The Toronto Outcome Measure for Craniofacial Prosthetics: A Condition-Specific Quality-of-Life Instrument

Jim D. Anderson, BSc, DDS, MScD[1]/John P. Szalai, PhD[2]

Purpose: *The objective was to develop a patient-based outcome measure of condition-specific quality of life that would minimize measurement error related to the instrument when used with patients requiring extraoral craniofacial prostheses.* ***Materials and Methods:*** *An item pool of potential questionnaire items covered 10 clinical/technical and social/psychologic domains. They sought how frequently the issue in the item affected patients and how important the problem in the item was. The 139 items were administered to 94 treated patients in 5 centers in the United States, Canada, and the United Kingdom. Items were eliminated using relevance (frequency × importance), frequency of answer endorsement, Cronbach's alpha (internal consistency), and correlation of items on the same subject. International cultural agreement was tested using analysis of variance and Tukey comparisons within each domain. Scoring was transformed to a scale (0 to 100).* ***Results:*** *The final instrument contained 52 items yielding a mean quality-of-life score of 72.5% and a standard deviation of 17.9. Very high internal consistency was demonstrated with a final Cronbach's alpha of 0.967. No international cultural disagreement was found in 9 of the 10 domains.* ***Discussion:*** *The relative weight of each of the domains is (partially) based on the relevance to the patients. Of the 52 items, 29 were identified that do not mention a prosthesis. This subscale has a Cronbach's alpha of 0.976. These items may therefore be useful where within-patient change is of interest.* ***Conclusion:*** *A patient-based outcome measure of condition-specific quality of life has been developed with control of bias and demonstrated performance characteristics.* (INT J ORAL MAXILLOFAC IMPLANTS 2003;18:531–538)

Key words: *craniofacial prosthetics, quality of life, questionnaire*

Since patients' reactions to facial surgery and prostheses have been shown to be poorly correlated to "objective" outcomes, and particularly to the opinion of the provider,[1] there is heightened interest in patient-based outcomes as a major component of clinical research in craniofacial prosthetics. Several such outcome measures have been reported in the literature.[2–7] A variety of approaches have been taken in their development, but evidence of their performance characteristics in this population of patients is very limited. The early instruments consisted of questionnaires that appear to have been composed by the providers with no pretesting of the questionnaire. Jani and Schaaf[3] explored the wear habits, comfort, technical adequacy, and maintenance of the prosthesis without reference to psychologic and social impact. Two reports from Sela and Lowental[4,5] focused on psychologic health while almost ignoring those factors related to the prosthesis itself that had been covered by Jani and Schaaf. Two later reports[6,7] that looked at overall use and patient satisfaction provide no reference to how the instruments were developed or their performance characteristics.

[1] Director, Craniofacial Prosthetic Unit, Toronto Sunnybrook Regional Cancer Centre, Toronto, Canada.
[2] Director, Research Design and Biostatistics, Senior Scientist, Clinical Epidemiology, Sunnybrook and Women's College Health Sciences Centre, Toronto, Canada.

Reprint requests: Dr J. D. Anderson, Craniofacial Prosthetic Unit, Toronto Sunnybrook Regional Cancer Centre, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5 Canada. Fax: +416-480-6801. E-mail: jim.anderson@utoronto.ca

Table 1 No. of items in Original Pool and Completed Instrument by Domain

Domains	No. of items in completed pool	No. retained after first reduction	No. of items in completed instrument
Clinical/technical			
Fit and retention	11	7	7
Comfort	11	5	5
Esthetics	11	7	6
Maintenance	7	3	2
Social/psychologic			
Body image	16	8	7
Social interactions/roles			
Leisure	18	10	10
Work/school	8	3	3
Family/friends/ strangers	30	4	3
Mood	19	7	7
Sexuality	8	2	2
Total	139	56	52

The first reduction was done by tests of relevance, frequency of endorsement, and the first Cronbach's alpha test for internal consistency. The instrument was completed after the further removal of highly correlated items on similar subjects and second Cronbach's alpha.

The most recent instrument developed by Sloan and others[2] was generated from the literature using items the authors believed to be important to patients or were expected to show moderate shifts (or both). To keep the instrument very short, only 1 item was used to measure a domain. The items were intended to stand alone to stimulate further investigations into quality of life, rather than to be used as a summative scale.

A measure of condition-specific quality of life would be beneficial for testing of new innovations in this field. The objective of this project was to create a patient-based outcome measure of condition-specific quality of life for patients requiring extraoral craniofacial prostheses.

METHODS

Preliminary Item Pool

Domains were selected to be consistent with the World Health Organization's definition of health: "a state of complete physical, mental, and social well-being, and not merely the absence of disease or infirmity." They were chosen for their likely importance to this population of patients, relevance to caregivers, and potential to show between-patient or within-patient differences. This was done by reviewing the craniofacial prosthesis literature[3–10] and the head and neck cancer and facial injury literature.[11–14] In addition, input was sought by consultation with a prosthodontist and a psychologist familiar with this population of patients and a member of a support group for people with facial differences. The domains chosen are listed in Table 1.

Like the domains, a preliminary pool of items was generated from the literature, adding to the items already generated in a pilot project.[15] The items were chosen from 3 sources in the literature:

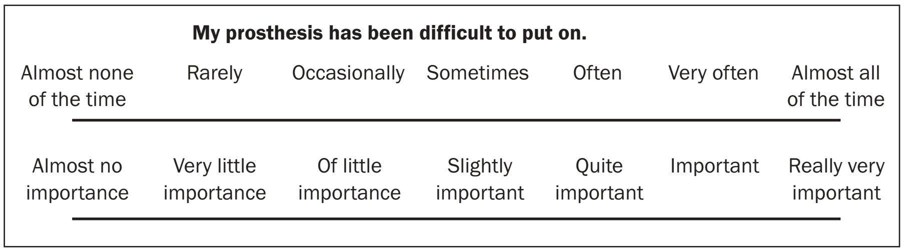
1. Previous instruments used in the maxillofacial prosthetic literature[3–10,16,17]
2. Previous generic or site-specific instruments used in cancer or other facial injury patients[18–23]
3. Previous generic instruments that are relevant to the domains chosen[24–27]

Similarly, the items were chosen or adapted from these sources for their likely importance to this population of patients, relevance to caregivers, and potential to show between-patient or within-patient differences. The domains were represented by subscales of items. This yielded a preliminary pool of 87 items.

Step 1: Development of the Item Pool

The wording of each item was modified to fit within a 13-year-old's reading level.[28] After institutional ethical review, a convenience sample of 9 treated patients was selected to represent a broad range of age (over 18), defect type, severity, and gender. After obtaining informed consent, the first of these patients was given the preliminary list of items and asked to review each domain and item during an interview with the clinic coordinator. The patient was asked to comment on the clarity of the items and the extent to which the items cover the relevant domains. In addition, the patient was asked to add domains or items that described experiences or feelings they had had that were not covered by the preliminary pool. Changes and additions were made to the item pool, and the process was repeated in an iterative fashion for the remaining patients in the sample. No new domains or items were added by the last respondents, thus supporting the face validity and content validity of the item pool.

Five providers were asked in a single mailed poll for comment on the clarity and the extent to which the items covered the relevant domains. Providers were also asked to add domains or items that described patient problems they had seen that were not covered by the preliminary pool. At the end of these 2 processes, the original pool of 87 items had been enlarged to 139 items (Table 1).



My prosthesis has been difficult to put on.

Almost none of the time	Rarely	Occasionally	Sometimes	Often	Very often	Almost all of the time
Almost no importance	Very little importance	Of little importance	Slightly important	Quite important	Important	Really very important

Fig 1 Sample item in the Item Reduction Questionnaire

All items were stated in the past tense so that they would relate to the patient's experiences and feelings over the last month. The questions were structured to accept an answer that reflected how often the patient had been affected by the issue raised in the item. To avoid double negative confusion and to improve item validity,[29,30] all the items were positively phrased but describe illness states.[18] For example, the item "I feel ill" was used, rather than "I feel well," "I do not feel well," or "l do not feel ill." It was assumed that social desirability bias[31] may be operating in this population of patients. If so, a ceiling effect would be created, with most patients' answers bunched at the "good health" end of the scales, leaving little room to differentiate between patients or to show improvement. The items were therefore deliberately framed to bias the answers toward an illness direction to minimize the ceiling effect.

Step 2: Item Reduction

Two sets of equally spaced 7-point adjectival scales[32] were then added to each item: a "frequency set" and an "importance set." The first set contained descriptors of how often the patient was affected in the last month by the issue raised in the item. The second set sought an estimate of the importance of that issue to the patient. To avoid end-aversion effects,[33] extreme anchors were avoided. The frequency set was anchored by "Almost all the time" and "Almost none of the time" rather than "Always" and "Never." Similarly, the importance set was anchored by "Really very important" and "Almost no importance." A typical item is shown in Fig 1.

To simplify mathematical complexities, all items were given equal weight, and response options were scored 1 through 7: A score of 1 was given to "Almost none of the time" and "Almost no importance," and a score of 7 was given to the opposite extremes. To minimize acquiescence bias[34] and halo effects[35] within a domain, the items in all the domains were mixed up at random, so that there was no relationship between consecutive items.

To support the generalizability of the finished instrument, item reduction was done using groups of patients in different English-speaking countries. The total sample of 94 patients included treated individuals from 5 craniofacial prosthetic centers in Canada, the United States, and the United Kingdom. No attempt was made to translate the instrument for testing in other cultures. The number of patients recruited at each center for the item reduction step is shown in Table 2.

Table 2 No. of Treated Patients Recruited at Each Center for Step 2 (Item Reduction)

Craniofacial center	No. of patients recruited
Craniofacial Prosthetic Unit, Toronto Sunnybrook Regional Cancer Centre, Toronto, Ontario, Canada	29
Department of Prosthetic Dentistry, King's Dental Institute, London, United Kingdom	8
Oral and Maxillofacial Surgery Department, Canniesburn Hospital, Bearsden, Glasgow, Scotland	14
COMPRU, Misericordia Hospital, Edmonton, Alberta, Canada	11
Mayo Clinic, Rochester, Minnesota	32
Total sample of treated patients	94

Separate ethical approvals were obtained at each center.

Analysis

Three main strategies were used to eliminate poorly performing items from the instrument, followed by minor refinements.

1. *Relevance.* The finished instrument should contain only items that were shown to be relevant to most patients. Since an item would be relevant to a patient only if the patient was frequently affected and/or the item was very important to the patient, relevance was defined as the product of frequency and importance.[36] Since each of the frequency and importance scales was scored 1 to 7, the range of relevance scores was 1 to 49 for each item.
2. *Frequency of Endorsement.* If nearly all the respondents reply to an item using the same answer option, then clearly the item is not useful for discriminating between patients and should therefore

be eliminated from the final instrument. The same reasoning applies if almost none of the patients use an answer option. The frequency set of responses was examined for frequency of endorsement of each of the response options. Where any response option at the extremes was endorsed more than 75% or less than 2% of the time, the item was discarded.[33] Similarly, where any 2 adjacent responses accounted for more than 80% of the answers, the item was discarded.

3. *Internal Consistency.* Since each of the subscales is intended to measure different aspects of the same attribute, the items within each of the subscales should correlate closely with each other. Items that do not correlate with each other may be measuring some other trait and thus should not be part of the subscale. Similarly, to the extent that each of the subscales is measuring a common attribute, the items among the different subscales should correlate (less closely) with each other. The frequency set of responses was examined within each subscale for internal consistency using Cronbach's alpha.[37] Where the internal consistency (alpha score) for a subscale was increased by deletion of an item, the item was discarded.

   The Cronbach's alpha test was repeated after the items had been removed from the item pool using these 3 strategies to determine if diminution of the internal consistency occurred because of the reduced number of items, and to identify any other inconsistent items. The test was repeated again for record purposes on the final set of items after the elimination of the highly correlated items (below).

Highly Correlated Items on the Same Subject. The subject matter of each of the items remaining from the reductions above was examined for apparent overlap between items. Where items with similar subject matter were found to be highly correlated with each other, this suggests that each item measured nearly the same thing and thus added very little new information. Where a pair of items had a correlation above 0.8, the item with the poorer relevance was discarded.

Cultural Agreement. Frequency responses from the 5 centers on the final number of items were then grouped by nationality (British, Canadian, and American) and examined by domain for any systematic differences that would indicate cultural differences using an analysis of variance followed by Tukey pairwise comparisons. For these tests, $P < .05$ was used to assert statistical significance.

Scoring

The total score per patient is the sum of the individual per-item scores (range 1 to 7) where "1" represents a "good quality of life" score and "7" represents a "bad quality of life" score. The range of total scores will thus go from n to 7n, where n represents the number of items in the final instrument. This range of scores is counterintuitive, because the high scores represent bad quality of life and the low scores represent good quality of life, and the range spans a series of numbers that offer no insight into the degree of good or bad health. Using an approach similar to other health and quality-of-life measures,[38] the scores therefore were transformed to a percentage scale so that a high score represents good quality of life. This inversion and transformation is represented by the following formula:

$$\text{Percentage score} = ([7n - \text{raw score}]/[7n - n]) \times 100.$$

The patient's raw score was subtracted from the highest score available, and that difference was divided by the number of possible scores over the range n to 7n. The result was expressed as a percentage of the maximum attainable score. This transformed score permitted a more intuitive interpretation of a patient's condition-specific quality of life.

RESULTS

The data from all centers were gathered and analyzed together. Because of a photocopying error, 14 patients from 1 center did not provide answers for the last 64 items in the original list of 139 items. Fortunately, the randomized ordering of the items minimizes the impact of this problem on any single domain. Apart from this block, the rate of missing data was 2.75%. All analyses were done without alteration of the missing data.

Table 3 lists the relevance raw data by domain. The relatively low means for each domain and the low average maximums suggest that the issues raised in the items generally affected the patients infrequently and/or that they were relatively unimportant to the patients. In other words, the patients were generally unaffected by their facial difference and their prosthesis. This suggests that the suspected ceiling effect toward the "good quality of life" end of the response options was indeed operating. Items scoring less than 10 on the relevance scale were eliminated. Twenty-five such items were eliminated in this way.

Table 3 Relevance (Frequency × Importance) Scores by Domain

Domain	Mean	SD	Minimum	Maximum
Fit and retention	12.2	6.2	1.0	37.4
Comfort	10.5	6.3	1.0	29.2
Esthetics	13.4	7.9	1.0	41.2
Maintenance	9.7	6.2	1.0	42.0
Body image	11.6	7.7	1.2	41.8
Leisure	10.0	7.5	1.0	38.0
Work/school	7.6	5.4	1.0	32.9
Family/friends/ strangers	7.0	4.0	1.1	19.0
Mood	8.5	5.7	1.0	32.3
Sexuality	8.3	6.3	1.0	31.9

The domain means are the average across all respondents of the per-item scores (range 1–49) in each domain. The minimum scores are the lowest domain means among the 94 respondents. Similarly, the maximum scores are the highest domain means among the 94 respondents.

The frequency of endorsement for an example item with poorly distributed answers is given in Table 4. When most patients give nearly the same answer, discrimination between patients becomes nearly impossible. Therefore, the item provides no useful information and is discarded. Thirty-one such items were discarded in this way.

The results of the first pass on 139 items showed high internal consistency within nearly every domain. Where the alpha scores were lower, items contributing to the lowered scores were removed because of their inconsistent behavior. Thirty-three items were removed in this way (6 items overlapped with the frequency of endorsement items).

By these 3 means, the item pool was reduced from the original 139 items to 56. A repeat of the Cronbach's alpha test on these 56 items revealed 1 more inconsistent item in the maintenance domain. This item was eliminated.

Inspection of the remaining 55 items revealed 13 pairs of items that appeared to address the same subject matter. For example, Item 36 ("It has been difficult for me to meet strangers") is very similar to Item 70 ("I have been anxious about meeting new people"). Three of these pairs had Pearson correlations with each other above 0.8. In every case, the item with the poorer relevance had the lower correlation with the other items in the domain and was eliminated. The other 10 pairs with the lower correlations with each other were retained, since the lower correlations suggested that each item added new information.

A final test of internal consistency using Cronbach's alpha on the remaining 52 items in the finished instrument is presented in Table 5. It reveals a high overall internal consistency of α = 0.967 (Table 5).

Table 4 Example of an Item with Poorly Spaced Frequency of Endorsement

Response no.	Descriptor	Frequency	% frequency
1	Almost none of the time	75	83.3
2	Rarely	11	12.2
3	Occasionally	2	2.2
4	Sometimes	2	2.2
5	Often	0	0
6	Very often	0	0
7	Almost all of the time	0	0

The statement in question was "My family thought of me as someone who is ill." More than 95% of the answers are concentrated in only 2 answer options, and 3 other options are not used, providing poor discrimination between patients.

Table 5 Results of the Final Cronbach's Alpha Test for Internal Consistency on the Completed Instrument of 52 Items

Domain	Cronbach's alpha
Fit and retention	0.749
Comfort	0.768
Esthetics	0.828
Maintenance	0.480
Body image	0.899
Leisure	0.948
Work/school	0.838
Family/friends/strangers	0.652
Mood	0.822
Sexuality	0.807
Overall	0.967

Note the overall Cronbach's alpha exceeds the 0.8 suggested for use in comparing groups and is even above the "desirable standard" of 0.95 for using this instrument to compare individuals against norms.[40]

The average frequency responses by domain for each of the national groups are presented in Table 6. Consistency was evident between the national groups, suggesting relatively small cultural differences among the national groups. The only exception was the maintenance domain, where there was no difference between the Canadian and American groups, but the British responses were inconsistent with both the Canadian and American responses, as indicated by the Tukey pairwise comparisons ($P < .05$).

The completed instrument thus had 52 items (Table 1). The mean raw score was 128.9 with a standard deviation of 56.3. The lowest and highest total raw scores were 41 and 330, respectively. The lowest total raw score (41) reflects the missing data, since it is lower than the number of items in the instrument. The usefulness of these raw scores as an

Table 6 British-Canadian-American Cultural Agreement on the 52 Items Remaining After Reduction from the Initial Item Pool

	Domain means (frequency set)			
Domain	UK	Canada	USA	*P* value*
Fit and retention	2.43	2.71	2.77	0.494
Comfort	3.10	2.65	2.41	0.141
Esthetics	2.87	3.21	3.06	0.676
Maintenance	3.40	2.26	2.42	0.012†
Body image	3.19	2.91	2.79	0.613
Leisure	2.52	2.59	2.40	0.855
Work/school	2.67	2.40	2.22	0.882
Family/friends/ strangers	2.74	2.61	2.43	0.719
Mood	2.38	2.53	2.43	0.885
Sexuality	2.50	2.29	2.22	0.800

All domains show no disagreement except the Maintenance domain.
*Analysis of variance, national group main effects; †statistically significant difference.

Table 7 Frequency Set Average Scores (Range 1 to 7) and Standard Deviations for Each Domain and the Whole 52-item Completed Instrument (Range 52 to 364), After Correction for Missing Data

Domain	Mean	SD	Minimum	Maximum
Fit and retention	2.7	1.1	1	6.1
Comfort	2.7	1.2	1	6.6
Esthetics	3.1	1.4	1	6.8
Maintenance	2.6	1.5	1	7.0
Body image	2.9	1.5	1	6.9
Leisure	2.5	1.4	1	7.0
Work/school	2.3	1.5	1	7.0
Family/friends/ strangers	2.6	1.4	1	7.0
Mood	2.5	1.2	1	6.1
Sexuality	2.3	1.5	1	7.0
Total instrument	137.7	55.8	56	336.6

The means are the average across all respondents of all the per-item scores in each domain. The minimums are the lowest domain means among the 94 respondents. The maximums are the highest domain means among the 94 respondents. For the total instrument, the mean and standard deviation are the average and standard deviation across all respondents of their total score for the 52 items. The minimum total score is the lowest total score recorded. The maximum total score is the highest total score recorded.

indication of population norms is therefore somewhat limited. To report more meaningful norms, and to provide a strategy for users to deal with missing data when the instrument is in clinical use, the patient's individual domain mean for that item was imputed for each missing data point. The resulting response averages and standard deviations for each domain and for the total instrument are given in Table 7. The overall average score was 137.7 on a scale of 52 to 364, and the standard deviation was 55.8. The minimum score was 56 and the highest score was 336.6.

To make the score more intuitive, the scores were transformed to a percentage scale and inverted so that "good quality of life" is represented by a high score and "bad quality of life" is represented by a low score. The inversion and transformation are represented by the following formula:

$$\text{Percentage score} = ([364 - \text{raw score}]/312) \times 100$$

where 312 is the number of possible scores across the range of 52 to 364. The transformed overall average score is 72.5% and the standard deviation is 17.9. The minimum percentage score reported was 8.8% and the maximum percentage score was 98.7%. Thus, the instrument provided a wide range of scores.

DISCUSSION

The number of patients needing extraoral craniofacial prostheses is relatively small, and these patients are scattered around the world. Similarly, the number of providers with appropriate expertise is small and widely scattered. As a result, it is difficult to generate sufficient numbers of patients for clinical testing of new techniques and materials. Therefore, in the development of this instrument, attempts were made to minimize the effects of measurement error related to the instrument so that smaller sample sizes would be required with its use in research applications. Similarly, control of bias at every step was important. The only exception was the deliberate bias introduced in framing the questions to favor the illness state, to make room to spread the answers. The ceiling effect found in the answers to the 139 items justified this bias.

Items were eliminated from the original item pool based on the patients' responses. In some domains, there was a relatively small reduction from the number of items in the original pool (4/11, or 36%, in Fit and retention), whereas in other domains the reduction was much larger (27/30, or 90%, in Family/friends/strangers). By virtue of the changed number of items left in the finished instrument, the relative weighting of the domains in the whole instrument was changed. Since there was no basis for it, no attempt was made to weight the domains in the original pool. On the other hand, since the patients were allowed to rate the relevance of each of the items, the number of items remaining is empirical evidence for the patient-based weighting of the domains.

The reduction in the number of items reduces the respondent burden, but it raises the question of whether some whole domains should be eliminated. In addition to the basis provided above, Moran and coworkers[39] suggested that domains with as few as 2 items be retained to preserve reliability and responsiveness at acceptable levels. This will also minimize the impact of idiosyncratic responses when the finished instrument is in use. Therefore, with these issues in mind, no domains were eliminated.

The Cronbach's alpha measure of internal consistency on the final 52 items (unmodified for missing data) was 0.967. This very high internal consistency is well above the suggested standard of 0.8 for comparing groups in clinical trials and is even above the "desirable standard" of 0.95 for using this instrument to compare individuals against norms.[40] It could be argued that the internal consistency would be spuriously inflated by imputing the patient's individual domain means for missing data. However, a comparison of the Cronbach's alpha where no changes were made for missing data (α = 0.967174) with the same measure where domain means were imputed (α = 0.967526) reveals that there was virtually no meaningful inflation of the internal consistency.

Where the data were analyzed by national group, the lack of difference in each domain among the groups (with 1 exception) was interpreted in cultural terms. Other factors may have contributed to this result, however, including statistical issues such as sample size. On the other hand, since all the respondents were treated patients, the surgical and prosthetic management could have been substantially different among the centers, which might have caused a spurious interpretation of cultural difference if such a difference were found. The fact that data came from more than 1 center in both Canada and the United Kingdom helped to diminish this effect. Since it is unlikely that the management protocols at each center are identical, it is therefore easier to interpret lack of difference among the national groups in cultural terms.

Many of the items in the original item pool were gathered from treated patients. Many items therefore refer to a prosthesis in the subject matter of the item. However, of the 52 remaining items, 29 do not make reference to a prosthesis. These items, which are consistent with each other (α = 0.976) and relevant to treated patients, and which provide variance in their answer performance, might be expected to perform similarly among members of the same population who are awaiting treatment. The performance of these items among pretreatment patients is unknown. However, their behavior in a similar population of patients (different only in that they have been treated), suggests that these would be good items to test for use in trials where within-patient change is the outcome of interest.

CONCLUSION

A patient-based outcome measure of condition-specific quality of life has been developed with control of bias and demonstrated performance characteristics. Evidence for the relevance to patients, variation in answer options, internal consistency, cultural agreement, and minimized redundancy has been presented as strategies to minimize measurement error related to the instrument.

ACKNOWLEDGMENTS

This work obviously required the cooperation and coordination of many people. We are particularly indebted to the principal investigators at the various centers. Dr Roger Watson provided accommodation and logistic support in the United Kingdom, as well as early suggestions and arrangements for ethical review in London. Dr Patricia Finlay in Glasgow, Dr Johan Wolfaardt in Edmonton, and Dr Steve Eckert in Rochester, Minnesota, all graciously took on the task of obtaining local ethical approval and supervised the data collection in their centers. The willingness of all these people was most generous, for which we are extremely grateful. Finally, we are grateful to Mr Marco Katic, who provided a considerable volume of statistical information and volunteered his time to provide assistance with its interpretation. We are also grateful to the patients, who took considerable time to complete the questionnaires with no expectation of compensation.

REFERENCES

1. Roefs AJM, Van Oort RP, Schaub RMH. Factors related to the acceptance of facial prostheses. J Prosthet Dent 1984;52:849–852.
2. Sloan JA, Tolman DE, Anderson JD, Sugar AW, Wolfaardt JF, Novotny P. Patients with reconstruction of craniofacial or intraoral defects: Development of instruments to measure quality of life. Int J Oral Maxillofac Implants 2001;16:225–245.
3. Jani RM, Schaaf NG. An evaluation of facial prostheses. J Prosthet Dent 1978;39:546–550.
4. Sela M, Lowental U. Therapeutic effects of maxillofacial prostheses. Oral Surg Oral Med Oral Pathol 1980;50:13–16.
5. Lowental U, Sela M. Evaluating cosmetic results in maxillofacial prosthetics. J Prosthet Dent 1982;48:567–570.
6. Tolman DE, Taylor PF. Bone-anchored craniofacial prosthesis study. Int J Oral Maxillofac Implants 1996;11:159–168.

7. Arcuri MR, LaVelle WE, Fyler A, Funk G. Effects of implant anchorage on midface prostheses. J Prosthet Dent 1997;78:496–500.
8. Nordlicht S. Facial disfigurement and psychiatric sequelae. NY State J Med 1979:1382–1384.
9. Chen M-S, Udagama A, Drane JB. Evaluation of facial prostheses for head and neck cancer patients. J Prosthet Dent 1981;46:538–544.
10. Curtis TA, Beumer J. Sexuality and head and neck cancer. In: Vaeth JM (ed). Body Image, Self-Esteem, and Sexuality in Cancer Patients, ed 2. Basel: Karger, 1986:30–40.
11. Robinson E, Rumsey N, Partridge J. An evaluation of the impact of social interaction skills training for facially disfigured people. Br J Plast Surg 1996;49:281–289.
12. Heimberg RG, Hope DA, Rapee RM, Bruch MA. The validity of the Social Avoidance and Distress Scale and the Fear of Negative Evaluation Scale with social phobic patients. Behav Res Ther 1988;26:407–413.
13. Kornblith AB, Zlotolow IM, Gooen J, et al. Quality of life of maxillectomy patients using an obturator prosthesis. Head Neck 1996;18:323–334.
14. McDonough EM, Varvares MA, Dunphy FR, Dunleavy T, Dunphy CH, Boyd JH. Changes in quality-of-life scores in a population of patients treated for squamous cell carcinoma of the head and neck. Head Neck 1996;18:487–493.
15. Kapasi A, Anderson JD. Quality of life with implant-retained facial prostheses. A pilot study. In: Beumer J (ed). Proceedings of the First International Congress on Maxillofacial Prosthetics. Palm Springs, CA 1994.
16. Sykes BE, Curtis TA, Cantor R. Psychosocial aspects of maxillofacial rehabilitation. Part II. A long-range evaluation. J Prosthet Dent 1972;28:540–545.
17. Rozen RD, Ordway DE, Curtis TA, Cantor R. Psychosocial aspects of maxillofacial rehabilitation. Part I. The effect of primary cancer treatment. J Prosthet Dent 1972;28:423–428.
18. Bjordal K, Ahiner-Elmqvist M, Tollesson E, et al. Development of a European Organization for Research and Treatment of Cancer (EORTC) questionnaire module to be used in quality of life assessments in head and neck cancer patients. EORTC Quality of Life Study Group. Acta Oncol 1994;33:879–885.
19. Hassan SJ, Weymuller EA Jr. Assessment of quality of life in head and neck cancer patients. Head Neck 1993;15:485–496.
20. Browman GP, Levine MN, Hodson Dl, et al. The Head and Neck Radiotherapy Questionnaire: A morbidity/quality-of-life instrument for clinical trials of radiation therapy in locally advanced head and neck cancer. J Clin Oncol 1993;11:863–872.
21. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. J Clin Oncol 1993;11:570–579.
22. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365–376.
23. Spitzer WO, Dobson AJ, Hall J, et al. Measuring the quality of life of cancer patients: A concise QL-index for use by physicians. J Chronic Dis 1981;34:585–597.
24. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–370.
25. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and final revision of a health status measure. Med Care 1981;19:787–805.
26. Watson D, Friend R. Measurement of social-evaluative anxiety. J Consulting Clin Psychol 1969;33:448–457.
27. Beck AT, Beck RW. Screening depressed patients in family practice. A rapid technic. Postgrad Med 1972;52:81–85.
28. Doak C, Doak L. Teaching Patients with Low Literary Skills, ed 2. Philadelphia: Lippincott, 1996.
29. Schriesheim CA, Hill KD. Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. Educ Psychol Meas 1981;41:1101–1114.
30. Holden RR, Fekken G, Jackson D. Structured personality test item characteristics and validity. J Res Personality 1985;19:386–394.
31. Edwards A. The Social Desirability Variable in Personality Assessments and Research. New York: Dryden, 1957.
32. Guilford J. Psychometric Methods. New York: McGraw-Hill, 1954.
33. Streiner DL, Norman GR. Health Measurement Scales. A Practical Guide to their Development and Use, ed 2. Oxford: Oxford University Press, 1995.
34. Couch A, Keniston K. Yeasayers and naysayers: Agreeing response set as a personality variable. J Abnormal Social Psychol 1960;60:151–174.
35. Thorndike E. A constant error in psychological ratings. J Appl Psychol 1920;4:25–29.
36. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. Can Med Assoc J 1986;134:889–895.
37. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
38. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF36).1. Conceptual framework and item selection. Med Care 1992;30:473–483.
39. Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. J Clin Epidemiol 2001;54:571–579.
40. Nunnally J Jr, Bernstein IH. Psychometric Theory, ed 3. New York: McGraw-Hill, 1994.