Assessment of facial function and quality of life in patients with peripheral facial palsy

Bruins, Tessa

DOI:
10.33612/diss.246089669

Chapter 5

Interpreting quality of life questionnaires in patients with long-standing facial palsy

Tessa E. Bruins
Martinus M. van Veen
Ingrid J. Kleiss
Dieuwke C. Broekstra
Pieter U. Dijkstra
Koen J.A.O. Ingels
Paul M.N. Werker

Published in Facial Plastic Surgery & Aesthetic Medicine, 2021

Abstract

Objective(s):
To interpret change in quality-of-life scores in facial palsy patients by calculating the smallest detectable change (SDC) and minimal important change (MIC) for the Facial Disability Index (FDI), Facial Clinimetric Evaluation (FaCE) scale, and Synkinesis Assessment Questionnaire (SAQ).

Materials and Methods:
The SDC, for individuals and groups, was calculated using previously collected test–retest data (2-week interval). The MIC (predictive modeling method) was calculated in a second similar facial palsy population using two measurements (1–1.5-year interval) and an anchor question assessing perceived change.

Results:
$SDC_{individual}$ of FaCE was 17.6 and SAQ was 28.2. SDC_{group} of FaCE was 2.9 and SAQ was 4.6 (n = 62). Baseline FaCE and SAQ scores were 43.3 (interquartile range [IQR]: 35.8;55.0) and 51.1 (IQR: 32.2;60.0), respectively. MIC for important improvement of FDI physical/social function, FaCE total, and SAQ total were 4.4, 0.4, 0.7, and 2.8, respectively (n = 88). MIC for deterioration was 8.2, −1.8, −8.5, and 0.6, respectively. Baseline scores were 70.0 (IQR: 60.0;80.0), 76.0 (68.0;88.0), 55.0 (IQR: 40.0;61.7), and 26.7 (IQR: 22.2;35.6), respectively. Number of participants reporting important change for the different questionnaires ranged from 3 to 23 per subscale.

Conclusion:
Interpreting change scores of the FDI, FaCE, and SAQ is appropriate for groups, but for individual patients it is limited by a substantial SDC.

Key points

1) *Question:* How much change in quality of life can reliably be detected in patients with long-standing facial palsy when they are asked to answer questions in a questionnaire?
2) *Findings:* Our study shows that only large changes in quality-of-life scores can reliably be detected, thus making the measurement of small but important individual changes impossible.
3) *Meaning:* The questionnaires of study are not suitable for clinical use in individuals and should mainly be used for research purposes.

Introduction

Facial palsy is a condition affecting facial expression to varying degrees. It has a major influence on both functional and psychosocial aspects of daily life and can, therefore, reduce quality of life considerably.[1-3] In addition to observer-graded facial function, patient-reported outcome measures (PROMs) assessing quality of life are important in evaluating the degree of facial palsy and treatment effects.

The most commonly used instruments for assessing quality of life in patients with facial palsy are the Facial Disability Index (FDI) and the Facial Clinimetric Evaluation (FaCE) scale.[4-6] In addition, the Synkinesis Assessment Questionnaire (SAQ) was designed to assess patient perception of synkinesis severity.[7,8] All three questionnaires are translated to Dutch and validated for the use in patients with facial palsy.[9-11]

Although the validity and reliability of the three questionnaires of interest were considered appropriate, interpretation of changed scores within an individual (interpretability) has not yet been addressed.[12] Estimating whether changed scores, for example, when evaluating treatment effect, actually represent "true" or important change without taking interpretability into account is difficult. Interpreting changed scores is currently not possible for the FaCE and SAQ. For the FDI, the interpretability has partly been resolved, since its measurement variation defined as the smallest detectable change (SDC) has been detected.[11] Changed scores larger than the SDC can be considered as "true change," falling outside the range of normal measurement variation.[13] A way to reduce measurement variation is to repeat measurements. The SDC on group level is, therefore, usually smaller than on individual level, meaning that for individuals a greater change in scores is necessary to ensure true change.[13] Another aspect of interpretability is the minimal important change (MIC), which is the change in scores that is clinically perceived as important by patients.[13] The aim of this study is to determine SDCs and MICs of the FDI, FaCE, and SAQ in two patient cohorts, to be able to help interpret change of scores at group and individual levels.

Materials and Methods

The medical ethics review board of the University Medical Center Groningen (UMCG), the Netherlands, approved the study (METc 2019/491). Written informed consent was obtained from all participants. Adults (≥18 years) diagnosed with peripheral facial palsy were included.

Smallest Detectable Change

To determine the SDC, participants filled out the FaCE and SAQ twice with a 2-week interval. This interval was chosen, assuming that the construct under study did not change and that differences in outcomes are based on variation of the construct within participants. The FaCE and SAQ data were collected during the translation and validation of these questionnaires between December 2012 and August 2014, conducted at the Radboud University Medical Center (RadboudUMC), the Netherlands.[9,10] In addition, patient characteristics such as gender, age, etiology, side of palsy, duration of palsy, and severity of palsy using the Sunnybrook Facial Grading System were collected.

According to COSMIN guidelines, a minimum group sample size of 30 and preferably ≥50 participants is adequate for studying measurement properties and interpretabilit.[14,15] We considered previously collected test–retest data with a minimum of 37 and a maximum of 52 participants per group, therefore, adequate.

Minimal Important Change

The calculation of MIC of the FDI, FaCE, and SAQ was based on two assessments, T1 and T2. Anchor questions at T2 assessed the perceived change of the participants' condition between T1 and T2 for every subdomain and total score on a 7-point Likert scale: "How much did […] change in comparison to approximately one year ago?" T1 was between March and May 2019 at the UMCG (METc 2018/562).[2] For this study, follow-up data (T2) were collected between June and August 2020. Patient characteristics equal to those previously mentioned were collected. With ~1–1.5 years between T1 and T2, depending on the participants' response time, it was assumed that quality of life of a substantial part of the participants had changed due to the changeable nature of the condition and possible effects of treatment.

We hypothesized that, taking the COSMIN guidelines into account,[14,15] 131 participants at T1 would result in a minimum of 30 participants with changed (improved/deteriorated) conditions, and 30 participants with unchanged conditions at T2, leaving room for dropout/nonparticipation.

Questionnaires

The FDI consists of a physical and a social/well-being scale. Both subscales range from 0 (worst) to 100 (best). Each scale contains five questions with a Likert scale ranging from 2 (worst) to 5 (best) and an option for "of other reasons" (0) and "of health" (1).[6]

The FaCE scale consists of 15 questions with a Likert scale ranging from 1 (worst) to 5 (best). A total score and six domain scores can be calculated: facial movement, facial comfort, oral

function, eye comfort, lacrimal control, and social function. Both FaCE total and domain scores range from 0 (worst) to 100 (best).[5]

The SAQ consists of nine questions with a Likert scale ranging from 1 (best) to 5 (worst). The scores are converted into a total score ranging from 20 (best) to 100 (worst).[7] See Appendix 5.1 for characteristics of the three questionnaires.

Statistics

Descriptive statistics are presented as frequency and percentage, median and interquartile range (IQR), and mean with SD as appropriate.

The SDC value at individual level was calculated using the formula 1.96 × × SEM.[13] The required standard error of measurement (SEM) was calculated using the pooled SD and the intraclass correlation coefficient (ICC) of test–retest data (Appendix 5.2).[13, 16] The SDC value at group level was calculated by $SDC_{individual} / \sqrt{n}$

For calculating the MIC, an anchor-based method was used, since it directly asks the patient about their perceptions.[13, 17] The predictive modeling approach was used, resulting in MIC_{pred}, which aims to predict to which group (improved/deteriorated or unchanged) the patient belongs (Appendix 5.2).[18] These groups were defined by the anchor, where participants expressing "somewhat better," "better," and "much better" form the improved group; "no change" forms the unchanged group; and "somewhat worse," "worse," and "much worse" form the deteriorated group. The anchor questions represented the gold standard and the change PROM score the test of interest. The statistical test used in the predictive modeling method is logistic regression.[18] IBM SPSS for Windows version 23.0 (IBM Corp, Armonk, NY) was used for all statistical analyses.

Results

The cohort at the RadboudUMC included 62 participants between December 2012 and August 2014. Forty-two participants were female (67.7%), median duration of palsy was 1.5 years (IQR: 0.9;3.3), and the most common etiology was Bell's palsy (n = 36, 58.1%) (Table 1). Complete test–retest data of 38 participants for the FaCE total score, 46–52 participants for the FaCE subdomains, and 37 participants for the SAQ total score were available (Fig. 1). Out of 62 participants, a Sunnybrook composite score was measured in 35 (median: 52.0, IQR: 30.0;60.0).

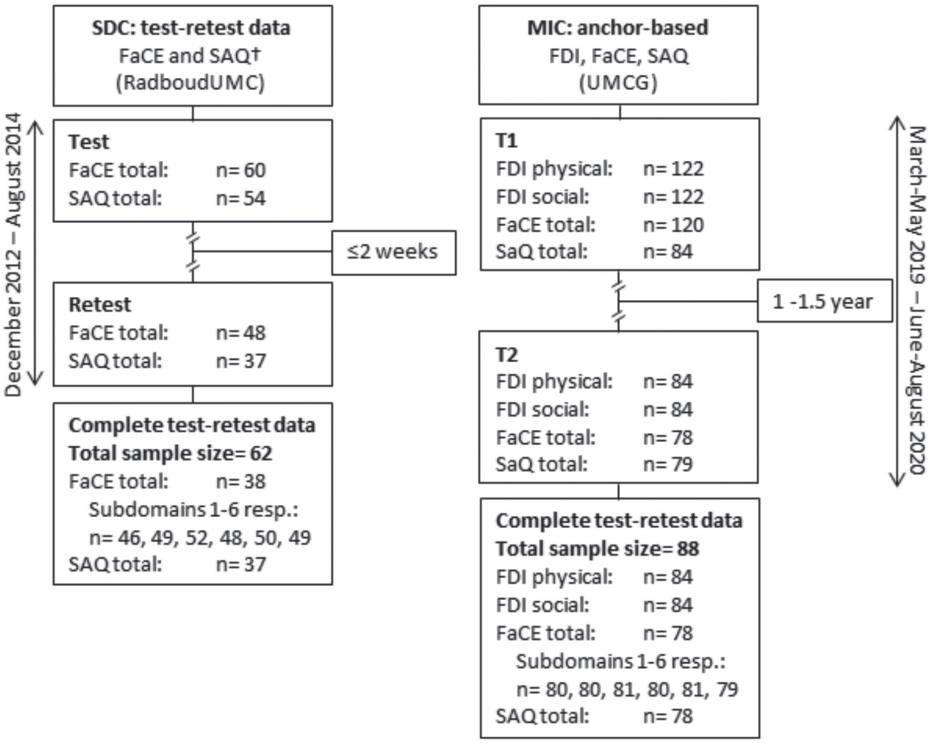

Figure 1. Data collection of the FDI, FaCE scale, and the SAQ to determine the SDC and the MIC. The FaCE has six subdomains: (1) facial movement, (2) facial comfort, (3) oral function, (4) eye comfort, (5) lacrimal control, and (6) social function. The SDC is determined with test–retest data and the MIC with two measurement moments, T1 and T2. In a previous study, test–retest data of the FDI are collected and the SDC has been calculated, we refer to those results. FaCE, Facial Clinimetric Evaluation; FDI, Facial Disability Index; MIC, minimal important change; SAQ, Synkinesis Assessment Questionnaire; SDC, smallest detectable change.

One out of 131 patients who completed the FDI, FaCE, and/or SAQ between March and May 2019 at the UMCG died, resulting in 130 eligible patients who received an invitation for participation and a reminder when appropriate. Fourteen patients were not interested in participating and 28 patients did not respond to the invitation and reminder, resulting in the inclusion of 88 participants (68%) (Fig. 1). Of the respondents at the UMCG, 42 participants were female (47.7%), median duration of palsy was 17.5 years (IQR: 9.6;34.7), and the most common etiology was benign tumor (n = 32, 36.4%) (Table 2). Between T1 and T2, 35% of the participants underwent treatment of any kind for their facial palsy. Of 60 participants, a Sunnybrook composite score was determined at T1, with a median score of 27.5 (IQR: 18.5;34.0). Age and duration of palsy of participants were significantly higher, and the Sunnybrook composite score at T1 was significantly lower than that of nonparticipants (Table 2).

Table 1. Patient characteristics Smallest Detectable Change

Variables	Median (IQR), n= 62
Age (years)	57.0 (46.5; 63.7)
Duration of palsy (years)	1.5 (0.9; 3.3)
Sunnybrook composite score, n= 35	52.0 (30.0; 60.0)
FaCE[a]	
Total, n= 41	43.3 (35.8; 55.0)
Facial movement, n= 50	29.2 (16.7; 50.0)
Facial comfort, n= 51	25.0 (8.3; 50.0)
Oral function, n= 54	50.0 (25.0; 62.5)
Eye comfort, n= 50	37.5 (0.0; 62.5)
Lacrimal control, n= 52	50.0 (25.0; 75.0)
Social function, n= 52	62.5 (45.31; 85.9)
SAQ total[a], n= 41	51.1 (32.2; 60.0)
	n (%)
Gender	
Female	42 (67.7)
Laterality of palsy	
Left	26 (41.9)
Right	35 (56.5)
Bilateral	1 (1.6)
Etiology	
Bell's palsy	36 (58.1)
Infection	15 (24.2)
Ramsay hunt	- 14 (22.6)
Herpes simplex virus	- 1 (1.6)
Trauma (iatrogenic)	5 (8.1)
Acoustic neuroma	4 (6.5)
Brain tumor	1 (1.6)
Otitis	1 (1.6)

[a]Baseline scores.
FaCE, Facial Clinimetric Evaluation scale; IQR, interquartile range; SAQ, Synkinesis Assessment Questionnaire.

Table 2. Patient characteristics Minimal Important Change

Variables	Participants, n= 88 Median (IQR)	Non-participants, n= 42 Median (IQR)	p[a]
Age (years)	63.9 (53.7; 74.3)	61.7 (42.4; 68.2)[b]	0.035*
Duration of palsy (years)	17.5 (9.6; 34.7)	12.6 (5.1; 20.9)[b]	0.033*
Sunnybrook composite score, T1[c]			
Score	27.5 (18.5; 34.0)	42.0 (31.0; 70.0)	<0.001***
Missing n(%)	28 (31.8)	18 (42.9)	
FDI[c]			
Physical function, n= 85 and 37 resp.	70.0 (60.0; 80.0)	70.0 (60.0; 80.0)	0.952
Social function, n= 85 and 37 resp.	76.0 (68.0; 88.0)	72.0 (60.0; 80.0)	0.058
FaCE[c]			
Total, n= 81 and 38 resp.	55.0 (40.0; 61.7)	51.7 (42.9; 62.9)	0.988
Facial movement, n= 83 and 38 resp.	25.0 (8.3; 41.7)	25.0 (0.0; 37.5)	0.925
Facial comfort, n= 85 and 41 resp.	58.3 (33.3; 83.3)	50.0 (25.0; 83.3)	0.796
Oral function, n= 86 and 42 resp.	62.5 (37.5; 87.5)	50.0 (12.5; 75.0)	0.516
Eye comfort, n= 86 and 42 resp.	37.5 (12.5; 62.5)	43.8 (12.5; 75.0)	0.467
Lacrimal control, n= 87 and 42 resp.	50.0 (25.0; 75.0)	50.0 (25.0; 75.0)	0.662
Social function, n= 86 and 41 resp.	75.0 (56.3; 87.5)	75.0 (46.9; 96.9)	0.823
SAQ total†, n= 82 and 3 resp.	26.7 (22.2; 35.6)	42.2 (26.7; -)	0.183
	n (%)	**n (%)**	**p[a]**
Gender			0.325
Female	42 (47.7)	23 (54.8)	
Laterality of palsy			0.329
Left	40 (45.5)	22 (52.4)	
Right	46 (52.3)	17 (40.5)	
Bilateral	2 (2.3)	3 (7.1)	
Etiology			0.542
Benign tumor	32 (36.4)	11 (26.2)	
Trauma	13 (14.8)	9 (21.4)	
Head and neck cancer	10 (11.4)	4 (9.5)	
Infection	9 (10.2)	3 (7.1)	
Congenital	9 (10.2)	5 (11.9)	
Bell's palsy	4 (4.2)	8 (19.0)	
Other	11 (12.5)	2 (4.8)	
Treatment between T1 and T2			
None	57 (64.8)		
Mime	5 (5.7)		
Botox (botulinum toxin)	10 (11.4)		
Botox and static reconstruction	1 (1.1)	-	-
Static reconstruction	9 (10.2)		
Dynamic reconstruction	1 (1.1)		
Other	5 (5.7)		

IQR: interquartile range.
† T1: first measuring moment, other date related values are shown for the second measuring moment.
‡ On the date that questionnaires were sent (03 June 2020).
§ Mann-Whitney U test for scale data and Chi-square test for categorical data.
* p<0.05, **p<0.01, ***p<0.001.

Smallest detectable change

The individual SDC ($SDC_{individual}$) of the FaCE total and SAQ total was 17.6 and 28.2, respectively, with a group SDC (SDC_{group}) score of 2.9 and 4.6, respectively (Table 3). For the FaCE subdomains, the SDC scores are given in Table 3. All $SDC_{individual}$ scores were higher (range: 17.6–43.4) than their SDC_{group} scores (range: 2.9–6.3). ICCs ranged from 0.65 (95% CI = 0.42–0.80) to 0.83 (95% CI = 0.66–0.88).

Table 3. SDC of the FaCE total score, FaCE subdomain scores and the SAQ total score

Variable	ICC (CI 95%)	$SDC_{individual}$	SDC_{group}
FaCE total, n= 38	0.83 (0.68; 0.91)	17.62	2.86
Facial movement, n= 46	0.67 (0.47; 0.80)	34.90	5.15
Facial comfort, n= 49	0.80 (0.66; 0.88)	33.66	4.81
Oral function, n= 52	0.74 (0.58; 0.85)	39.82	5.52
Eye comfort, n= 48	0.76 (0.60; 0.86)	43.41	6.27
Lacrimal control, n= 50	0.67 (0.48; 0.80)	- [a]	- [a]
Social function, n= 49	0.76 (0.60; 0.86)	35.16	5.02
SAQ total, n= 37	0.65 (0.42; 0.80)	28.15	4.63

[a]Subdomain includes only one item.
ICC, intraclass correlation coefficient; SDC, smallest detectable change.

Minimal important change

The MIC determined by the predictive modeling method (MIC_{pred}) of the improved and deteriorated group of the FDI physical function was 4.4 and 8.2, respectively (Table 4). The MIC_{pred} of the improved and deteriorated group of the FDI social function was 0.4 and −1.8, respectively. For the FaCE total score and the SAQ total score, the improved/deteriorated MIC_{pred} values were 0.7/−8.5 and 2.8/0.6, respectively. The sample size of participants who reported important change varied from 3 to 23. All answers to the anchor questions are given in Appendix 5.3.

Table 4. Characteristics of the participants according to the anchor: improved, deteriorated or unchanged.

Domains	n (%)	Changed score Mean (SD)	MIC_{pred} [a]
FDI (n= 84)			
FDI physical improved	14 (16.7)	5.0 (10.4)	4.4
FDI physical unchanged	53 (63.1)	2.3 (9.9)	-
FDI physical deteriorated	17 (20.2)	0.7 (10.6)	8.2
FDI social improved	10 (11.9)	0.0 (10.2)	0.4
FDI social unchanged	68 (81.0)	1.0 (12.6)	-
FDI social deteriorated	6 (7.1)	-4.7 (14.6)	-1.8
FaCE (n= 81)			
FaCE total improved	11 (13.6)	1.7 (16.3)	0.7
FaCE total unchanged	56 (69.1)[b]	-0.5 (8.8)	-
FaCE total deteriorated	13 (16.0)	-0.9 (9.9)	-8.5
Facial movement improved	10 (12.3)	5.0 (7.0)	1.8
Facial movement unchanged	60 (74.1) [b]	-1.4 (17.2)	-
Facial movement deteriorated	10 (12.3)	5.0 (14.3)	1.7
Facial comfort improved	12 (14.8)	9.0 (23.4)	3.3
Facial comfort unchanged	54 (66.7) [b]	-2.2 (21.8)	-
Facial comfort deteriorated	14 (17.3)	-3.0 (22.1)	-2.1
Oral function improved	9 (11.1)	15.3 (18.5)	5.8
Oral function unchanged	58 (71.6)	-3.9 (22.1)	-
Oral function deteriorated	14 (17.3)	-6.3 (14.5)	-4.7
Eye comfort improved	10 (12.3)	8.8 (22.1)	4.3
Eye comfort unchanged	47 (58.0) [b]	-0.3 (16.0)	-
Eye comfort deteriorated	23 (23.4)	-3.8 (16.2)	-2.0
Lacrimal control improved	5 (6.2)	0.0 (17.7)	-1.5
Lacrimal control unchanged	57 (70.4)	-3.1 (23.7)	-
Lacrimal control deteriorated	19 (23.5)	-2.6 (20.2)	-2.6
Social function improved	7 (8.6)	20.5 (27.4)	8.5
Social function unchanged	71 (87.7)	-1.3 (14.4)	-
Social function deteriorated	3 (3.7)	0.0 (10.8)	0.1
SAQ (n= 79)			
SAQ total improved	4 (5.1)	3.9 (4.9)	2.8
SAQ total unchanged	69 (87.3)	1.5 (10.4)	-
SAQ total deteriorated	6 (7.6)	-0.7 (3.0)	0.6

Improved: anchor options 1 (much better), 2 (better), and 3 (somewhat better). Unchanged: anchor option 4 (no change). Deteriorated: anchor options 5 (somewhat worse), 6 (worse), and 7 (much worse).
[a]MIC_{pred} is the MIC based on a predictive modeling approach.
[b]One participant had missing data on questions 1–6 of the FaCE, resulting in one missing total, facial movement, facial comfort, and eye comfort score.
MIC, minimal important change; SD, standard deviation.

Discussion

This study showed the relevance of determining the interpretability of PROMs by calculating the SDC and MIC in three quality-of-life instruments validated for patients with facial palsy. All instruments had substantial $SDC_{individual}$ values, whereas the SDC_{group} values were relatively low. The MIC values were lower than the $SDC_{individual}$ values.

The results of this study are in line with previously reported $SDC_{individual}$ and SDC_{group} values of the FDI physical function (17.6 and 1.9, respectively) and social function (17.7 and 1.9, respectively).[11] Ideally, an instrument should be able to detect small changes, expressed as a low SDC value compared with the range of the scale. The relatively large individual scores suggest that the FDI, FaCE, and SAQ are less suitable for monitoring change in individuals with facial palsy. In the field of facial palsy, it is unlikely to measure changed scores that exceed the SDC, especially when large changes are not expected such as in patients with long-standing facial palsy. However, the relatively small values for groups indicate that these PROMs are useful for interpreting change on a group level. It has been argued that the $SDC_{individual}$ of quality-of-life instruments will often be large due to its subjective nature, but it nonetheless is a shortcoming of this PROM.[19]

The SDC was considerably larger than the MIC on all total and subscales, meaning that there is a relatively large range of scores where important change cannot be distinguished from measurement variation. To ensure that important change reflects true change, the changed score must exceed the SDC. There are no publications examining the MIC of facial palsy-specific PROMs to directly compare these results with. A study evaluating the effect of mime therapy in patients with facial palsy found a mean changed FDI physical score of 16.7 and FDI social score of 12.1.[20] A study examining quality of life after surgical treatment of the periocular complex and a study examining quality of life after free gracilis muscle transfer found mean changed FaCE total scores of 8.6 (11.5 for improved and −10.6 for deteriorated patients) and 16.2, respectively.[21, 22] Of three studies examining synkinesis before and after botulinum toxin treatment, one study found a mean changed FaCE total score of 6.7[23] and two studies found mean changed SAQ total scores of −11.0 and −8.7.[24, 25] All the mentioned changes were significant, but the change of a large proportion of the study participants would fall within the ranges of measurement variation when evaluated according to the $SDC_{individual}$, placing the outcome of these studies in a different light.

There are limitations to this study. Both study populations are heterogeneous, for instance regarding age, duration of palsy, severity of palsy, and etiology. In addition, the UMCG participants of both T1 and T2 had a significantly higher age, longer duration of palsy, and lower Sunnybrook scores than people who only participated at T1, indicating selection bias. Although it could be argued that a heterogeneous sample represents the population visiting a tertiary referral center, certain variables may influence quality of life in patients with facial palsy to such an extent that it can be viewed as a separate population and thus requiring separate interpretability values.[26] Although literature shows conflicting results, in general, a negative association between higher age, female gender, a shorter duration of palsy, and lesser severity of palsy and quality of life in patients with facial palsy

is described.[1,2,27] In this study it was not possible to create subgroups, because the groups would be too small to subanalyze.

Another limitation of this study is the generalizability of the results to the facial palsy population as a whole. Both study populations included patients with mostly long-standing facial palsy, making the results of this study less applicable to patients with acute facial palsy.

According to COSMIN guidelines, the sample size for determining the SDC in this study can be classified as moderate to good.[14,15] However, the sample sizes of all "improved" and "deteriorated" groups used to determine the MIC were small, due to less participants with changed conditions than expected. This makes the interpretation of the MIC in this study to some extend uncertain. An explanation for the small proportion of participants who experienced important change is the relative long median duration of palsy of 17.5 years (IQR: 9.6;34.7). A longer duration of palsy means that large changes of facial function are less likely and will probably also have less impact, since participants may have adjusted to their condition. Changes in patients with long-standing facial palsy can be expected after treatment, but in our study only 35% of participants received treatment between T1 and T2. Patients in an acute stage of Bell's palsy are more likely to experience important change. In the Netherlands, this population is primarily treated by general practitioners, which makes it difficult to recruit a large and representative sample at tertiary referral centers. Another option is to measure change in quality of life before and after treatment, which should be done separately for different interventions given the context-specific nature of measurement properties.[26] Further research is necessary before definitive conclusions about the MIC can be drawn.

A limitation when using an anchor-based method is the possibility of recall bias. Previous research suggests that a longer recall period results in less reliable estimates of change and thus increases recall bias.[28,29] Furthermore, over time people may internally change their standards of quality of life (response shift). A systematic review and meta-analysis examining response shift in quality-of-life research substantiated this and found that patients may adapt their standards, resulting in an underestimation of the "true" quality of life.[30] We consider the determination of the MIC as essential when interpreting changed scores of quality-of-life instruments, because it provides meaning to the change. However, for future research, we recommend a smaller follow-up time to reduce recall bias and response shift. To obtain a sufficient sample size, a center with a high turnover of patients with facial palsy, or a multicenter study should be considered.

Conclusions

The FDI, FaCE, and SAQ are useful for research purposes comparing groups. However, the use of these instruments for monitoring patients with long-standing facial palsy individually is limited by a large individual SDC for all instruments, and results should, therefore, be interpreted with care. The SDC of the three questionnaires is larger than the MIC. Therefore, the change score must exceed the SDC to ensure important change also reflects true change.

References

1. Tavares-Brito J, van Veen MM, Dusseldorp JR et al. Facial Palsy-Specific Quality of Life in 920 Patients: Correlation With Clinician-Graded Severity and Predicting Factors. Laryngoscope 2019;129:100-104.
2. Bruins TE, van Veen MM, Mooibroek-Leeuwerke T et al. Association of Socioeconomic, Personality, and Mental Health Factors With Health-Related Quality of Life in Patients With Facial Palsy. JAMA Otolaryngoloy - Head-Neck Surgery 2020;146:331-337.
3. Volk GF, Granitzka T, Kreysa H et al. Initial severity of motor and non-motor disabilities in patients with facial palsy: an assessment using patient-reported outcome measures. Eur Arch Otorhinolaryngol 2016;274:45-52 LID - 10.1007/s00405-00016-04018-00401 [doi].
4. Ho AL, Scott AM, Klassen AF et al. Measuring quality of life and patient satisfaction in facial paralysis patients: a systematic review of patient-reported outcome measures. Plast Reconstr Surg 2012;130:91-99.
5. Kahn JB, Gliklich RE, Boyev KP et al. Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. Laryngoscope 2001;111:387-398.
6. VanSwearingen JM, Brach JS. The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. Phys Ther 1996;76:1288-1298; discussion 1298-1300.
7. Mehta RP, WernickRobinson M, Hadlock TA. Validation of the Synkinesis Assessment Questionnaire. 2007:923-926.
8. Berner JE, Kamalathevan P, Kyriazidis I et al. Facial synkinesis outcome measures: A systematic review of the available grading systems and a Delphi study to identify the steps towards a consensus. J Plast Reconstr Aesthet Surg 2019;72:946-963.
9. Kleiss IJ, Beurskens CH, Stalmeier PF et al. Synkinesis assessment in facial palsy: validation of the Dutch Synkinesis Assessment Questionnaire. Acta Neurol Belg 2016;116:171-178.
10. Kleiss IJ, Beurskens CH, Stalmeier PF et al. Quality of life assessment in facial palsy: validation of the Dutch Facial Clinimetric Evaluation Scale. Eur Arch Otorhinolaryngol 2015;272:2055-2061.
11. van Veen MM, Bruins TE, Artan M et al. Health-related quality of life in facial palsy: translation and validation of the Dutch version Facial Disability Index. 2020:256 LID - 210.1186/s12955-12020-01502-12950 [doi] LID - 12256.
12. Mokkink LB, Terwee CB, Patrick DL, Alonso J et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. 2010:737-745 LID - 710.1016/j.jclinepi.2010.1002.1006 [doi].
13. De Vet H, Terwee, C, Mokkink, L, Knol, D. Measurement in Medicine: A Practical Guide. (Cambridge: Cambridge University Press). 2011.
14. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. 2011:651-657 LID - 610.1007/s11136-11011-19960-11131 [doi].
15. Mokkink LB, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. (Amsterdam Public Health research institute; Amsterdam University Medical Centers, location VUmc). 2019.
16. Altman DG, Bryant TN, Gardner MJ. Statistics with confidence: confidence intervals and statistical guidelines. London. BMJ books. 2005
17. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395-407.
18. Terluin B, Eekhout I, Terwee CB et al. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. Journal of Clinical Epidemiology 2015;68:1388-1396 LID - S0895-4356(1315)00160-00162 [pii] LID - 00110.01016/j.jclinepi.02015.00103.00015 [doi].

19. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171-184.
20. Beurskens CH, Heymans PG. Positive effects of mime therapy on sequelae of facial paralysis: stiffness, lip mobility, and social and physical aspects of facial disability. Otol Neurotol 2003;24:677-681.
21. Henstrom DK, Lindsay RW, Cheney ML et al. Surgical treatment of the periocular complex and improvement of quality of life in patients with facial paralysis. Arch Facial Plast Surg 2011;13:125-128.
22. Lindsay RW, Bhama P, Hadlock TA. Quality-of-life improvement after free gracilis muscle transfer for smile restoration in patients with facial paralysis. JAMA Facial Plast Surg 2014;16:419-424.
23. Couch SM, Chundury RV, Holds JB. Subjective and objective outcome measures in the treatment of facial nerve synkinesis with onabotulinumtoxinA (Botox). Ophthalmic Plast Reconstr Surg 2014;30:246-250.
24. Shinn JR, Nwabueze NN, Du L et al. Treatment Patterns and Outcomes in Botulinum Therapy for Patients With Facial Synkinesis. JAMA Facial Plast Surg 2019;21:244-251.
25. Neville C, Venables V, Aslet M et al. An objective assessment of botulinum toxin type A injection in the treatment of post-facial palsy synkinesis and hyperkinesis using the synkinesis assessment questionnaire. J Plast Reconstr Aesthet Surg 2017;70:1624-1628.
26. Revicki D, Hays RD, Cella D et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-109.
27. Kleiss IJ, Hohman MH, Susarla SM et al. Health-related quality of life in 794 patients with a peripheral facial palsy using the FaCE Scale: a retrospective cohort study. Clin Otolaryngol 2015;40:651-656.
28. Schmitt JS, Abbott JH. Patient global ratings of change did not adequately reflect change over time: a clinical cohort study. Phys Ther 2014;94:534-542.
29. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. J Health Econ 2014;35:34-46.
30. Schwartz CE, Bode R, Repucci N et al. The clinical significance of adaptation to changing health: a meta-analysis of response shift. Qual Life Res 2006;15:1533-1550.

Appendix 5.1. Characteristics of the Facial Disability Index, the Facial Clinimetrical Evaluation scale and the Synkinesis Assessment Questionnaire

Questionnaire	(sub)scales	No of items	Scoring	Range of score
FDI	Physical function	5	$\frac{\text{Total score (questions 1–5)}-\text{N}}{\text{N}} \text{ x } \frac{100}{4}$ N= number of questions answered	0 (worst) - 100 (best)
	Social/well-being function	5	$\frac{\text{Total score (questions 6–10)}-\text{N}}{\text{N}} \text{ x } \frac{100}{5}$ N= number of questions answered	0 (worst) - 100 (best)
	Total	15	$\frac{((\text{Sum of all 15 items})-\text{\# valid})}{4}$ x (# valid) x 100	0 (worst) - 100 (best)
FaCE	Facial movement	3	$\frac{((\text{Items 1+2+3})-\text{\# valid})}{4}$ x (# valid) x 100	0 (worst) - 100 (best)
	Facial comfort	3	$\frac{((\text{Items 4+6+16})-\text{\# valid})}{4}$ x (# valid) x 100	0 (worst) - 100 (best)
	Oral function	2	$\frac{((\text{Items 11+12})-\text{\# valid})}{4}$ x (# valid) x 100	0 (worst) - 100 (best)
	Eye comfort	2	$\frac{((\text{Items 5+7})-\text{\# valid})}{4}$ x (# valid) x 100	0 (worst) - 100 (best)
	Lacrimal control	1	$\frac{(\text{Item 8}-\text{\# valid})}{4}$ x (# valid) x 100	0 (worst) - 100 (best)
	Social function	4	$\frac{((\text{Items 9+10+14+15})-\text{\# valid})}{4}$ x(# valid) x 100	0 (worst) - 100 (best)
SAQ	Total	9	$\frac{(\text{Sum of scores 1 to 9})}{45}$ x 100	20 (best) - 100 (worst)

FDI: Facial Disability Index, FaCE: Facial Clinimetric Evaluation scale, SAQ: Synkinesis Assessment Questionnaire.

Appendix 5.2. Formulas for calculating the smallest detectable change and minimal important change

Definitions

SDC= Smallest Detectable Change

SEM= Standard Error of Measurement

SD_{pooled}= Pooled Standard Deviation

ICC= Intraclass Correlation Coefficient

MIC= Minimal Important Change

Formulas for calculating SDC

$SDC_{individual} = 1.96 \times \sqrt{2} \times SEM$

$SEM = SD_{pooled} \times \sqrt{(1 - ICC)}$

$SD_{pooled} = \sqrt{\frac{(n_1 - 1) \times SD_1{}^2 + (n_2 - 1) \times SD_2{}^2)}{(n_1 + n_2 - 2)}}$

ICC= two-way random effects model with single measures and absolute agreement

$SDC_{group} = SDC_{individual} / \sqrt{n}$

Calculating MIC

The MIC_{pred} is characterized by the change score corresponding with a likelihood ratio of 1 as proposed in the paper of Terluin B et al.[1] Following the guidelines of Terluin B et al., MIC_{pred} cutoff is then calculated as: $X = (\ln(oddspre) - C) / B_X$, where C represents the intercept and B_X the regression coefficient.

5

Reference

1. Terluin B, Eekhout I, Terwee CB et al. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. Journal of Clinical Epidemiology 2015;68:1388-1396 LID - S0895-4356(1315)00160-00162 [pii] LID - 00110.01016/j.jclinepi.02015.00103.00015 [doi].

Appendix 5.3. Answers participants on the anchor questions

Anchor	Much better, N (%)	Better, N (%)	Somewhat better, N (%)	No change, N (%)	Somewhat worse, N (%)	Worse, N (%)	Much worse, N (%)
FDI physical function, n= 84	0 (0.0)	5 (6.0)	9 (10.7)	53 (63.1)	16 (19.0)	1 (1.2)	0 (0.0)
FDI social function, n= 84	0 (0.0)	4 (4.8)	6 (7.1)	68 (81.0)	6 (7.1)	0 (0.0)	0 (0.0)
FaCE total, n= 81	0 (0.0)	2 (2.5)	9 (11.1)	57 (70.4)	12 (14.8)	1 (1.2)	0 (0.0)
Facial movement, n= 81	0 (0.0)	3 (3.7)	7 (8.6)	61 (75.3)	9 (11.1)	1 (1.2)	0 (0.0)
Facial comfort, n= 81	1 (1.2)	2 (2.3)	9 (11.1)	55 (67.9)	13 (16.0)	1 (1.2)	0 (0.0)
Oral function, n= 81	0 (0.0)	3 (3.7)	6 (7.4)	58 (71.6)	13 (16.0)	1 (1.2)	0 (0.0)
Eye comfort, n= 81	0 (0.0)	3 (3.7)	7 (8.6)	48 (59.3)	15 (18.5)	8 (9.9)	0 (0.0)
Lacrimal control, n= 81	0 (0.0)	0 (0.0)	5 (6.2)	57 (70.4)	12 (14.8)	6 (7.4)	1 (1.2)
Social comfort, n= 81	0 (0.0)	2 (2.5)	5 (6.2)	71 (87.7)	3 (3.7)	0 (0.0)	0 (0.0)
SAQ total, n= 79	0 (0.0)	2 (2.5)	2 (2.5)	69 (78.4)	6 (7.6)	0 (0.0)	0 (0.0)

FDI: Facial Disability Index, FaCE: Facial Clinimetric Evaluation scale, SAQ: Synkinesis Assessment Questionnaire.

Part II

Factors associated with quality of life