IMPACT OF COMPLETING FACE-Q CRANIOFACIAL MODULE SCALES

IMPACT OF COMPLETING FACE-Q CRANIOFACIAL MODULE SCALES ON CHILDREN AND YOUNG ADULTS WITH FACIAL DIFFERENCES: AN INTERNATIONAL STUDY

By LUCAS GALLO, MD

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements of the Degree of Master of Science

McMaster University MASTER OF SCIENCE (2022)

Hamilton, Ontario (Health Research Methodology)

TITLE: Impact of Completing FACE-Q Craniofacial Module Scales on Children and Young Adults with Facial Differences: An International Study

AUTHOR: Lucas Gallo MD (McMaster University)

SUPERVISOR: Dr. Anne Klassen

NUMBER OF PAGES: xii, 20

LAY ABSTRACT

The FACE-Q Craniofacial module measures outcomes that matter to children and young adults with diverse facial conditions. To date, it remains unclear whether asking detailed questions about facial appearance and function can negatively impact patients, particularly children. The purpose of this study was to investigate the impact of completing the FACE-Q Craniofacial module and to identify factors associated with a negative impact in children and young adults. Participants aged 8-29, who completed at least one scale of the FACE-Q Craniofacial module as part of the international field-test study between December 2016-2019, were asked three impact questions following scale completion. Most patients responded neutrally to all questions with negative responses representing only a small proportion of patients (<13%). Increased severity of the facial condition, more scales completed, and lower FACE-Q scale scores were associated with a negative impact. Ultimately, this study demonstrates the FACE-Q Craniofacial module is acceptable for most participants.

ABSTRACT

BACKGROUND: The FACE-Q Craniofacial module measures outcomes that matter to patients with diverse craniofacial conditions. However, it is not known whether completing a patient reported outcome measure (PROM) has a negative impact on patients, particularly children. This study aims to investigate the impact of completing the FACE-Q Craniofacial module and identify factors associated with a negative impact.

METHODS: Participants were aged 8-29 years, with a facial difference, who completed at least one module of the FACE-Q Craniofacial module as part of the international field-test study between December 2016-2019. Participants were asked three questions: 'Did you like or dislike answering this questionnaire?'; 'Did answering these questions change how you feel about how you look?'; and 'Did answering this questionnaire make you feel unhappy or happy?' Univariate and multivariable logistic regression analyses were used to evaluate variables associated with a negative response.

RESULTS: The sample included 927 participants. Most patients responded neutrally to all impact questions: 42.7% neither disliked nor liked the questionnaire; 76.6% felt the same about how they looked; and 72.7% felt neither unhappy/happy after completion. Negative responses represented a small proportion of patients across all three impact questions (<13.2%). Increased craniofacial severity, more scales completed, and lower scores on all FACE-Q

scales were associated with negative responses for all three impact questions ($p<0.01$).

CONCLUSION: This study provides evidence that the FACE-Q Craniofacial module is acceptable for most participants. Clinicians and study investigators should follow up with patients after completing this PROM to address areas of concern in scale scores.

ACKNOWLEDGEMENTS

I would like to thank my primary supervisor, Dr. Anne Klassen, and the thesis committee: Dr. Sophocles Voineskos and Dr. Achilleas Thoma, for their support throughout the program.

I am grateful to the Q-portfolio team at McMaster University whose help was instrumental to the completion of this study, including Charlene Rae, Rakhshan Kamran, Shelby Deibert and Dr. Karen W.Y. Wong Riff.

Finally, I would like to thank the staff and resident physicians of the Division of Plastic Surgery at McMaster University for their ongoing mentorship and support throughout the completion of this Master of Science program.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PROM: Patient reported outcome measure

IRB: Institutional review board

REDcap: Research Electronic Data Capture

SE: Standard error

OR: Odds ratio

MID: Minimal important difference

DECLARATION OF ACADEMIC ACHIEVEMENT

Lucas Gallo is primarily responsible for the development of the research question, study design, data analysis, and manuscript preparation.

CHAPTER 1: INTRODUCTION

1.0 Background

Visible facial differences, secondary to a congenital or acquired disease process, can significantly impact an individual's quality of life.[1,2] These conditions are frequently associated with complex interventions that aim to restore both form and function to patients, and can dramatically alter facial appearance. Outcome measures designed for this patient population contain few items that address facial appearance or function and do not capture all pediatric ages.[1] To address these limitations, our team developed a novel patient reported outcome measure (PROM), known as the FACE-Q Craniofacial module, to measure outcomes that matter to patients with facial differences.[1] The 27 independently functioning scales/checklists that compose the FACE-Q Craniofacial module were developed and validated in a sample of children and young adults aged 8 - 29 years old.[1,3-5]

While PROMs are increasingly used within clinical and academic settings to promote shared patient decision-making, to evaluate quality of care, and to measure outcomes in clinical effectiveness trials, few studies have evaluated the impact of PROM completion on the patient.[6,7] Specifically, it remains unclear whether asking detailed questions about facial appearance, function, and psychosocial function can negatively impact patients. Investigating these effects are especially important in vulnerable populations, for example pediatric patients. Of note, Klassen et al.[7] evaluated the impact of completing the CLEFT-Q on a sample of 2056 pediatric and young adult patients with cleft lip and/or palate. Most participants (88%) liked answering the CLEFT-Q, with 67% and 23% reporting that they felt 'the same' or 'better' about how they look after completing the appearance scales, respectively.[7] It is unclear whether the results of this study can be applied to a broader patient population with diverse facial conditions.

Institutional review boards (IRBs) and regulators of federally sponsored research in the United States often request information on the potential harm when asking individuals to participate in surveys and interviews that may involve emotionally distressing topics.[8] While some survey research respondents report negative emotional reactions, a large meta-analysis demonstrated benefits to psychological health and overall functioning when disclosing information, thoughts, and feelings about personal and meaningful topics.[9]

1.1 Study Objectives

Additional evidence regarding the impact of asking detailed questions about facial differences is needed to guide the use of the FACE-Q Craniofacial module. As such, the primary aim of this study was to describe the impact of completing the FACE-Q Craniofacial module in a sample of pediatric and young adult patients with diverse facial conditions. The secondary aim of this study was to perform an exploratory analysis to identify demographic and clinical factors associated with reporting a negative impact following PROM completion.

CHAPTER 2: STUDY DESIGN AND METHODS

2.0 Study Design

This cross-sectional study was approved by the research ethics board at McMaster University and by the ethics board at each participating site. Consent was obtained prior to study enrollment.

2.1 Participant Inclusion and Exclusion Criteria

A convenience sample of pediatric and young adult patients was recruited from craniofacial centers as part of the larger international field-test study to validate the FACE-Q Craniofacial module. English language, as well as translated versions of the FACE-Q scales, were administered in the clinical setting. Patients were included if: 1) they were 8 – 29 years old; 2) had any condition, congenital or acquired, that results in a visible facial appearance and/or functional difference, 3) completed one or more of the FACE-Q Craniofacial module scales. Exclusion criteria included: isolated orthodontic (i.e., dental) conditions, failure to complete the impact questions, or inability to self-report.

2.2 Study Procedure

Data was collected between December 2016 and 2019, using either electronic tablets or paper booklets. A trained research assistant added patient clinical data regarding the type of facial condition, the severity (none, minor or major) of the facial difference by facial part (i.e., birthmark, cheeks, chin, ears, eyes/eyelids, forehead/eyebrows, head shape, jaws, lips, nose, nostrils, smile, teeth), and treatment status (i.e., whether additional surgical or non-surgical treatment(s) were anticipated). Demographic data, core FACE-Q Craniofacial module scales (Appearance Distress, Psychological, Social, and Facial Appearance), as well as additional scales relevant to the participant's diagnosis, were completed by participants in the clinical setting. At

the end of the survey, participants answered three questions designed by the research team to evaluate the impact of completing the FACE-Q (**Figure 1**). Each question consisted of five ordinal response options. At the end of the questionnaire, participants were invited to provide free-text comments. Data were entered into Research Electronic Data Capture (REDcap)[10] hosted at McMaster University (Canada).

Figure 1. *Participant impact questions*

1. ***Did you like or dislike answering this questionnaire?***
   Response options included: 'I disliked it a lot', 'I disliked it a little', 'I neither disliked or liked it', 'I liked it a little', and 'I liked it a lot'.
2. ***Did answering this questionnaire change how you feel about how you look (i.e., your appearance)?***
   Response options included: 'I feel a lot worse about how I look', 'I feel a little worse about how I look', 'I feel the same about how I look', 'I feel a little better about how I look', and 'I feel a lot better about how I look'.
3. ***Did answering this questionnaire make you feel unhappy or happy?***
   Response options included: 'It made me feel very unhappy', 'It made me feel a little unhappy', 'I feel the same as usual', 'It made me feel a little happy', and 'It made me feel very happy'.

2.3 Data Analysis

Descriptive statistics were used to evaluate demographic and clinical characteristics as well as frequencies for each impact question. To evaluate variables associated with a negative participant impact, a univariate logistic regression analysis was performed by converting the five ordinal response options into a dichotomous dependent outcome corresponding to a negative or a neutral/positive impact for each question. Independent variables were selected based on their hypothesized association with a negative participant impact, including: age (continuous, years at time of PROM completion), gender (male/female), extent of facial involvement (unilateral/bilateral), facial surgery within the last six months (yes/no), additional surgical or non-surgical treatment(s) anticipated (yes/no), composite score of facial severity (continuous, 0

to 26, where *higher* scores reflect *increased* severity), number of FACE-Q scales completed (continuous, 1-27 number of scales/checklists) as well as the final scores on core scales completed by participants (i.e., Appearance Distress, Psychological, Social, and Facial Appearance scales). All scales were scored from 0 – 100, where *higher* scores correspond to *better* outcomes. A Wald test or General Likelihood Ratio test was used to evaluate the statistical significance of each independent variable ($p<0.05$).

Statistically significant variables from a univariate logistic regression analysis were included in a multivariable logistic regression model. Backward stepwise regression analysis was performed to identify significant variables, adjusting for other variables within the model. A threshold of $p=0.10$ was used for variable inclusion. Goodness-of-fit of the multivariable models were evaluated using the Hosmer-Lemeshow test, where $p\geq0.05$ implies the model is appropriate. Given a sample size rule of thumb of 10 events (negative impact) per independent variable, the participant sample was powered to address the first impact question (i.e., "Did you like or dislike answering this questionnaire?").[11,12]

Multicollinearity was evaluated using the standard errors (SE) of independent variables within the logistic regression analysis, where a SE >2 was used to denote multicollinear independent variables. Listwise deletion (i.e., complete case analysis) was performed to address missing data within regression analyses. Statistical significance was considered $p<0.05$. All analyses were performed using SPSS ® version 26.0 (IBM Corporation, Armonk NY, USA for Windows ®).

CHAPTER 3: STUDY RESULTS

3.0 Participant Characteristics

The study sample consisted of n=927 participants. Most patients identified as female (n=514, 55.4%), were 8 – 17 years old (n=681, 73.5%), and from Canada or the United States (n=636, 68.6%). Participants completed an average of 10 FACE-Q scales. Demographic characteristics and scale summary statistics are presented in **Table 1** and **Table 2**, respectively.

Table 1. *Demographic characteristics (n=927)*

Participant Characteristics		***N***	***N %***
Gender	Male	409	44.1%
	Female	514	55.4%
	Other	3	0.3%
	Missing	1	0.1%
Age group, years	8-10	218	23.5%
	11-13	207	22.3%
	14-17	256	27.6%
	18-29	246	26.5%
Country	Australia	31	3.3%
	Canada	522	56.3%
	Chile	7	0.8%
	Ireland	113	12.2%
	USA	114	12.3%
	United Kingdom	139	15.0%
	Nepal	1	0.1%
Facial condition	Soft tissue	96	8.4%
	Skeletal	480	38.3%
	Facial Paralysis	42	3.4%
	Congenital skin lesion	162	13.0%
	Trauma	74	5.9%
	Ear	73	5.8%
Extent of facial involvement	Unilateral	424	45.7%
	Bilateral	378	40.8%
	Missing	125	13.5%

Facial surgery within last 6 months	Yes	143	15.4%
	No	774	83.5%
	Missing	10	1.1%
Treatment status	Additional treatment anticipated	274	29.6%
	Prior treatment, no further treatment anticipated	479	51.7%
	No treatment, no further treatment anticipated	174	18.8%

Table 2. *Distribution of participant variables (n=927)*

Participant Variable	N	Mean (SD)	Range
Age, years	927	14.6 (4.7)	8 – 29
Craniofacial severity score*	927	4.7 (4.4)	0 – 24
Number of FACE-Q Subscales completed	927	9.6 (4.1)	3 – 23
Facial appearance scale score**	892	61.3 (21.0)	0 – 100
Appearance distress scale score**	925	74.5 (20.3)	0 – 100
Psychological scale score**	923	75.2 (20.9)	0 – 100
Social scale score**	923	75.0 (18.1)	17 – 100

N, denotes frequency of individual participants; *, scored from 0-24, where *higher* scores reflect *increased* severity; **, subscales were scored from 0 – 100, where *higher* scores correspond to *better* outcomes.

3.1 Regression Analysis

1. Did you like or dislike answering this questionnaire?

Most participants reported a neutral (n=396, 42.7%) or positive impact (n=409, 44.1%) associated with completing the questionnaire (**Figure 2)**. Of the participants who described a negative impact (n=122, 13.2%), n=89 (9.6%) and n=33 (3.6%) participants reported that they disliked the questionnaire ‘a little’ and ‘a lot’, respectively (**Table 3**).

Figure 2. *Impact of FACE-Q Craniofacial module*

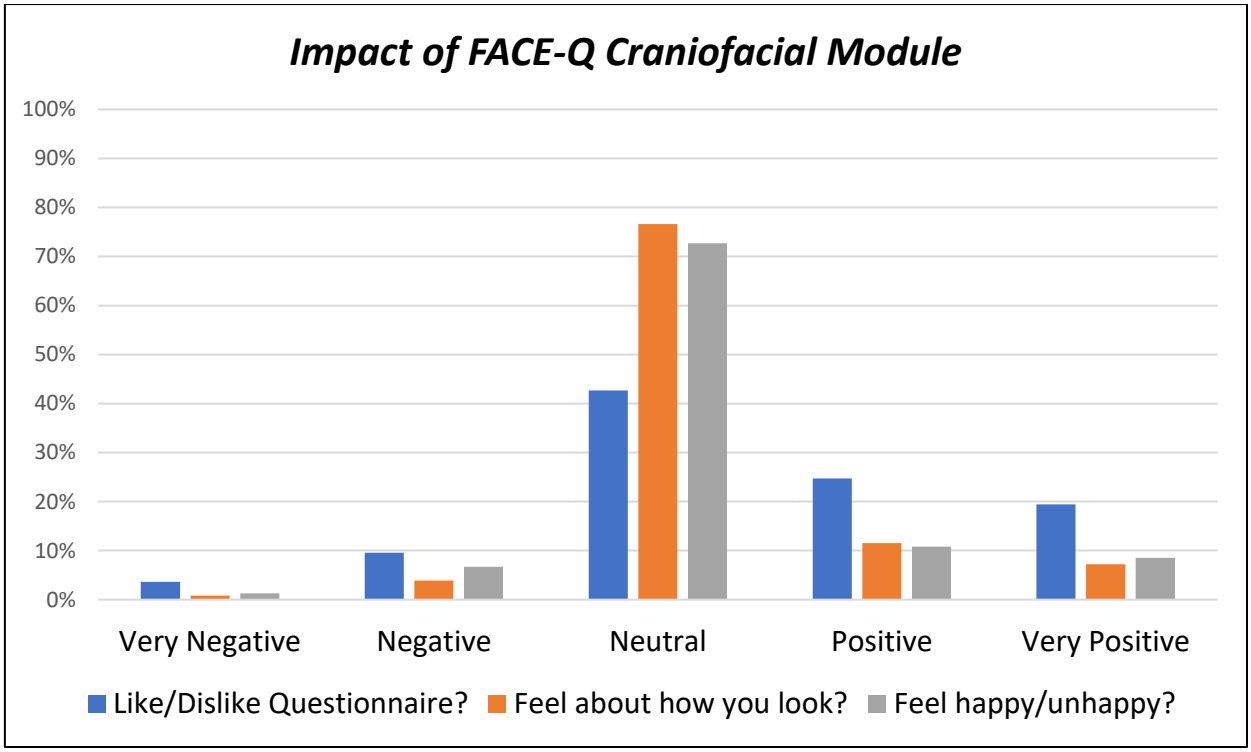

Table 3. *Distribution of participant responses to impact questions*

Participant Response		**Like/Dislike questionnaire? (n=927)**		**Feel about how you look? (n=927)**		**Feel happy/unhappy? (n=927)**	
		N	***N%***	***N***	***N%***	***N***	***N%***
Negative Impact	***Very Negative***	33	3.6%	7	0.8%	12	1.3%
	Negative	89	9.6%	36	3.9%	62	6.7%
Neutral Impact	***Neutral***	396	42.7%	710	76.6%	674	72.7%
Positive Impact	***Positive***	229	24.7%	107	11.5%	100	10.8%
	Very Positive	180	19.4%	67	7.2%	79	8.5%

N, denotes frequency of individual participants; N%, represents percent of the total sample.

In univariate analyses, participants who had an increased facial severity score [β 0.131, OR 1.14 (1.10 to 1.18), p<0.01], completed more FACE-Q scales/checklists [β 0.147, OR 1.16 (1.11 to 1.21), p<0.01], or scored lower (i.e., worse outcomes) on FACE-Q core scales (i.e., Appearance Distress, Psychological, Social, and Facial Appearance scales) were more likely (p<0.01) to report they disliked the questionnaire (**Table 4**).

When statistically significant variables were included in a multivariable model with backward stepwise logistic regression analysis, only participant facial severity score [β 0.068, OR 1.07 (1.00 to 1.14), p=0.04] and the Appearance Distress scale score [β -0.017, OR 0.98 (0.97 to 0.99), p<0.01] remained significant **(Table 5 and 6**).

2. *Did answering this questionnaire change how you feel about how you look (i.e., your appearance)?*

Overall, most participants reported a neutral impact (n=710, 76.6%) about their feelings regarding their appearance following completion of the questionnaire. In addition, n=174 (18.8%) participants reported a positive impact, indicating that they felt 'a little better' (n=107, 11.5%) or 'a lot better' (n=67, 7.2%) about their appearance. Only a minority of participants (n=43, 4.6%) reported a negative impact following completion of the questionnaires, with n=36 (3.9%) and n=7 (0.8%) reporting they felt 'a little worse' and 'a lot worse', respectively **(Figure 2 and Table 3)**.

In univariate analyses, increased facial severity score [β 0.099, OR 1.10 (1.05 to 1.16), p< 0.01], increased number of FACE-Q scales/checklists completed [β 0.113, OR 1.12 (1.05 to 1.19), p< 0.01], and lower scores on all selected FACE-Q scales corresponded to a significant (p<0.01) increased odds of reporting a negative impact on perceived self-appearance **(Table 4).** When these variables were included in multivariable logistic regression model with backward

stepwise logistic regression analysis, only the Appearance Distress scale score remained statistically significant [β -0.03, OR 0.97 (0.93 to 0.99), p<0.01] **(Table 5 and 6)**.

3. *Did answering this questionnaire make you feel unhappy or happy?*

Most participants reported a neutral impact (n=674, 72.7%) following questionnaire completion, indicating they 'felt the same as usual'. In keeping with the other impact questions, relatively few participants reported a positive (n=179, 19.3%) or negative (n=74, 8.0%) impact, with n=62 (6.7%) and n=12 (1.3%) reporting the questionnaire made them feel 'a little unhappy' and 'very unhappy', respectively **(Figure 2 and Table 3)**.

Again, the results of the univariate logistic regression analysis demonstrated that increased facial severity [β 0.078, OR 1.08 (1.04 to 1.13), p< 0.01], increased number of FACE-Q scales /checklists completed [β 0.081, OR 1.08 (1.03to 1.14), p< 0.01], as well as lower scores on the FACE-Q scales were associated with a significant (p<0.01) increase in the odds of reporting a negative impact **(Table 4)**. In addition, those participants who reported no anticipated surgical or nonsurgical treatment for their facial condition were 1.98 times more likely (95% CI 1.21 to 3.24) to report a negative impact relative to participants who indicated that treatment was ongoing.

When applied to a multivariable logistic regression model with backward stepwise logistic regression, only the Social scale score [β -0.03, OR 0.97 (0.95 to 0.99), p< 0.01] and need for additional treatment [β 0.742, OR 2.10 (1.24 to 3.56), p< 0.01] remained significant after adjusting for other variables in the model **(Table 5 and 6)**.

Table 4. *Univariate logistic regression analysis*

Independent Variable	Like/Dislike Questionnaire?				Feel about how you look?				Feel happy/unhappy?			
	β	**SE**	**P-value**	**OR (95% CI)**	**β**	**SE**	**P-value**	**OR (95% CI)**	**β**	**SE**	**P-value**	**OR (95% CI)**
Age, years	0.023	0.02	0.24	1.02 (0.98 to 1.07)	0.026	0.03	0.41	1.03 (0.97 to 1.09)	0.022	0.03	0.37	1.02 (0.97 to 1.07)
Gender (ref = male)	0.074	0.20	0.70	1.08 (0.74 to 1.58)	0.191	0.31	0.54	1.21 (0.66 to 2.23)	-0.168	0.25	0.50	0.85 (0.52 to 1.37)
Facial involvement (ref = unilateral)	-0.196	0.20	0.34	0.82 (0.55 to 1.22)	-0.016	0.33	0.96	0.99 (0.52 to 1.86)	-0.093	0.26	0.72	0.91 (0.55 to 1.50)
Surgery within last 6 months (ref = yes)	0.342	0.25	0.17	1.41 (0.87 to 2.29)	-0.585	0.53	0.27	0.56 (0.20 to 1.59)	-0.594	0.41	0.15	0.55 (0.25 to 1.23)
Treatment status (ref = No treatment, no further treatment anticipated)			0.29				0.75				**0.02****	
Additional treatment anticipated	0.076	0.31		1.08 (0.59 to 1.97)	-0.286	0.44		0.75 (0.32 to 1.78)	-0.269	0.31		0.76 (0.42 to 1.41)
Prior treatment, no further treatment anticipated	0.350	0.28		1.42 (0.83 to 2.43)	-0.285	0.40		0.75 (0.35 to 1.63)	-0.832	0.31		0.44 (0.24 to 0.79)
Craniofacial severity score	0.131	0.02	**<0.01****	1.14 (1.10 to 1.18)	0.099	0.03	**<0.01****	1.10 (1.05 to 1.16)	0.078	0.02	**<0.01****	1.08 (1.04 to 1.13)
Number of FACE-Q scales completed	0.147	0.02	**<0.01****	1.16 (1.11 to 1.21)	0.113	0.03	**<0.01****	1.12 (1.05 to 1.19)	0.081	0.03	**<0.01****	1.08 (1.03 to 1.14)
Appearance Distress scale score	-0.026	0.01	**<0.01****	0.98 (0.97 to 0.98)	-0.046	0.01	**<0.01****	0.96 (0.94 to 0.97)	-0.032	0.01	**<0.01****	0.97 (0.96 to 0.98)
Psychological scale score	-0.018	0.01	**<0.01****	0.98 (0.97 to 0.99)	-0.039	0.01	**<0.01****	0.96 (0.95 to 0.98)	-0.030	0.01	**<0.01****	0.97 (0.96 to 0.98)
Social scale score	-0.021	0.01	**<0.01****	0.98 (0.97 to 0.99)	-0.042	0.01	**<0.01****	0.96 (0.94 to 0.98)	-0.041	0.01	**<0.01****	0.96 (0.95 to 0.97)
Face appearance scale score	-0.021	0.01	**<0.01****	0.98 (0.97 to 0.99)	-0.039	0.01	**<0.01****	0.96 (0.95 to 0.98)	-0.032	0.01	**<0.01****	0.97 (0.96 to 0.98)

**, denotes statistical significance; ref, denotes reference value; SE, refers to standard error; P-value, corresponds to Wald test (t distribution) or General Likelihood Ratio Test (categorical variables, chi-square distribution); OR represents odds ratio with 95% confidence intervals. Goodness-of-fit of the multivariable models were evaluated using the Hosmer-Lemeshow test, all models were deemed to be appropriate. For continuous variables, the listed ORs correspond to a 1 unit increase in the variable value (i.e., increase in 1 year or 1 point on the applicable scale).

Table 5. *Multivariable logistic regression analysis of statistically significant variables*

Independent Variable	Like/Dislike Questionnaire?				Feel about how you look?				Feel happy/unhappy?			
	β	**SE**	**P-value**	**OR (95% CI)**	**β**	**SE**	**P-value**	**OR (95% CI)**	**β**	**SE**	**P-value**	**OR (95% CI)**
Craniofacial severity score	0.069	0.03	**0.04****	1.07 (1.00 to 1.14)	0.038	0.06	0.50	1.04 (0.93 to 1.16)	0.026	0.05	0.58	1.03 (0.94 to 1.12)
Number of FACE-Q scales completed	0.073	0.04	0.054	1.08 (0.99 to 1.16)	0.020	0.06	0.75	1.02 (0.90 to 1.15)	0.046	0.05	0.36	1.05 (0.95 to 1.16)
Appearance Distress scale score	-0.011	0.01	0.13	0.99 (0.97 to 1.00)	-0.026	0.01	**0.02****	0.97 (0.95 to 1.00)	-0.008	0.01	0.41	0.99 (0.97 to 1.01)
Psychological scale score	-0.005	0.01	0.89	1.00 (0.98 to 1.01)	-0.011	0.01	0.38	0.99 (0.97 to 1.01)	-0.005	0.01	0.59	0.99 (0.98 to 1.01
Social scale score	-0.008	0.01	0.37	0.99 (0.98 to 1.01)	-0.009	0.01	0.54	0.99 (0.97 to 1.02)	-0.024	0.01	**0.03****	0.98 (0.96 to 0.99)
Face appearance scale score	-0.002	0.01	0.83	1.00 (0.98 to 1.01)	-0.008	0.01	0.51	0.99 (0.97 to 1.02)	-0.008	0.01	0.41	0.99 (0.97 to 1.01)
Treatment status (ref = No treatment, no further treatment anticipated)	***Not applicable***										**<0.01****	
Additional treatment anticipated									-0.633	0.35		0.53 (0.27 to 1.04)
Prior treatment, no further treatment anticipated									-1.085	0.33		0.34 (0.18 to 0.65)

**, denotes statistical significance; ref, denotes reference value; SE, refers to standard error; P-value, corresponds to Wald test (t distribution) or General Likelihood Ratio Test (categorical variables, chi-square distribution); OR represents odds ratio with 95% confidence intervals. Goodness-of-fit of the multivariable models were evaluated using the Hosmer-Lemeshow test, all models were deemed to be appropriate. For continuous variables, the listed ORs correspond to a 1 unit increase in the variable value (i.e., increase in 1 year or 1 point on the applicable scale).

Table 6. *Backward stepwise logistic regression analysis of statistically significant variables*

Independent Variable	**Like/Dislike Questionnaire?**			
	β	**SE**	**P-value**	**OR (95% CI)**
Craniofacial severity score	0.068	0.03	**0.04****	1.07 (1.00 to 1.14)
Number of FACE-Q scales completed	0.073	0.04	0.055	1.08 (0.99 to 1.16)
Appearance Distress scale score	-0.017	0.01	**<0.01****	0.98 (0.97 to 0.99)
Independent Variable	**Feel about how you look?**			
	β	**SE**	**P-value**	**OR (95% CI)**
Craniofacial severity score	0.054	0.03	0.07	1.05 (0.99 to 1.12)
Psychological scale score	-0.018	0.01	0.06	0.98 (0.96 to 1.00)
Appearance Distress scale score	-0.030	0.01	**<0.01****	0.97 (0.95 to 0.99)
Independent Variable	**Feel happy/unhappy?**			
	β	**SE**	**P-value**	**OR (95% CI)**
Number of FACE-Q scales completed	0.075	0.03	**<0.01****	1.08 (1.02 to 1.14)
Social scale score	-0.030	0.01	**<0.01****	0.97 (0.95 to 0.99)
Face appearance scale score	-0.014	0.01	0.08	0.99 (0.97 to 1.00)
Treatment status (ref = No treatment, no further treatment anticipated)			**<0.01****	
Additional treatment anticipated	-0.670	0.34		0.51 (0.26 to 1.00)
Prior treatment, no further treatment anticipated	-1.090	0.33		0.34 (0.18 to 0.64)

**, denotes statistical significance; ref, denotes reference value; SE, refers to standard error; P-value, corresponds to Wald test (t distribution) or General Likelihood Ratio Test (categorical variables, chi-square distribution); OR represents odds ratio with 95% confidence intervals. Goodness-of-fit of the multivariable models were evaluated using the Hosmer-Lemeshow test, all models were deemed to be appropriate. For continuous variables, the listed ORs correspond to a 1 unit increase in the variable value (i.e., increase in 1 year or 1 point on the applicable scale).

3.2 Analysis of Free-Text Responses

The free-text responses offered diverse insights into the questionnaire itself, the participants' experiences while completing the questionnaire, and their feelings after questionnaire completion. Many participants noted their dislike of the length of the questionnaire; others talked about the repetitiveness of the questions being asked and concerns with specific questions (e.g., if they found a question odd, difficult to answer, did not make sense, etc.). For example, one participant (20-years-old) who completed 19 FACE-Q scales shared, "*It was too long. I never want to answer such a long survey before coming to an appointment*." Another participant (25-years-old who completed 16 scales) stated, "*I think you are asking the same thing again and again and it is too repetitive.*" However, the length of the questionnaire did not dissuade everyone from liking this survey. For example, one participant (10-years-old who completed eight scales) commented, "*I thought this questionnaire was pretty cool. Despite the fact that it was long*." Other participants described the questionnaire as being fun, interesting, and a great experience, in addition to having good and important questions. For example, one participant (13-years-old who completed six scales) said, "*I liked doing this questionnaire. I hope it helps other kids to understand that they are perfect just how they are*."

Regarding their experiences completing the questionnaire, some participants expressed their dislike of thinking about and answering questions about their appearance. Other participants mentioned that the questionnaire was a reminder of their craniofacial conditions and that they are different from others. For example, one participant (14-years-old) remarked, "*It made me feel upset because it just kept reminding me that I look different than everyone because of my scars*." Another participant (20-years-old) said, "*I don't like thinking about all of this. I try to forget that I look different.*" However, other participants enjoyed being able to think more about themselves

and learn new things. Furthermore, for many participants, this questionnaire was seen as an opportunity to help others, open up about their own experiences, and receive help themselves. For example, one participant (20-years-old) stated, "*It just reminded me of how I look different than people but at the same time it is good to help other people or kids who feel the same way. I need to be able to talk about it more.*" Another participant (11-years-old) commented, "*It [made] me feel like I could answer questions I don't really answer all the time and have it go to a good cause that can help other kids like me.*" A participant (12-years-old) also responded, "*Knowing that others have answered these [surveys] I feel that I can fit in because now I know I can also be the same as everyone. Thank you.*"

Lastly, some participants described feeling bad, less confident, stressed out, tired, or losing interest while completing the survey. For example, one participant (9-years-old) commented, "*I [feel] bad answering about my looks*." Another participant (19-years-old) expressed, "*I felt that the questions were meant to bring down my confidence by asking the same thing again and again.*" However, most participants expressed that they were thankful for the questionnaire and happy that this research is being done to help those with craniofacial conditions. Some participants also shared that they felt better after completing the questionnaire. For example, one participant (10-years-old) said, "*It was really fun answering all the questionnaire because it makes me feel better about myself*."

CHAPTER 4: DISCUSSION

4.0 Study Implications

This study found that the majority of participants provided 'neutral' responses to PROM impact questions, with relatively few patients reporting a negative response. This finding suggests that the FACE-Q Craniofacial module is acceptable for most participants. Our results appear to be in keeping with those of Klassen et al.[7], who similarly described a small proportion of participants reporting a negative impact following completion of the CLEFT-Q. Unlike previous analyses, we were unable to demonstrate an association between demographic characteristics such as age or gender and negative participant impact.[7]

Specifically, this study reveals that participants with more severe facial differences, and those who scored lower on the Appearance Distress, Psychological, Social, and Facial Appearance scales were more likely to report a negative impact to completing the PROM. The number of FACE-Q scales completed was also associated with negative responses. This finding was also reflected in the participant comments where many participants expressed the questionnaire was long and repetitive. It is important to note that participants completed the field-test version of the FACE-Q Craniofacial module, which contains more items per subscale with some repetitive items. The final versions of the scales have fewer items.[3,5] Also in practice a participant would likely not complete as many scales as they did in this study. Rather a selection of fewer scales relevant to the purpose of the clinic visit or study would be administered. Therefore, use of the FACE-Q Craniofacial module within different contexts may produce fewer negative response, especially regarding how much participants liked the questionnaire. Also of note, those participants who reported no ongoing anticipated treatment had 1.98 times higher odds of reporting a negative impact only when asked whether the FACE-Q questionnaire made

them feel unhappy/happy; however, the reason(s) 'why' a participant reported a negative impact could not be determined given the cross-sectional nature of this study design and warrants further investigation.

4.1 Application to Clinical Practice

When using this PROM in clinical practice it is important to consider negative impacts of answering potentially distressing questions. The authors recommend that patients who score low on FACE-Q scales, have more severe facial differences, or are asked to complete multiple PROMs as part of their clinic visit should be followed up by a member of the healthcare team to: 1) assess the impact of completing the FACE-Q; 2) address concerns in their scale scores; and 3) ensure that patients are referred appropriately for further supports if required. Our findings about PROM impact are also relevant to other PROMs used in cleft and craniofacial patient care and research, especially given the focus on standardization and use of core outcome sets.[13]

4.2 Study Limitations

There are several limitations to this study. First, participants were recruited at each site using a convenience sample and therefore, it is possible that selection bias took place in recruitment. Secondly, there is no established minimal important difference (MID) for the FACE-Q scales; thus, we were unable to account for the MID when performing our logistic regression analysis. Thirdly, a sample size rule of thumb of 10 events (negative impact) per independent variable was powered to address the first impact question (i.e., "Did you like or dislike answering this questionnaire?"). As such, the remaining impact questions may be underpowered to detect statistically significant independent variables. Moreover, several factors limit the ability to interpret the free-text comments given by participants. Specifically, less than a quarter of the total participants provided comments. In addition, participants provided these

comments after completing the questionnaire, which many described as being a lengthy process. Therefore, even though these comments help shed some light on participants' experiences after completing this questionnaire, both the limited number of comments received and the fact that many of the participants were fatigued when generating these comments impact the quality of this free-text data. Further qualitative research is needed to gain a richer understanding of participants' experiences during and after PROM completion.

4.3 Conclusion

The FACE-Q Craniofacial module is a rigorously developed and validated PROM designed for use in children and young adults with a variety of congenital or acquired conditions that result in a facial difference. Ultimately, this study provides evidence that the FACE-Q Craniofacial module is acceptable for most participants. Clinicians and study investigators should follow-up with patients after completing these PROMs to address areas of concern in their scale scores; and to ensure that a patient is offered further supports, if necessary.

4.4 Funding

The research described in this manuscript was supported by a grant from the Canadian Institute of Health Research (FRN 148779).

4.5 Disclosures

Anne Klassen (Supervisor) and Karen Wong Riff are co-developers of the FACE-Q Craniofacial Module and receive a share of any license revenues based on the inventor sharing policies of the institutions that own them. Anne Klassen is an owner of EVENTUM Research which provides consulting services to the pharmaceutical industry.

REFERENCES:

1. Longmire N, Wong Riff K, O'Hara J, Aggarwala S, Allen G, Bulstrode N et al. Development of a New Module of the FACE-Q for Children and Young Adults with Diverse Conditions Associated with Visible and/or Functional Facial Differences. *Facial Plastic Surgery*. 2017;33(05):499-508.
2. Masnari O, Schiestl C, Rossler J, Gutlein S, Neuhaus K, Weibel L et al. Stigmatization Predicts Psychological Adjustment and Quality of Life in Children and Adolescents With a Facial Difference. *Journal of Pediatric Psychology*. 2012;38(2):162-172.
3. Klassen A, Rae C, Riff K, Denadai R, Murray D, Bracken S et al. FACE-Q Craniofacial Module: Part 2 Psychometric properties of newly developed scales for children and young adults with facial conditions. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2021;.
4. Klassen A, Rae C, Gallo L, Norris J, Bogart K, Johnson D et al. Psychometric Validation of the FACE-Q Craniofacial Module for Facial Nerve Paralysis. *Facial Plastic Surgery & Aesthetic Medicine*. 2021;.
5. Klassen A, Rae C, Wong Riff K, Bulstrode N, Denadai R, Goldstein J et al. FACE-Q Craniofacial Module: Part 1 validation of CLEFT-Q scales for use in children and young adults with facial conditions. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2021;.
6. Turner G, Litchfield I, Finnikin S, Aiyegbusi O, Calvert M. General practitioners' views on use of patient reported outcome measures in primary care: a cross-sectional survey and qualitative study. *BMC Family Practice*. 2020;21(1).

7. Klassen A, Dalton L, Goodacre T, Harman K, Slator R, Tsangaris E et al. Impact of Completing CLEFT-Q Scales That Ask About Appearance on Children and Young Adults: An International Study. *The Cleft Palate-Craniofacial Journal*. 2020;57(7):840-848.
8. Labott S, Johnson T, Fendrich M, Feeny N. Emotional Risks to Respondents in Survey Research: Some Empirical Evidence. *Journal of Empirical Research on Human Research Ethics*. 2013;8(4):53-66.
9. Frattaroli J. Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin*. 2006;132(6):823-865.
10. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics.* 2009;42(2):377-381.
11. Peduzzi P, Concato J, Kemper E, Holford T, Feinstein A. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*. 1996;49(12):1373-1379.
12. Vittinghoff E, McCulloch C. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology*. 2007;165(6):710-718.
13. Stock N, Hammond V, Hearst D, Owen T, Edwards Z, Ridley M et al. Achieving Consensus in the Measurement of Psychological Adjustment to Cleft Lip and/or Palate at Age 8+ Years. *The Cleft Palate-Craniofacial Journal*. 2020;57(6):746-752.