sample for validation of the CLEFT-Q scales, which may result in selection bias. The large sample size and the variety of patients from many cleft care centres in different countries may help to reduce this bias somewhat. Most patients were also those seeking clinical care, as recruitment often took place in the clinical setting. While this type of recruitment may also be a source of bias since patients who are no longer in treatment were not included, the CLEFT-Q is designed for clinical use and the patients included reflect people like those that will be using the scales in the future. Despite the large sample size, there were small numbers in some of the groups of cleft types due to rarity in presentation or perhaps less frequent presentation for clinical care.

Among the main strengths of this study are its internationality and the large sample size. Including patients from thirty centres in twelve countries (see Table 1) ensured a heterogeneous sample, with patients of various backgrounds following various treatment protocols. Another strength is that individual components of overall outcome of cleft care were measured separately using the twelve CLEFT-Q scales rather than combined in a single cleft outcome measure, allowing for the independent evaluation of these components from the patient's perspective. The quality of the data was further strengthened by using scales

developed using modern psychometrics, meaning that the scores are valid at the individual level, and that the scores provide interval-level measurement(28).

As benchmarking efforts continue and multi-centred studies become more common, quality improvement will depend on established normative or standard data. In the case of CL/P, the condition presents with broad variation, and our study shows that PROs depend on cleft type. Evaluations of cleft care centres should account for cleft type in order to produce valid comparisons. In addition, accurate representation of outcomes requires thoughtful application of all relevant CLEFT-Q scales. While PROs are an essential component of final outcomes of cleft care, caution must be exercised when using PROs as performance measures, as PROs may be influenced by external factors (e.g., socioeconomic status) that are not modifiable by the health care team(43).

Conclusions

The CLEFT-Q is a PRO instrument that includes twelve independently functioning scales. In a large international sample, scores on all scales varied with cleft type severity, with patients with bilateral complete CLP scoring the lowest of all groups. Data are presented using radar charts, providing a method of visualizing all relevant outcomes together that may improve understanding of patterns of scores across different outcomes. The CLEFT-Q scores may be used to inform future quality improvement efforts and benchmarking according to cleft type.

Table 1: Demographic characteristics of participants in the CLEFT-Q field-test.

	N	%
Country		
Canada	624	25.6
United States	362	14.9
England	339	14.0
India	232	9.5
Colombia	210	8.6
Netherlands	206	8.5
Ireland	100	4.1
Sweden	100	4.1
Spain	93	3.8
Chile	89	3.7
Turkey	54	2.2
Australia	25	1.0
Language		
English	1450	59.6
Spanish	354	14.7
Hindi	232	9.5
Dutch	206	8.5
Swedish	100	4.1
Turkish	54	2.2
Catalan	38	1.4
Age in years		
8-9	426	17.5
10-11	411	16.9
12-13	372	15.3
14-15	385	15.8
16-17	293	12.0
18-20	300	12.3
≥21	246	10.1
Missing	1	0.1
Gender		
Male	1351	55.5
Female	1081	44.4
Missing	2	0.1

Table 2: Distribution of cleft types in the sample population: a) overall cleft types, and b) distribution of patients with cleft palate by the Veau classification.

Table 2a: Distribution of overall cleft types.

<i>Cleft Type</i>	<i>N</i>	<i>%</i>
Cleft Lip and Palate	1399	57.5
Unilateral Complete (CLP-UC)	834	34.3
Unilateral Incomplete (CLP-UI)	108	4.4
Bilateral Complete (CLP-BC)	355	14.6
Bilateral Incomplete (CLP-BI)	28	1.2
Bilateral Asymmetric (CLP-BA)	22	0.01
Missing	52	0.02
Cleft Lip	467	19.2
Unilateral Complete (CL-UC)	195	8.0
Unilateral Incomplete (CL-UI)	183	7.5
Bilateral Complete (CL-BC)	33	1.4
Bilateral Incomplete (CL-BI)	22	0.01
Bilateral Asymmetric (CL-BA)	8	0.00
Missing	26	0.01
Cleft Palate (CP)	568	23.3
Total	2434	100

Table 2b: Distribution of patients with cleft palate by the Veau classification.

<i>Cleft Palate Type</i>	<i>N</i>	<i>%</i>
Veau I	258	13.1
Veau II	287	14.6
Veau III	943	47.9
Veau IV	394	20.0
Missing	85	4.3
Total	1967	100

Table 3: Mean scores for the CLEFT-Q scales by cleft subtype. The cleft subtypes are ordered by overall types, with CP first, followed by the CLP and then CL subtypes. Cleft subtypes include: CP, cleft palate; CLP-UC, unilateral complete cleft lip and palate; CLP-UI, unilateral incomplete cleft lip and palate; CLP-BC, bilateral complete cleft lip and palate; CLP-BI, bilateral incomplete cleft lip and palate; CLP-BA, bilateral asymmetric cleft lip and palate; CL-UC, unilateral complete cleft lip; CL-UI, unilateral incomplete cleft lip; CL-BC, bilateral complete cleft lip; and CL-BI, bilateral incomplete cleft lip. Patients with CL-BA, bilateral asymmetric cleft lip, were excluded from these analyses as there were only 8 patients in this group. n, number of cases; SD, standard deviation.

	CP	CLP-UC	CLP-UI	CLP-BC	CLP-BI	CL-UC	CL-UI	CL-BC	CL-BI	CLP-BA
Face										
n	549	830	107	355	28	193	178	33	22	21
mean	71	60	66	58	63	63	67	60	70	70
SD	19	19	18	19	24	19	18	21	20	19
Nose										
n	499	807	105	343	28	193	175	32	21	21
mean	72	53	60	52	59	56	64	53	66	63
SD	21	22	20	22	28	22	22	27	22	25
Nostrils										
n	493	802	104	341	28	191	174	32	20	20
mean	75	47	55	50	62	50	60	48	67	70
SD	25	27	29	29	32	27	27	29	29	30
Lips										
n	439	797	102	338	28	193	173	32	20	20
mean	74	59	64	53	57	61	68	56	65	71
SD	23	24	24	25	29	23	21	27	24	28
Cleft Scar										
n	-	771	98	330	27	188	164	32	20	19
mean	-	58	62	54	57	56	65	53	70	73
SD	-	28	27	30	29	28	25	31	25	29
Teeth										
n	535	798	103	337	28	191	177	32	20	20
mean	60	52	54	49	48	60	62	53	58	59
SD	24	23	20	24	25	23	23	26	29	24

Jaws										
n	308	537	58	238	18	112	103	24	12	-
mean	71	65	70	62	71	75	78	62	83	-
SD	26	27	22	28	30	24	23	27	23	-
Speech Function										
n	492	744	94	326	27	80	26	14	-	-
mean	70	68	68	65	69	81	88	71	-	-
SD	22	21	19	20	23	21	15	21	-	-
Speech Distress										
n	498	757	96	325	27	80	27	14	-	-
mean	71	68	66	66	67	79	84	67	-	-
SD	21	21	20	21	20	19	16	28	-	-
Social										
n	515	790	100	337	28	187	168	33	20	17
mean	75	72	74	69	68	76	77	71	75	79
SD	17	17	18	18	17	16	15	18	15	18
School										
n	410	558	70	234	22	142	135	21	19	14
median	76	75	77	72	70	80	79	75	73	80
SD	18	18	19	18	23	16	16	17	15	19
Psychological										
n	515	783	98	337	28	188	168	33	20	17
median	77	72	78	70	67	75	76	71	78	79
SD	18	19	17	20	21	18	17	21	15	21

Figure 1: Boxplots showing the distribution of scores for each CLEFT-Q scale.

The bottom of the box represents the 25th percentile and the top of the box represents the 75th percentile. The whiskers represent 1.5 times the interquartile range. A circle represents an outlier, and an asterisk represents an extreme outlier.

Figure 1a: Boxplot of CLEFT-Q Face scores by cleft type.

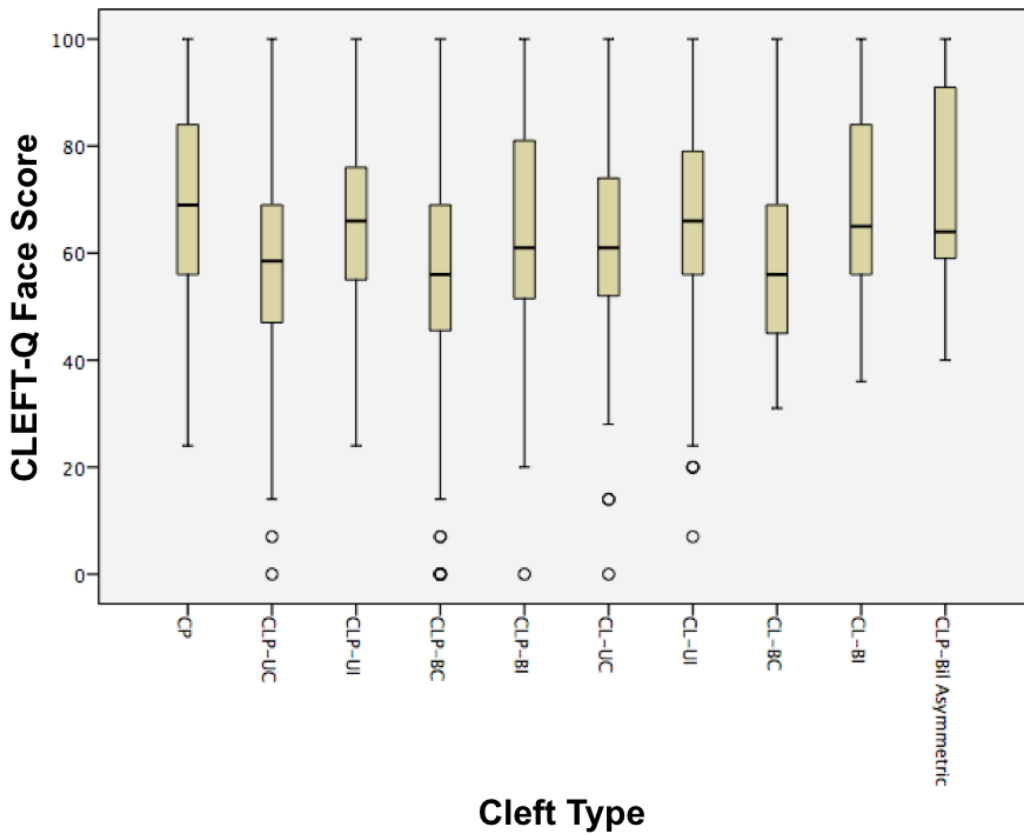


Figure 1b: Boxplot of CLEFT-Q Nose scores by cleft type.

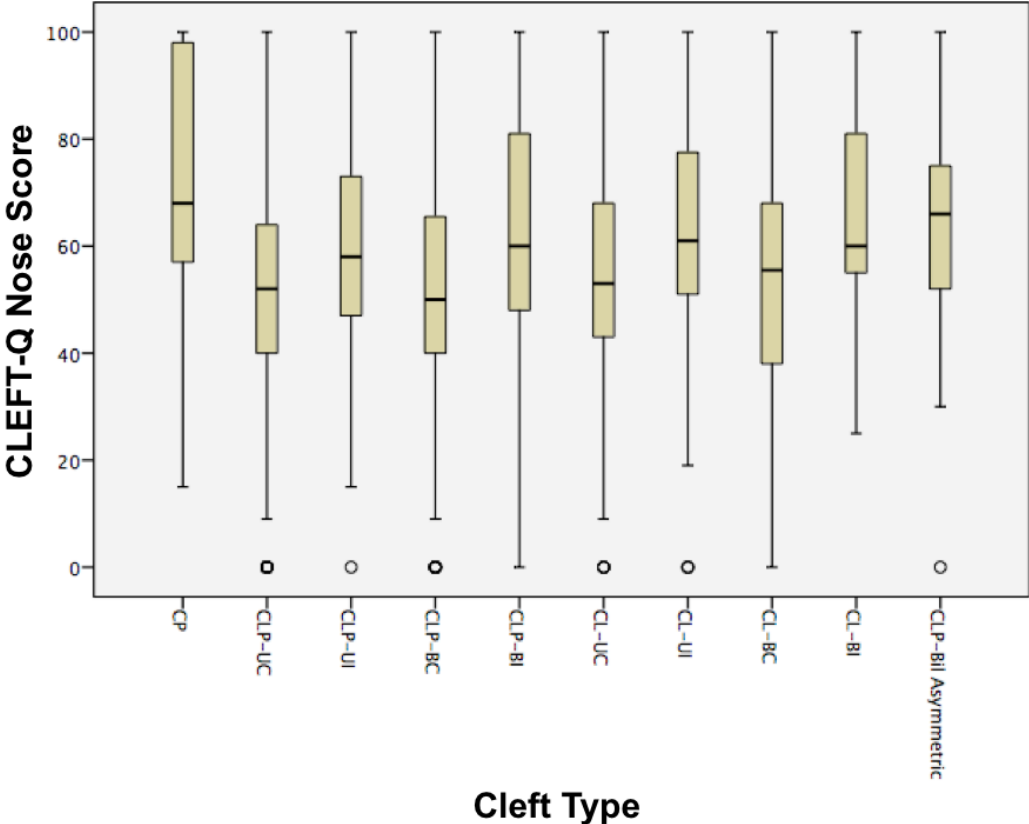


Figure 1c: Boxplot of CLEFT-Q Nostrils scores by cleft type.

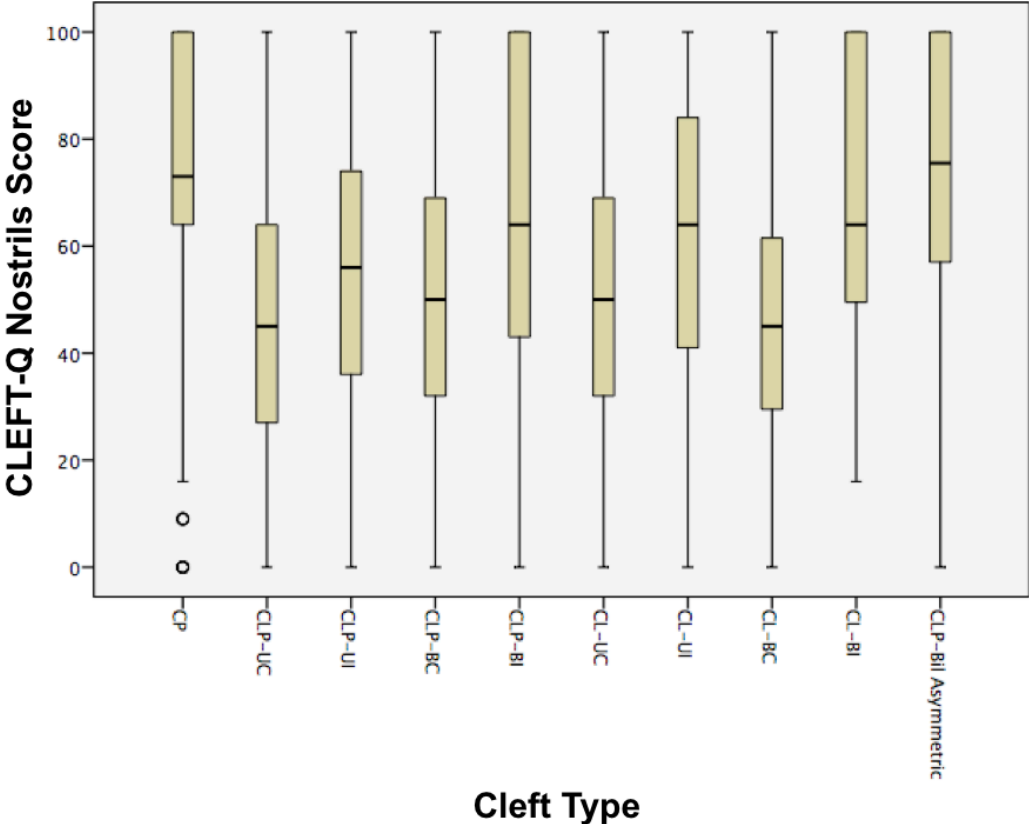


Figure 1d: Boxplot of CLEFT-Q Lips scores by cleft type.

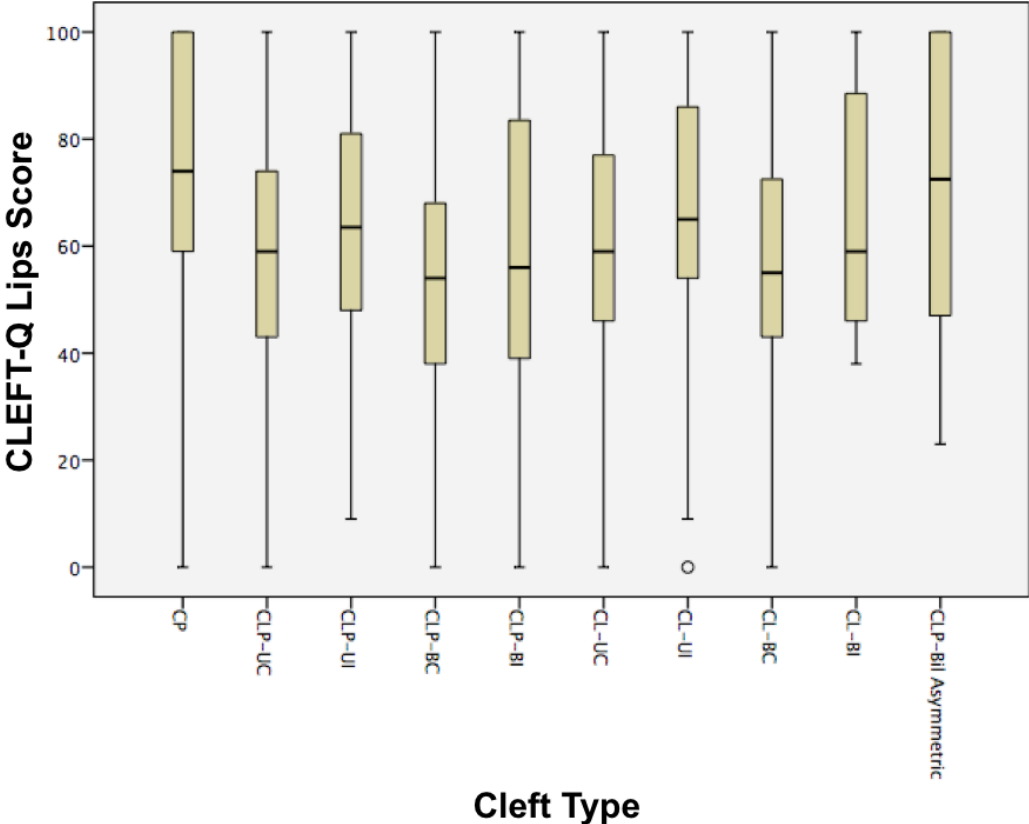


Figure 1e: Boxplot of CLEFT-Q Cleft Lip Scar scores by cleft type.

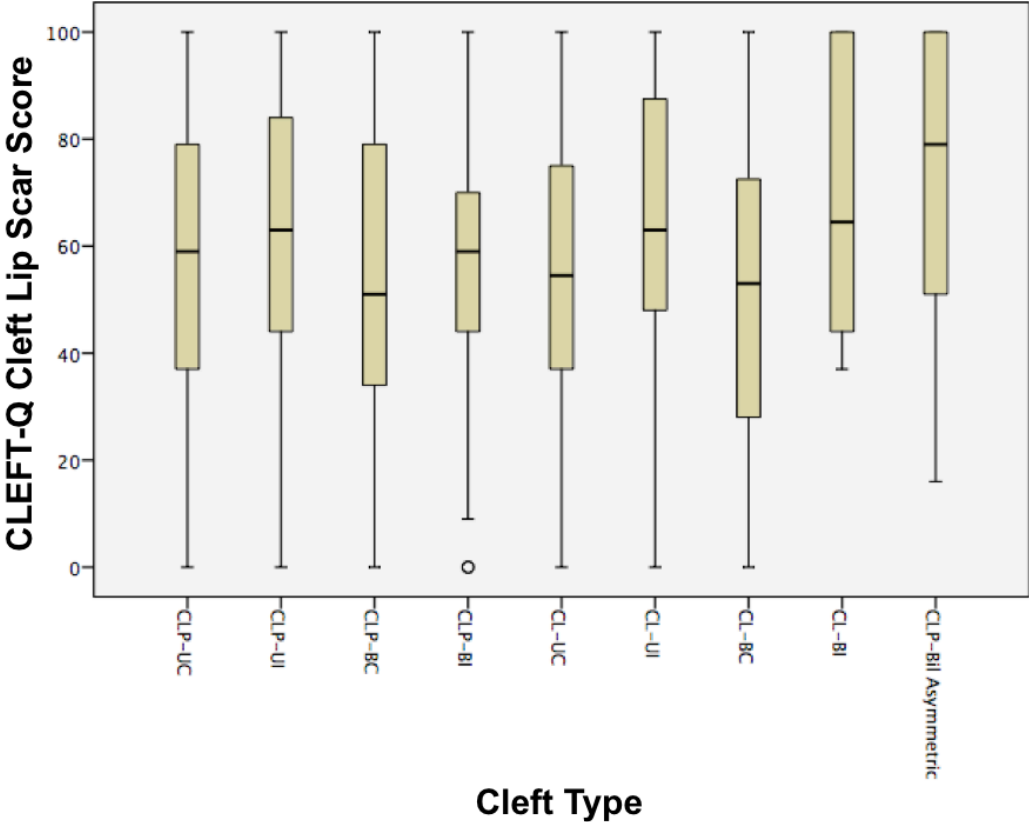


Figure 1f: Boxplot of CLEFT-Q Teeth scores by cleft type.

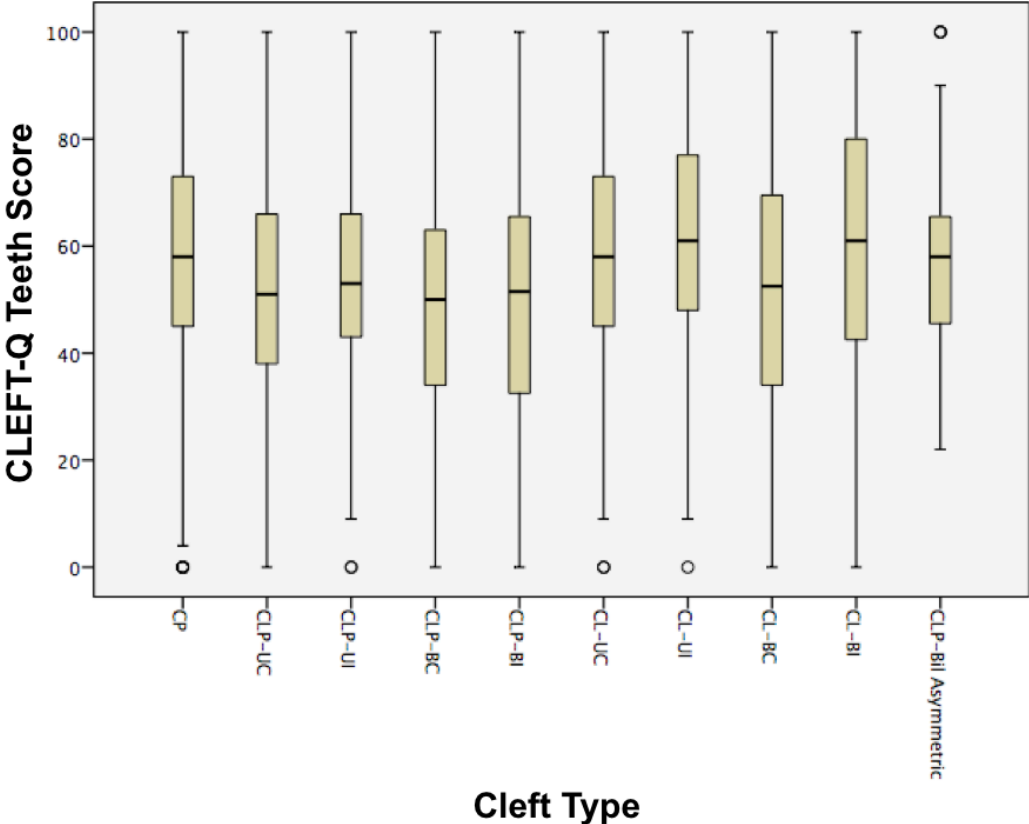


Figure 1g: Boxplot of CLEFT-Q Jaws scores by cleft type.

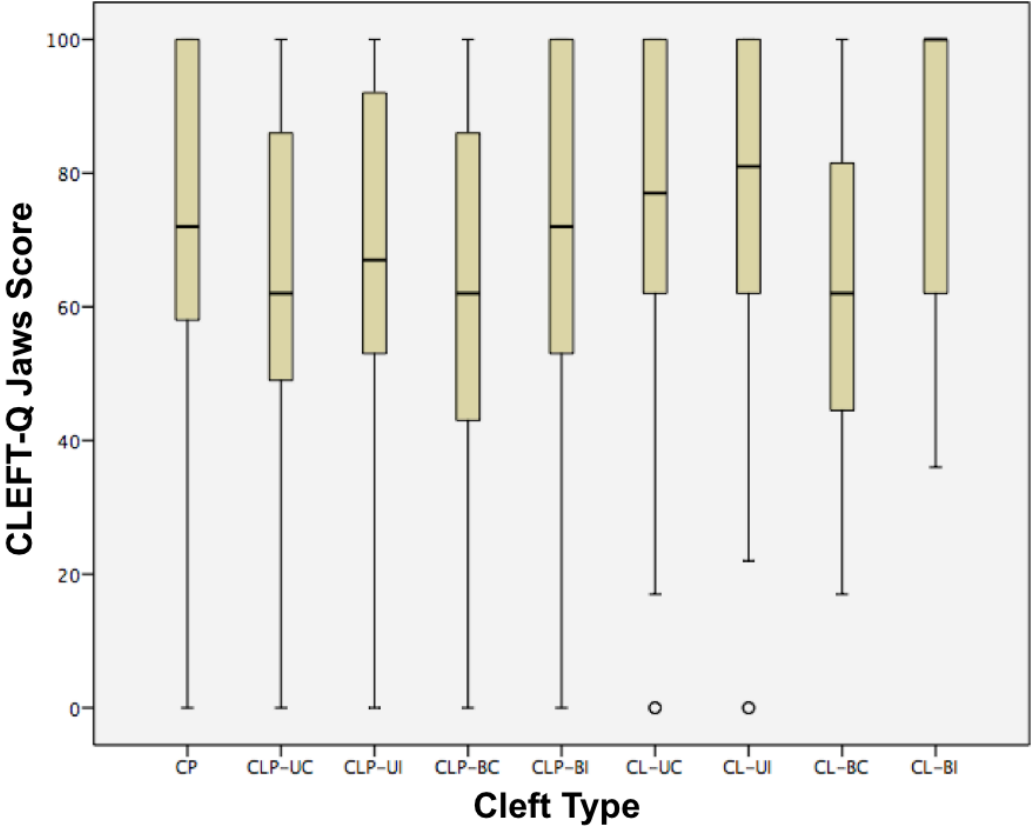


Figure 1h: Boxplot of CLEFT-Q Speech Function scores by cleft type.

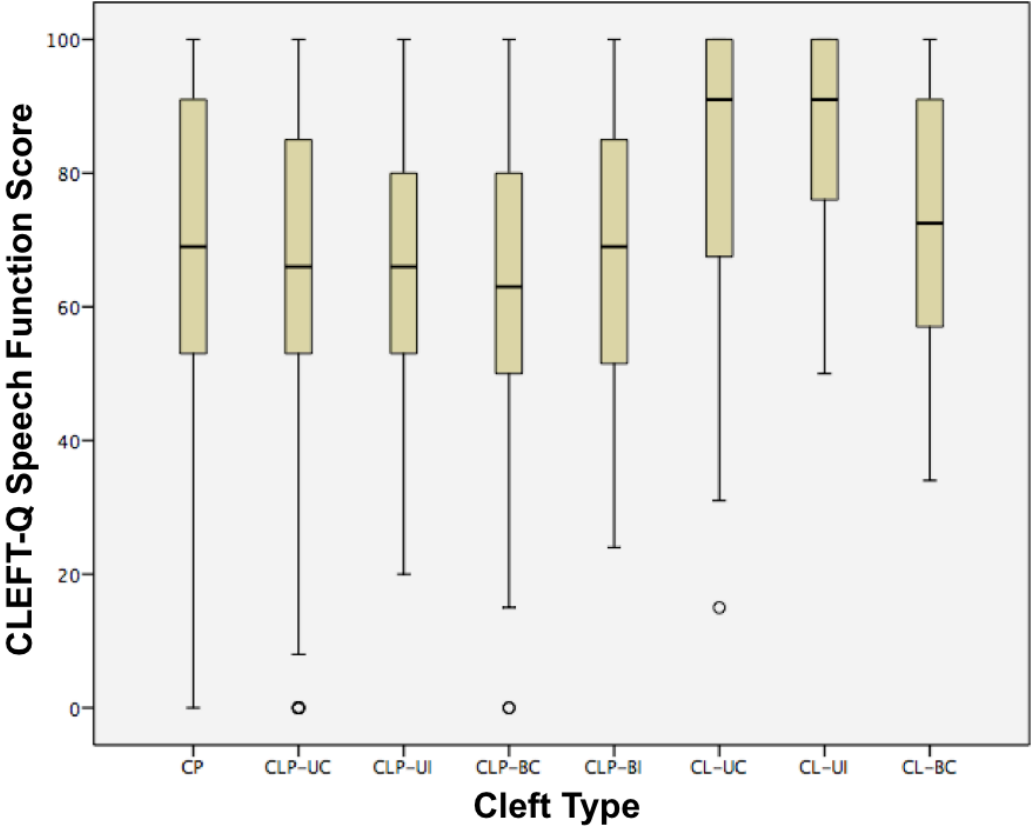


Figure 1i: Boxplot of CLEFT-Q Speech-Related Distress scores by cleft type.

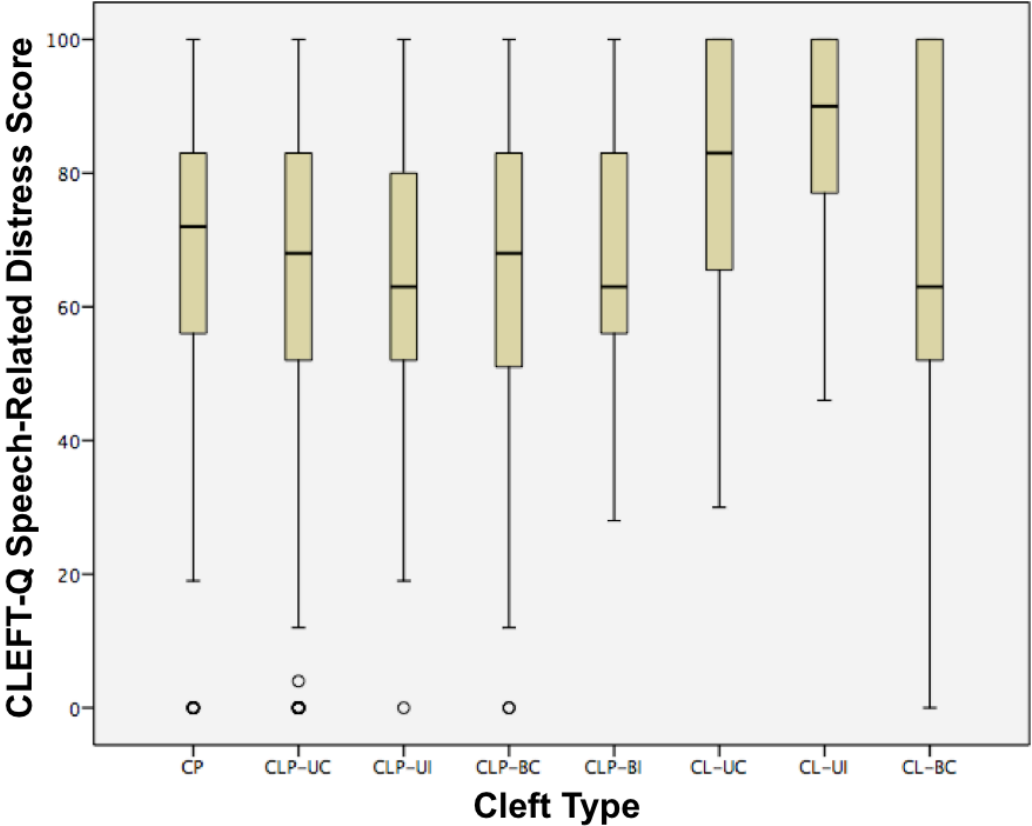


Figure 1j: Boxplot of CLEFT-Q Social scores by cleft type.

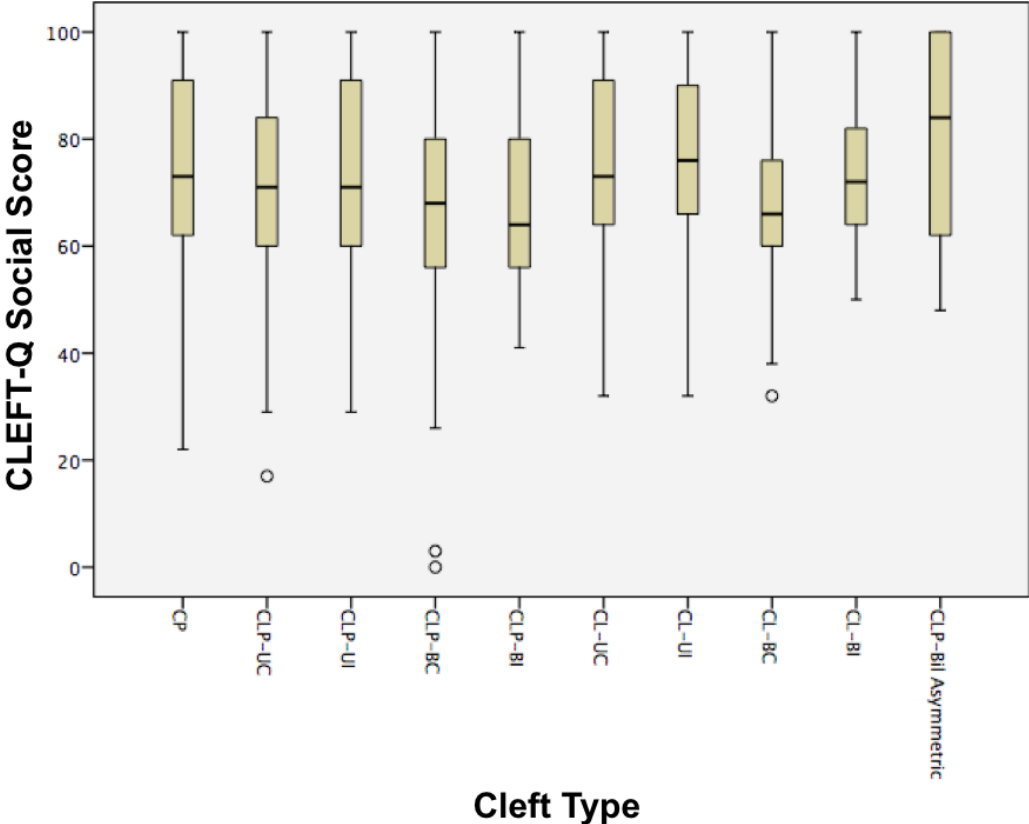


Figure 1k: Boxplot of CLEFT-Q School scores by cleft type.

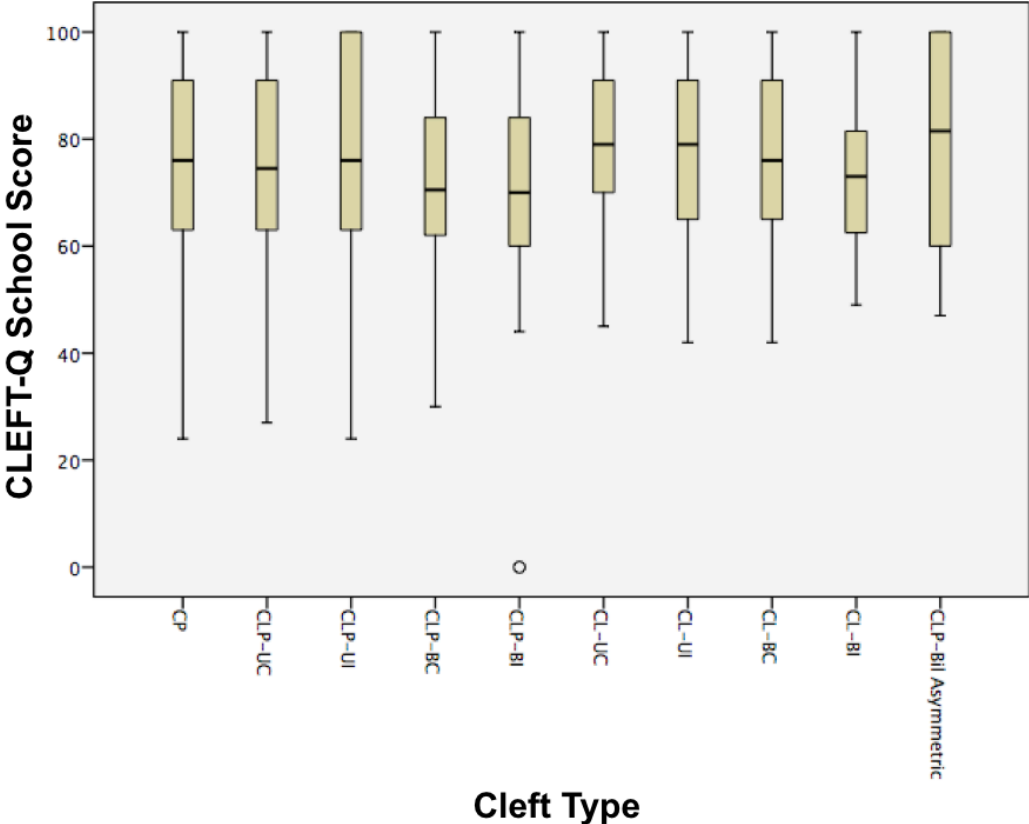


Figure 11: Boxplot of CLEFT-Q Psychological scores by cleft type.

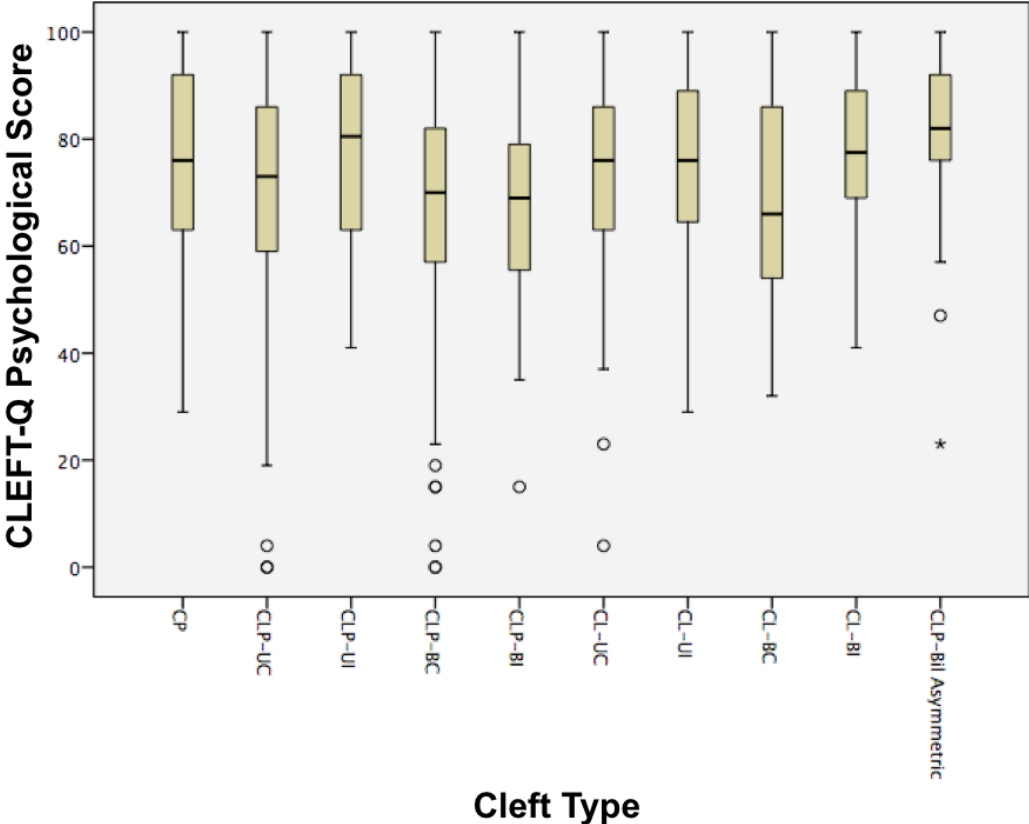


Table 4: Statistics for one-way ANOVA or Kruskal-Wallis H tests evaluating differences in CLEFT-Q scale scores between cleft types. When one-way ANOVA was used, results are shown as: F statistic (degrees of freedom between groups, degrees of freedom within groups). When the Kruskal-Wallis H test was used, results are shown as: H test statistic (degrees of freedom).

<i>Scale</i>	<i>Test Statistic</i>	<i>Sig.</i>
Face	F(9, 2306) = 17.953	0.000
Nose	F(9, 2214) = 31.806	0.000
Nostrils	F(9,2195) = 39.142	0.000
Lip	F(9, 2132) = 21.565	0.000
Cleft Lip Scar	F(8, 1640) = 3.959	0.000
Teeth	F(9, 2231) = 9.542	0.000
Jaws	F(8, 1401) = 6.508	0.000
Speech Function	F(7, 1795) = 8.830	0.000
Speech Function (Veau)	F(3, 1670) = 5.985	0.000
Speech-related Distress	F(7, 1816) = 7.106	0.000
Speech-related Distress (Veau)	F(3, 1691) = 7.832	0.000
Social	F(9, 2185) = 5.145	0.000
School	H(9) = 21.973	0.009
Psychological	H(9) = 42.697	0.000

Figure 2: Speech function and speech-related distress in patients with CP by Veau type. The bottom of the box represents the 25th percentile and the top of the box represents the 75th percentile. The whiskers represent 1.5 times the interquartile range. A circle represents an outlier, and an asterisk represents an extreme outlier.

Figure 2a: Boxplot of CLEFT-Q Speech Function scores by Veau type.

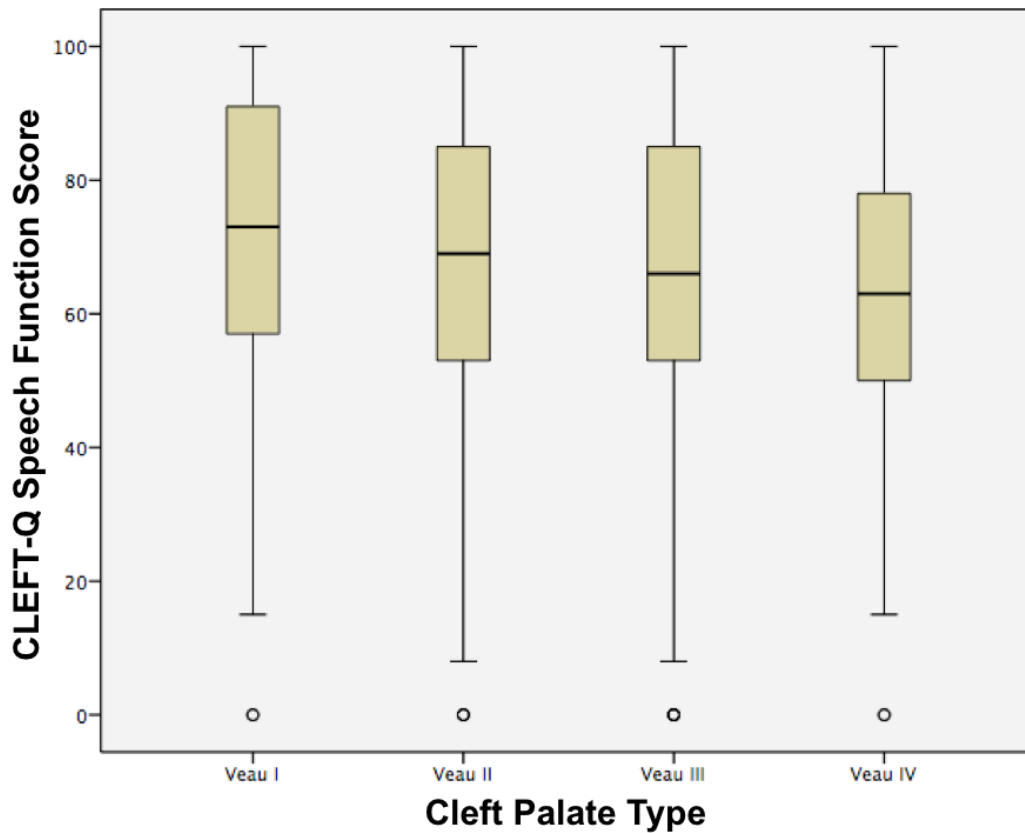


Figure 2b: Boxplot of CLEFT-Q Speech-Related Distress scores by Veau type.

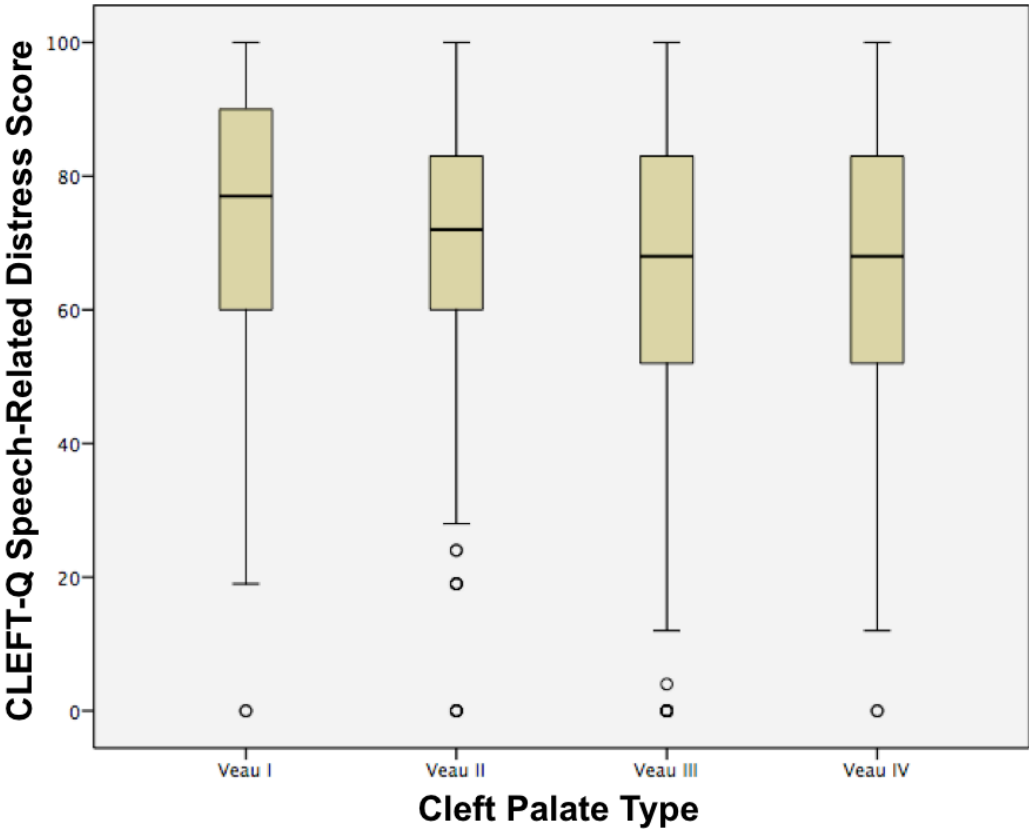


Figure 3: Scores for all CLEFT-Q scales by broad cleft types on radar charts.

Figure 2a shows scores for groups with all CLP subtypes; Figure 2b shows scores for patients with CP; and Figure 2c shows scores for groups with all CL subtypes. Figure 2d shows an example of CLEFT-Q scores for a 14 year old male with CLP-UC with the mean CLEFT-Q scores for patients with CLP-UC as a reference. The vertical axis or ‘spoke’ for each scale represents the CLEFT-Q score.

Figure 3a

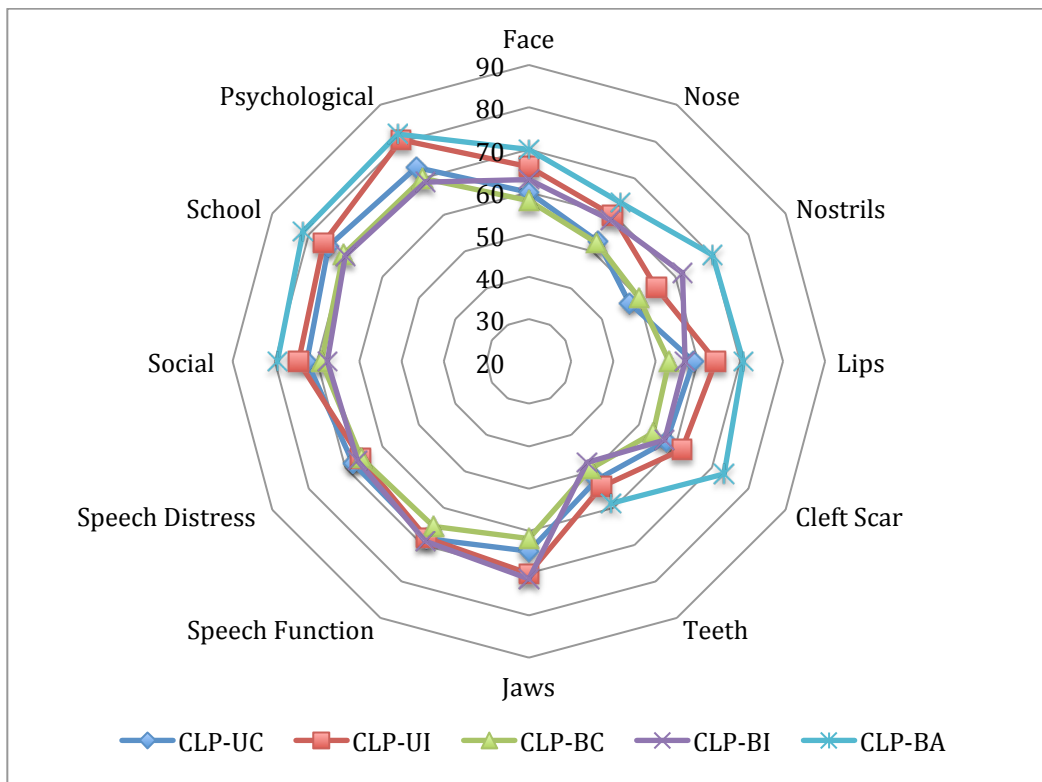


Figure 3b

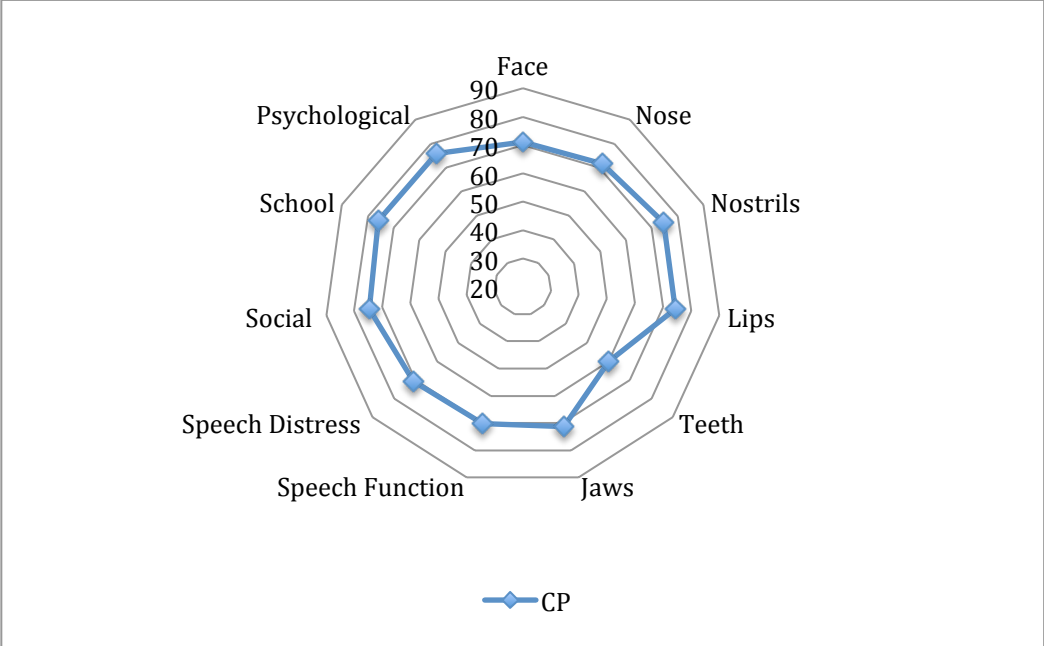


Figure 3c

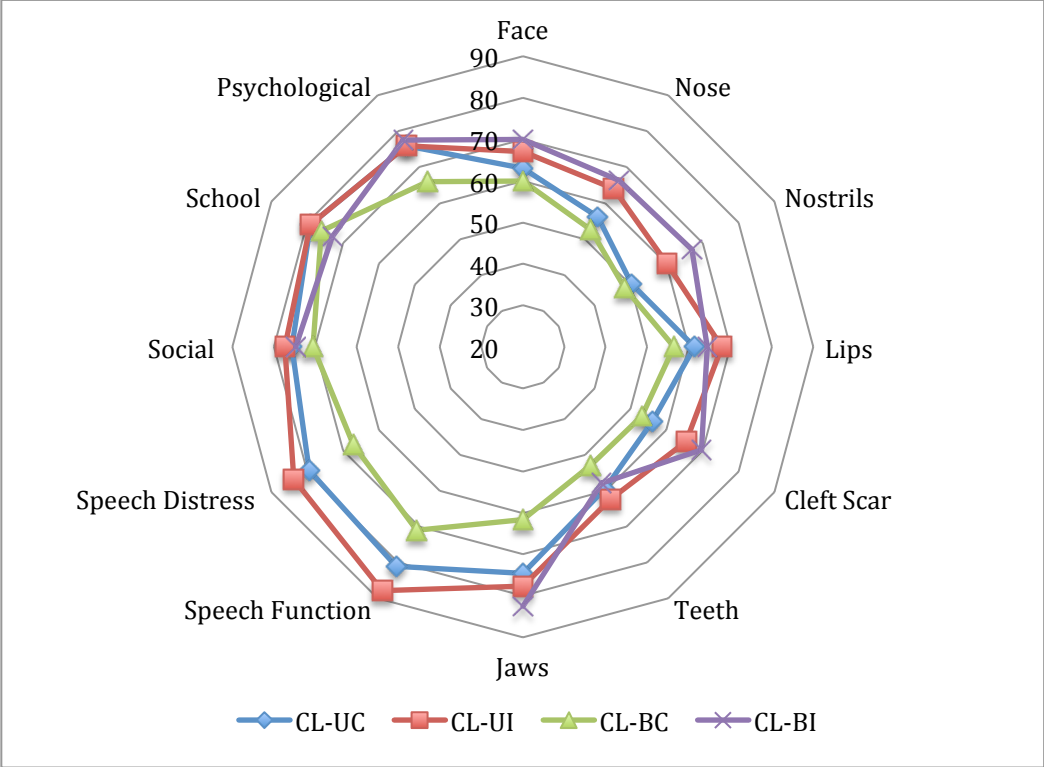
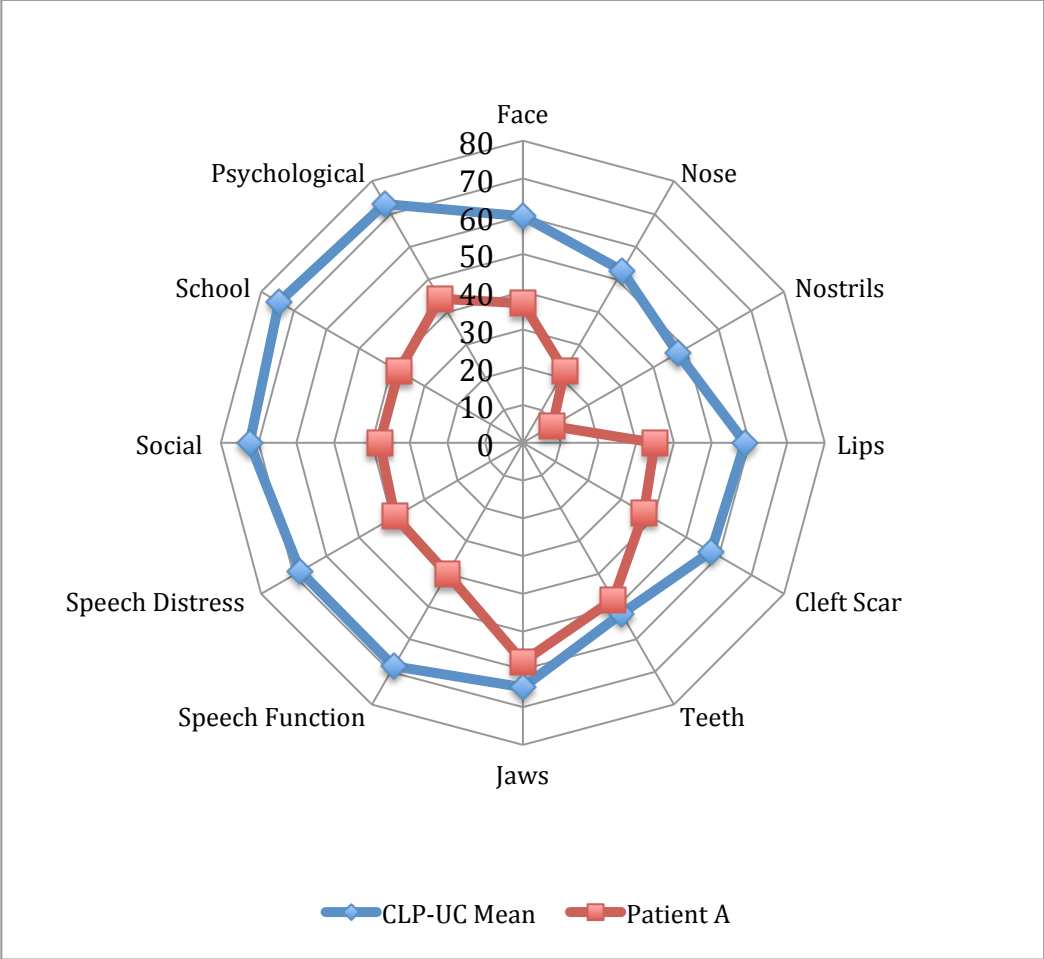


Figure 3d



References

1. Shaw, W. C., Brattstrom, V., Molsted, K., Pahl-Andersen, B., Roberts, C. T., Semb, G. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 5: discussion and conclusions. *Cleft Palate Craniofac J* 2005;42:93-98.
2. Shaye, D., Liu, C. C., Tollefson, T. T. Cleft Lip and Palate: An Evidence-Based Review. *Facial Plast Surg Clin North Am* 2015;23:357-372.
3. de Ladeira, P. R., Alonso, N. Protocols in cleft lip and palate treatment: systematic review. *Plast Surg Int* 2012;2012:562892.
4. Tsichlaki, A., O'Brien, K., Johal, A., Fleming, P. S. A scoping review of outcomes related to orthodontic treatment measured in cleft lip and palate. *Orthod Craniofac Res* 2017;20:55-64.
5. Shaye, D. Update on outcomes research for cleft lip and palate. *Curr Opin Otolaryngol Head Neck Surg* 2014;22:255-259.
6. Eckstein, D. A., Wu, R. L., Akinbiyi, T., Silver, L., Taub, P. J. Measuring quality of life in cleft lip and palate patients: currently available patient-reported outcomes measures. *Plast Reconstr Surg* 2011;128:518e-526e.
7. Klassen, A. F., Tsangaris, E., Forrest, C. R., et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. *J Plast Reconstr Aesthet Surg* 2012;65:547-557.

8. Broder, H. L., Wilson-Genderson, M., Sischo, L. Oral health-related quality of life in youth receiving cleft-related surgery: self-report and proxy ratings. *Qual Life Res* 2017;26:859-867.
9. McComb, R. W., Marrinan, E. M., Nuss, R. C., Labrie, R. A., Mulliken, J. B., Padwa, B. L. Predictors of velopharyngeal insufficiency after Le Fort I maxillary advancement in patients with cleft palate. *J Oral Maxillofac Surg* 2011;69:2226-2232.
10. Snyder, C. F., Aaronson, N. K., Choucair, A. K., et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res* 2012;21:1305-1314.
11. Wong Riff, K. W. Y., Tsangaris, E., Goodacre, T. E. E., et al. What matters to patients with cleft lip and/or palate: an international qualitative study informing the development of the CLEFT-Q. *Cleft Palate-Craniofacial Journal* 2017;in press.
12. Black, N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
13. OECD. Recommendations to OECD Ministers of Health from the high level reflection group on the future of health statistics: Strengthening the international comparison of health system performance through patient-reported indicators. Paris: Organisation for Economic Co-operation and Development, 2017.

14. Russell, K., Long, R. E., Jr., Hathaway, R., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 5. General discussion and conclusions. *Cleft Palate Craniofac J* 2011;48:265-270.
15. Bearn, D., Mildinhall, S., Murphy, T., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 4: outcome comparisons, training, and conclusions. *Cleft Palate Craniofac J* 2001;38:38-43.
16. Semb, G., Enemark, H., Friede, H., et al. A Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 1. Planning and management. *J Plast Surg Hand Surg* 2017;51:2-13.
17. Allori, A. C., Kelley, T., Meara, J. G., et al. A Standard Set of Outcome Measures for the Comprehensive Appraisal of Cleft Care. *Cleft Palate Craniofac J* 2017;54:540-554.
18. Glener, A. D., Allori, A. C., Shamma, R. L., et al. A Population-Based Exploration of the Social Implications Associated with Cleft Lip and/or Palate. *Plast Reconstr Surg Glob Open* 2017;5:e1373.
19. Kramer, F. J., Gruber, R., Fialka, F., Sinikovic, B., Hahn, W., Schliephake, H. Quality of life in school-age children with orofacial clefts and their families. *J Craniofac Surg* 2009;20:2061-2066.

20. Damiano, P. C., Tyler, M. C., Romitti, P. A., et al. Type of oral cleft and mothers' perceptions of care, health status, and outcomes for preadolescent children. *Cleft Palate Craniofac J* 2006;43:715-721.
21. Broder, H., Strauss, R. P. Self-concept of early primary school age children with visible or invisible defects. *Cleft Palate J* 1989;26:114-117; discussion 117-118.
22. Berger, Z. E., Dalton, L. J. Coping with a cleft: psychosocial adjustment of adolescents with a cleft lip and palate and their parents. *Cleft Palate Craniofac J* 2009;46:435-443.
23. Feragen, K. B., Borge, A. I., Rumsey, N. Social experience in 10-year-old children born with a cleft: exploring psychosocial resilience. *Cleft Palate Craniofac J* 2009;46:65-74.
24. Persson, M., Aniansson, G., Becker, M., Svensson, H. Self-concept and introversion in adolescents with cleft lip and palate. *Scand J Plast Reconstr Surg Hand Surg* 2002;36:24-27.
25. Kramer, F. J., Gruber, R., Fialka, F., Sinikovic, B., Schliephake, H. Quality of life and family functioning in children with nonsyndromic orofacial clefts at preschool ages. *J Craniofac Surg* 2008;19:580-587.
26. Lorot-Marchand, A., Guerreschi, P., Pellerin, P., et al. Frequency and socio-psychological impact of taunting in school-age patients with cleft lip-palate surgical repair. *Int J Pediatr Otorhinolaryngol* 2015;79:1041-1048.

27. Wong Riff, K. W., Tsangaris, E., Goodacre, T., et al. International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open* 2017;7:e015467.
28. Klassen, A., Wong Riff, K., Longmire, N., et al. Psychometric findings and normative values for the CLEFT-Q based on 2,434 children and young adult patients with cleft lip and/or palate from 12 countries. *Canadian Medical Association Journal* 2017;submitted.
29. Tsangaris, E., Wong Riff, K. W. Y., Goodacre, T., et al. Establishing Content Validity of the CLEFT-Q: A New Patient-reported Outcome Instrument for Cleft Lip/Palate. *Plast Reconstr Surg Glob Open* 2017;5:e1305.
30. Tsangaris, E., Wong Riff, K. W. Y., Vargas, F., et al. Translation and cultural adaptation of the CLEFT-Q for use in Colombia, Chile and Spain. *Health and Quality of Life Outcomes* 2017;submitted.
31. Tsangaris, E., Wong Riff, K. W. Y., Dreise, M., et al. Translation and cultural adaptation of the CLEFT-Q into Arabic, Dutch, Hindi, Swedish, and Turkish. *Plastic and Reconstructive Surgery Global Open* 2017;submitted.
32. Veau, V. *Division Palatine: Anatomie, Chirurgie, Phonétique*. Paris: Masson; 1931.
33. Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J. G. Research electronic data capture (REDCap)--a metadata-driven methodology and

workflow process for providing translational research informatics support. *J*

Biomed Inform 2009;42:377-381.

34. Klassen, A., Wong Riff, K., Longmire, N., et al. Psychometric findings and normative values for the CLEFT-Q based on 2,434 children and young adult patients with cleft lip and/or palate from 12 countries. *Canadian Medical Association Journal* 2017;in press.

35. George, D., Mallery, P. *SPSS for Windows step by step : a simple guide and reference, 17.0 update*, 10th ed. Boston: Allyn & Bacon; 2010.

36. Kaplan, R. S., Blackstone, R. P., Haas, D. A., Thaker, N. G. Measuring and communicating health care value with charts. *Harv Bus Rev*, 2015.

37. Sharma, V. P., Bella, H., Cadier, M. M., Pigott, R. W., Goodacre, T. E., Richard, B. M. Outcomes in facial aesthetics in cleft lip and palate surgery: A systematic review. *J Plast Reconstr Aesthet Surg* 2012;65:1233-1245.

38. Asher-McDade, C., Roberts, C., Shaw, W. C., Gallagher, C. Development of a method for rating nasolabial appearance in patients with clefts of the lip and palate. *Cleft Palate Craniofac J* 1991;28:385-390; discussion 390-381.

39. Long, R. E., Jr., Hathaway, R., Daskalogiannakis, J., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J* 2011;48:239-243.

40. Choa, R. M., Slator, R., Jeremy, A., et al. Identifying the effect of cleft type, deprivation and ethnicity on speech and dental outcomes in UK cleft patients: a multi-centred study. *J Plast Reconstr Aesthet Surg* 2014;67:1637-1643.
41. Mahoney, M. H., Swan, M. C., Fisher, D. M. Prospective analysis of presurgical risk factors for outcomes in primary palatoplasty. *Plast Reconstr Surg* 2013;132:165-171.
42. Sullivan, S. R., Marrinan, E. M., LaBrie, R. A., Rogers, G. F., Mulliken, J. B. Palatoplasty outcomes in nonsyndromic patients with cleft palate: a 29-year assessment of one surgeon's experience. *J Craniofac Surg* 2009;20 Suppl 1:612-616.
43. Browne, J. P., Cano, S. J., Smith, S. Using Patient-reported Outcome Measures to Improve Health Care: Time for a New Approach. *Med Care* 2017.

CHAPTER FIVE

Discussion and Conclusions

Discussion

On that first trip to India, patients and families were understandably curious about the group of foreigners that had descended upon their quiet village. Children would ask to play cricket with us, or to take photos with our digital cameras and then beam at the instant images that arose. One of the children had travelled for days to try and receive care. He had an unrepaired unilateral cleft lip, which made his smile look like a joyous toothy grin. He played alongside the other children, wanting to take photo after photo. Here was a boy who seemed to behave as if he was no different from the others. He stands out in my memory because the overall sense we got from the patients was one of urgency and sadness, and he seemed to carry less of a burden. Patients' perspectives are different for reasons that we cannot understand, and clinicians can only stand to benefit from an improved understanding of how patients experience their conditions.

The patient perspective has not been commonly included in evaluations of outcomes of cleft care. What initially began as a search for a way to describe the impact of treatment in low- and middle-income countries evolved into an understanding of the problem addressed in this thesis – namely the apparent inability to describe, from the patient's point of view, the impact of having a cleft and having it treated. The largest multi-centred studies of outcomes in cleft care to date either did not include PROs, or included PROs derived from ad hoc measures(1-35). Examining the literature, it appeared that this was due to three

factors: first, there were no cleft-specific PRO measures available; second, the science of measuring PROs was poorly understood by clinicians, leading to skepticism over the usefulness of PROs; and third, the clinical utility of PROs was not recognized. The three papers in this thesis sought to address these gaps in knowledge.

The first paper (Chapter Two) described an extensive qualitative study to identify the concepts important to patients with CL/P in high-income and low- and middle-income countries. We developed a conceptual framework of these concepts to guide the development of the CLEFT-Q. The qualitative study added significantly to a preliminary conceptual framework derived from a systematic review of concepts previously measured(36), further confirming that existing literature was missing measurements of concepts of interest to patients. The study identified appearance of the face, health-related quality of life including speech-related distress, psychological, social and school function, and facial function including speech and eating or drinking as the key concepts of interest. These findings are not surprising in themselves, but it is perhaps surprising that in the past they have not been evaluated from the patient perspective. The omission is not from a lack of recognition *per se*; narrative reviews have shown that several studies have addressed these concerns(37, 38), but critical review of these studies shows that these concepts were not measured in a rigorous way. The systematic review of concepts measured in the past that informed the preliminary conceptual framework excluded all studies that did not use validated PRO measures,

explaining the discrepancy between what has been *studied*(37, 38) and what has been *measured rigorously* in the past(36).

Importantly, none of the studies that met inclusion criteria in the systematic review were performed in low- and middle-income countries(36). We found that the concepts of interest identified in this study were similar between patients from high-income and low- and middle-income countries, meaning that a common metric could be designed for use in all countries.

The conceptual framework included thirteen concepts that were classified into the three top-level domains of appearance, health-related quality of life, and facial function. Each of these thirteen concepts was then turned into an independently functioning scale in the CLEFT-Q.

The second paper in the thesis (Chapter Three) describes the protocol for the entire development of the CLEFT-Q. The overall CLEFT-Q project has been a lengthy endeavor, notably because of our efforts to design the measure for international use. The purpose of this paper was to provide clinicians and other stakeholders with insight into the rigor required to design a scientifically sound, clinically meaningful PRO measure and an understanding of why this rigor is important. This paper discusses how the CLEFT-Q project adhered to recognized standards of measure development(39-45), and explicitly describes how the measure meets quality criteria. While clinicians may not need to understand the details of psychometric methods, it is important to recognize that the technique of measuring PROs has evolved over time, and that modern psychometric methods

provide more scientifically sound measurement. Since clinicians are often selecting measures for use both in clinical practice and in research studies, having a detailed description of how the CLEFT-Q was developed should be useful in future efforts to evaluate PROs in cleft care.

The third paper (Chapter Four) found that in the field test of 2,434 patients from twelve countries, CLEFT-Q outcomes varied by cleft type. Prior to this analysis, the field test data were used to perform the Rasch analysis, and all scales with the exception of the eating and drinking scale were found to fit the Rasch model(46). The eating and drinking scale was kept as a checklist rather than a scale. As part of the psychometric analysis, normative values were calculated by age, gender, and broad cleft type (cleft lip only, cleft lip and alveolus, cleft lip and palate, and cleft palate only). However, the clinical classification of cleft types is more detailed and thus the third paper provides a more clinically applicable analysis of how CLEFT-Q outcomes vary by cleft type.

Patients with the most severe cleft types, the complete unilateral and bilateral cleft lip and palate, scored lower on all CLEFT-Q appearance scales than patients with no visible cleft or a mild form of a cleft, the incomplete unilateral cleft lip. Patients with more severe clefts of the palate scored lower on the CLEFT-Q speech function and speech-related distress scales. On the CLEFT-Q psychological, social, and school function scales, patients with complete bilateral cleft lip and palate, the most severe presentation, scored the lowest. These findings serve several purposes: first, they show that the CLEFT-Q is able to

detect differences in outcomes between patients with different cleft types, and second, they show how PROs provide clinically useful information. The presentation of multiple outcomes using radar charts also gives clinicians an idea of the clinical utility of the results, since visualizing the outcomes simultaneously in a meaningful way is informative.

The CLEFT-Q scales have been incorporated by the International Consortium of Health Outcomes Measurement, or ICHOM, into an international effort to standardize measurement in cleft care(47). One of the aims of ICHOM is to facilitate benchmarking of outcomes such that providers and centres can be compared to each other. While establishing normative values is important to allow for these comparisons, the variability in presentation of CL/P may result in inaccurate comparisons if cleft types are not taken into consideration. Rigorous measurement is of great importance in this situation as the ability to detect differences is key in accounting for clinical variation.

The overall objective of the thesis was to show that through adherence to rigorous methods of development, a PRO measure could provide clinically meaningful outcome evaluation in cleft care. This can then facilitate studies of patterns of care, evaluations of quality of outcomes, and differentiation of outcomes by cleft type in the future. The three papers address this objective in different ways: the first showed that patient input provided a comprehensive framework of concepts that should be measured in patients with CL/P; the second served to inform clinicians about the science behind rigorous measurement by

describing how the CLEFT-Q was developed; and the third showed that variation in outcomes for different cleft types was detected using the CLEFT-Q. Ultimately, the work presented in the thesis should help in future efforts to incorporate the patient perspective in outcome evaluation.

One of the main reasons why the CLEFT-Q was developed was to aid in describing the status of patients with CL/P in low- and middle-income countries. Creating a ruler or a common metric that is validated in patients from countries of different levels of income is the first step towards describing the impact of having a cleft and the impact of its treatment for all patients. The hope is that eventually, what my surgical colleagues and I witnessed firsthand in India will be able to be measured in order to determine how best to provide care. A common metric will highlight disparities in care and will help in advocacy efforts for care in a condition that does not lend itself well to standard objective measurements such as infection rates or mortality.

Strengths and Limitations

One of the strengths of the studies presented is the scope of patients included in the development of the CLEFT-Q. Of the multi-centred studies of outcomes in cleft care to date, none has included as many patients from as many different countries as the CLEFT-Q project. While this was not meant to be a study of outcomes, the network of collaborators is now well established and may lead to more multi-centred studies in the future.

Collecting data for the qualitative component of the study from both high-income and low- and middle-income countries allowed for the incorporation of patient input from these countries at the outset of developing the CLEFT-Q. This strategy maximized the chance that the final measure would function well across countries.

Another significant strength of the study was the methodology behind the development of the CLEFT-Q. Using modern psychometrics allow for the most precise measurement possible, meaning that the rulers of the CLEFT-Q are well designed to capture relevant outcomes in cleft care.

A limitation of the work presented is that in the case of the qualitative study and the field-test of the CLEFT-Q, participants were recruited to provide a convenience sample. This may have skewed the population towards those seeking treatment, since a large proportion of participants were recruited in the clinical setting. In high-income countries, this may not be as much of a concern since the CLEFT-Q will be used primarily as an outcome measure to evaluate clinical care, and virtually all people with clefts receive care. However, in low- and middle-income countries, not including patients with CL/P who were unable to access clinical care may present a greater limitation. Theoretically, this would manifest in floor effects on the scales if we presume that patients who are unable to access care would score the lowest. Future studies in low- and middle-income countries should determine whether this is indeed the case.

Another limitation of the study is that the clinical meaning of the CLEFT-Q scores is yet to be determined. As the CLEFT-Q project continues, further evaluation will include calculations of minimally important differences on the scales.

Knowledge Translation

An integrated knowledge translation approach was taken in the work presented in the thesis. Clinicians were involved in the study from the outset and regular communication with our clinical partners from participating sites helped to disseminate information about the study and why rigorous methods were important. We regularly attended scientific meetings to present the study in order to maintain contact with the cleft care community. The second paper in the thesis was published after we identified a need for further understanding of PRO measure development as the study progressed. The incorporation of the CLEFT-Q scales into the ICHOM standard set for CL/P was also an important step in knowledge translation. This set is now being used in several centres around the world.

Moving forward, the next step is to begin using the CLEFT-Q in routine clinical practice. In this digital age, creating a mobile application or a website to administer the CLEFT-Q would perhaps ease the burden of administration, allowing patients to fill out the scales at home before a clinical visit. Future studies will also need to assess whether or not measuring PROs in fact leads to improved outcomes. Continuing multi-centred collaboration should help to collect

data faster, given the relatively low numbers of patients seen in any one given centre. This collaboration will also serve to facilitate quality improvement initiatives in the future that incorporate PROs.

Conclusions

PROs can be evaluated in patients with CL/P in a clinically meaningful way using the CLEFT-Q, a PRO measure designed using rigorous methodology for establishing scientifically sound measurement. This study represents the first step in changing the way that outcomes of cleft care are measured to include the patient perspective. By providing the rulers to measure PROs, the stage is set for a wave of studies, clinical audits, and quality improvement initiatives that can provide an accurate reflection of patient status. Perhaps most importantly, PROs can be incorporated into clinical practice so that individual patients receiving cleft care can potentially benefit from their measurement.

References

1. Shaw, W. C., Asher-McDade, C., Brattström, V., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 1. Principles and study design. *Cleft Palate Craniofac J* 1992;29:393-397.
2. Mølsted, K., Asher-McDade, C., Brattström, V., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 2. Craniofacial form and soft tissue profile. *Cleft Palate Craniofac J* 1992;29:398-404.
3. Mars, M., Asher-McDade, C., Brattström, V., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 3. Dental arch relationships. *Cleft Palate Craniofac J* 1992;29:405-408.
4. Asher-McDade, C., Brattström, V., Dahl, E., et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 4. Assessment of nasolabial appearance. *Cleft Palate Craniofac J* 1992;29:409-412.
5. Shaw, W. C., Semb, G., Nelson, P., et al. The Eurocleft project 1996-2000: overview. *J Craniomaxillofac Surg* 2001;29:131-140; discussion 141-132.
6. Semb, G., Brattstrom, V., Molsted, K., Prah-Andersen, B., Shaw, W. C. The Eurocleft study: intercenter study of treatment outcome in patients with

complete cleft lip and palate. Part 1: introduction and treatment experience. *Cleft Palate Craniofac J* 2005;42:64-68.

7. Brattstrom, V., Molsted, K., Prah-Andersen, B., Semb, G., Shaw, W. C. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 2: craniofacial form and nasolabial appearance. *Cleft Palate Craniofac J* 2005;42:69-77.

8. Molsted, K., Brattstrom, V., Prah-Andersen, B., Shaw, W. C., Semb, G. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 3: dental arch relationships. *Cleft Palate Craniofac J* 2005;42:78-82.

9. Semb, G., Brattstrom, V., Molsted, K., et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 4: relationship among treatment outcome, patient/parent satisfaction, and the burden of care. *Cleft Palate Craniofac J* 2005;42:83-92.

10. Shaw, W. C., Brattstrom, V., Molsted, K., Prah-Andersen, B., Roberts, C. T., Semb, G. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 5: discussion and conclusions. *Cleft Palate Craniofac J* 2005;42:93-98.

11. Sandy, J. R., Williams, A. C., Bearn, D., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 1: background and methodology. *Cleft Palate Craniofac J* 2001;38:20-23.

12. Williams, A. C., Bearn, D., Mildinhal, S., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 2: dentofacial outcomes and patient satisfaction. *Cleft Palate Craniofac J* 2001;38:24-29.
13. Sell, D., Grunwell, P., Mildinhal, S., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 3: speech outcomes. *Cleft Palate Craniofac J* 2001;38:30-37.
14. Bearn, D., Mildinhal, S., Murphy, T., et al. Cleft lip and palate care in the United Kingdom--the Clinical Standards Advisory Group (CSAG) Study. Part 4: outcome comparisons, training, and conclusions. *Cleft Palate Craniofac J* 2001;38:38-43.
15. Long, R. E., Jr., Hathaway, R., Daskalogiannakis, J., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J* 2011;48:239-243.
16. Hathaway, R., Daskalogiannakis, J., Mercado, A., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 2. Dental arch relationships. *Cleft Palate Craniofac J* 2011;48:244-251.
17. Daskalogiannakis, J., Mercado, A., Russell, K., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft

lip and palate part 3. Analysis of craniofacial form. *Cleft Palate Craniofac J* 2011;48:252-258.

18. Mercado, A., Russell, K., Hathaway, R., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 4. Nasolabial aesthetics. *Cleft Palate Craniofac J* 2011;48:259-264.

19. Russell, K., Long, R. E., Jr., Hathaway, R., et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 5. General discussion and conclusions. *Cleft Palate Craniofac J* 2011;48:265-270.

20. Persson, M., Sandy, J. R., Waylen, A., et al. A cross-sectional survey of 5-year-old children with non-syndromic unilateral cleft lip and palate: the Cleft Care UK study. Part 1: background and methodology. *Orthod Craniofac Res* 2015;18 Suppl 2:1-13.

21. Al-Ghatam, R., Jones, T. E., Ireland, A. J., et al. Structural outcomes in the Cleft Care UK study. Part 2: dento-facial outcomes. *Orthod Craniofac Res* 2015;18 Suppl 2:14-24.

22. Smallridge, J., Hall, A. J., Chorbachi, R., et al. Functional outcomes in the Cleft Care UK study--Part 3: oral health and audiology. *Orthod Craniofac Res* 2015;18 Suppl 2:25-35.

23. Sell, D., Mildinhall, S., Albery, L., Wills, A. K., Sandy, J. R., Ness, A. R. The Cleft Care UK study. Part 4: perceptual speech outcomes. *Orthod Craniofac Res* 2015;18 Suppl 2:36-46.

24. Waylen, A., Ness, A. R., Wills, A. K., Persson, M., Rumsey, N., Sandy, J. R. Cleft Care UK study. Part 5: child psychosocial outcomes and satisfaction with cleft services. *Orthod Craniofac Res* 2015;18 Suppl 2:47-55.
25. Semb, G., Enemark, H., Friede, H., et al. A Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 1. Planning and management. *J Plast Surg Hand Surg* 2017;51:2-13.
26. Rautio, J., Andersen, M., Bolund, S., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 2. Surgical results. *J Plast Surg Hand Surg* 2017;51:14-20.
27. Bannister, P., Lindberg, N., Jeppesen, K., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 3. Descriptive study of postoperative nursing care following first stage cleft closure. *J Plast Surg Hand Surg* 2017;51:21-26.
28. Lohmander, A., Persson, C., Willadsen, E., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 4. Speech outcomes in 5-year-olds - velopharyngeal competency and hypernasality. *J Plast Surg Hand Surg* 2017;51:27-37.
29. Willadsen, E., Lohmander, A., Persson, C., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 5. Speech outcomes in 5-year-olds - consonant proficiency and errors. *J Plast Surg Hand Surg* 2017;51:38-51.

30. Heliövaara, A., Küsseler, A., Skaare, P., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 6. Dental arch relationships in 5 year-olds. *J Plast Surg Hand Surg* 2017;51:52-57.
31. Karsten, A., Marcusson, A., Hurmerinta, K., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 7. Occlusion in 5 year-olds according to the Huddart and Bodenham index. *J Plast Surg Hand Surg* 2017;51:58-63.
32. Mølsted, K., Humerinta, K., Küsseler, A., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 8. Assessing naso-labial appearance in 5-year-olds - a preliminary study. *J Plast Surg Hand Surg* 2017;51:64-72.
33. Feragen, K. B., Rumsey, N., Heliövaara, A., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and Palate: 9. Parental report of social and emotional experiences related to their 5-year-old child's cleft diagnosis. *J Plast Surg Hand Surg* 2017;51:73-80.
34. Feragen, K. B., Semb, G., Heliövaara, A., et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 10. Parental perceptions of appearance and treatment outcomes in their 5-year-old child. *J Plast Surg Hand Surg* 2017;51:81-87.
35. Shaw, W., Semb, G. The Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 11. What next? *J Plast Surg Hand Surg* 2017;51:88-93.

36. Klassen, A. F., Tsangaris, E., Forrest, C. R., et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. *J Plast Reconstr Aesthet Surg* 2012;65:547-557.
37. Hunt, O., Burden, D., Hepper, P., Johnston, C. The psychosocial effects of cleft lip and palate: a systematic review. *Eur J Orthod* 2005;27:274-285.
38. Stock, N. M., Feragen, K. B. Psychological adjustment to cleft lip and/or palate: A narrative review of the literature. *Psychol Health* 2016;31:777-813.
39. Administration, U. F. a. D. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. In F. a. D. A. Department of Health and Human Services, Center for Drug Evaluation and Research ed. Rockville, MD, USA, 2009.
40. Aaronson, N., Alonso, J., Burnam, A., et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193-205.
41. Patrick, D. L., Burke, L. B., Gwaltney, C. J., et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument. *Value Health* 2011;14:967-977.
42. Patrick, D. L., Burke, L. B., Gwaltney, C. J., et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good

Research Practices Task Force report: part 2--assessing respondent understanding.

Value Health 2011;14:978-988.

43. Reeve, B. B., Wyrwich, K. W., Wu, A. W., et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889-1905.

44. Terwee, C. B., Bot, S. D., de Boer, M. R., et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.

45. Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-745.

46. Klassen, A., Wong Riff, K., Longmire, N., et al. Psychometric findings and normative values for the CLEFT-Q based on 2,434 children and young adult patients with cleft lip and/or palate from 12 countries. *Canadian Medical Association Journal* 2017;in press.

47. Allori, A. C., Kelley, T., Meara, J. G., et al. A Standard Set of Outcome Measures for the Comprehensive Appraisal of Cleft Care. *Cleft Palate Craniofac J* 2017;54:540-554.