Orthodontics & Craniofacial Research

ORIGINAL ARTICLE

A. T. Shelton
N. Y. Houghton
D. O. Morris
G. L. Latchford
H. L. Bekker
T. Munyombwe

The development and validation of a psychological questionnaire for patients undergoing orthognathic treatment

Authors' affiliations:
A. T. Shelton, Department of Orthodontics, Montagu Hospital, Mexborough, UK
A. T. Shelton, D. O. Morris, Department of Orthodontics, Leeds Dental Institute, Leeds, UK
N. Y. Houghton, Department of Orthodontics, St Lukes Hospital, Bradford, UK
G. L. Latchford, Leeds Institute of Health Sciences, Leeds, UK
H. L. Bekker, Department of Behavioural Sciences, University of Leeds, Leeds, UK
T. Munyombwe, Department of Biostatistics, University of Leeds, Leeds, UK

Correspondence to:
A. T. Shelton
Department of Orthodontics,
Montagu Hospital, Adwick Road,
Mexborough, Doncaster, S64 0AZ
E-mail: andrewshelton22@yahoo.co.uk

Date:
Accepted 12 October 2014

DOI: 10.1111/ocr.12061

Shelton A. T., Houghton N. Y., Morris D. O., Latchford G. L., Bekker H. L., Munyombwe T. The development and validation of a psychological questionnaire for patients undergoing orthognathic treatment
Orthod Craniofac Res 2015; **18**: 51–64. © 2014 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

Structured Abstract

Objectives – To develop a questionnaire to assess the psychosocial aspects which orthognathic patients considered important regarding their dento-facial deformity.

Setting and sample population – A multicentre, prospective, questionnaire development and validation study based in the UK.

Material and methods – Questionnaire development involved item (question) selection through literature review, consultation and feedback from a questionnaire development group and semi-structured interviews. A 'final' questionnaire was tested on a cross-sectional sample of 110 pre-operative and 74 post-operative orthognathic patients and a longitudinal sample of 23 orthognathic patients. Validity was tested using Rasch analysis.

Results – Reliability for the Hospital Anxiety and Depression Scale (HADS) section was unsatisfactory (ICC = 0.232–0.829, Cronbach alpha = 0.625–0.670), but for the well- being (ICC = 0.857, Cronbach alpha = 0.827–0.895) and expectations (ICC = 0.861, Cronbach alpha = 0.804–0.882) sections were satisfactory. The well-being section was the only section found to be valid for the pre-and post-operative samples. Responsiveness was satisfactory for the well-being scale ($p = 0.001$).

Conclusions – A new condition-specific orthognathic questionnaire has been developed which has been shown to be reliable, valid and responsive for the well-being scale. The HADS, as tested by Rasch analysis, was found not to be valid for this orthognathic population.

Key words: orthognathic; psychological; questionnaire

Introduction

Unlike routine orthodontic treatment, which produces minimal changes of the facial features of a patient, orthognathic treatment can result in sudden and often dramatic three-dimensional facial changes (1). The overall success of this type of combined treatment is based upon correct patient selection and diagnosis (2, 3). It is also increasingly important, that a sensitive and responsive means of quantifying psychological changes in patients is developed, to give an accurate outcome measure. A systematic review of the literature (4) found that most of the psychological evaluation techniques used in previous studies were not originally designed to evaluate orthognathic patients and many had not been properly validated.

There is currently no valid nationally or internationally agreed questionnaire capable of assessing the psychological profile of patients undergoing orthognathic treatment. The only condition-specific (dento-facial deformity) questionnaire for patients undergoing orthognathic treatment assesses 'quality of life' (5, 6). It is likely that there would be overlap between a quality of life questionnaire and a psychological questionnaire. Quality of life is concerned with how a person perceives aspects of their life and how any illness impacts on aspects of their life that are important to them. Psychology, on the other hand, includes aspects of mental illness such as anxiety and depression. It is underpinned by the theory of why people behave in the way they do, for example perception of risk, attitude, beliefs, social norms and reasoning.

There is a plethora of studies in the literature (1, 7–9) that have attempted to assess the psychological impact of orthognathic treatment. These studies however, have a potential weakness that the instruments used (questionnaires), have not been properly validated on an orthognathic population. There is evidence (10) to suggest that condition-specific approaches are more sensitive at detecting changes in the orthognathic population. Hunt et al. (5). concluded that it was important to achieve greater consistency with the techniques used to evaluate psychosocial status in orthognathic patients.

Therefore the aims of this study were

- To develop a questionnaire to identify and quantify the psychological aspects which orthognathic patients considered important regarding their dento-facial deformity. It is also hoped the questionnaire will act as a 'screening tool' to aid in psychological referral.
- To test the reliability, validity and responsiveness of the final questionnaire in both a cross-sectional and a longitudinal prospective manner.

Material and methods

The study design is shown in Fig. 1. Ethical approval and Research and Development approval were obtained (National Research Ethics Service) for the questionnaire development (03/114) and the testing of the questionnaire (08/H1302/11). Informed written consent was obtained from every subject who participated in any part of the study. For the questionnaire development, the following inclusion criteria were followed:

- Patients with a dento-facial deformity who had attended the joint orthodontic-orthognathic multidisciplinary clinic. This group included patients who were seeking orthognathic treatment and were in the process of or had completed orthognathic treatment.
- Patients over 16 years of age
- Patients who spoke English as their first language
- Patients without any congenital deformities, for example cleft lip/palate or cranio-facial syndromes
- Patients without pre-existing body dysmorphic syndrome

The patients were selected via their attendance at the joint clinic. Consecutive patients who had attended the joint clinic were invited to participate in the semi-structured interviews during the timescale (6 months) of data

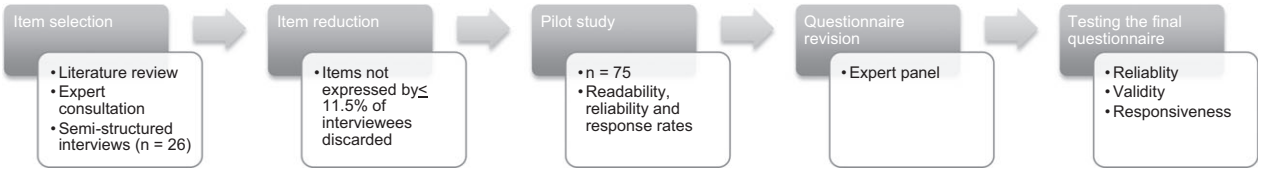

Fig. 1. Study design.

collection. This system was used to maximize the sample size.

Questionnaire development

Item selection – literature review

An item (question) pool was generated from a thorough review of the literature, consultation and feedback from the questionnaire development group (two consultant orthodontists and two consultant clinical psychologists) and informal interviews with 30 patients (attending a joint surgical-orthodontic clinic). The literature review search strategy utilized Medline, Embase, PsychINFO and Cochrane databases. The following keywords were used in various combinations: Orthognath, *Surg*, Orthodont, *Psycholog*, Osteotomy, Le Fort, Maxilla, *Mandib*, Advancement, Set-back, Impaction, Questionnaires, Self esteem, Anxiety, Depression and Self concept. There were no date restrictions, but only English language articles were used.

Item selection – semi-structured interviews

The interviewed group consisted of 15 pre-surgical and 15 post-surgical patients, at various stages of their treatment. Post-surgical patients were at least 3 months post- operation. This was to avoid potential bias in responses from the estimated small proportion of patients who suffer from depressive episodes in the short period following surgery (11, 12). The qualitative interviews were carried out by author NH (orthodontic background) and were designed following guidance from a clinical psychologist. All interviews were undertaken away from the clinical area in a private room. From these interviews, various themes surrounding the psychological aspects of orthognathic surgery were identified and used to form the basis of the initial questionnaire. The patients were asked details about their dento-facial deformity and how it affected their life. They were asked about other people's perception of their problem and why they were seeking treatment now. The problems of deformity were discussed in terms of functional, aesthetic and psychological factors, including depression and the effect on self-esteem and self-confidence. Some patients brought up the topic of teasing and how this had affected their lives. The support network at home was discussed along with the impact of surgery (both actual and anticipated) upon their lives.

The techniques of 'fair dealing' and 'reflexivity' were used in this study. Fair dealing (28) involves a 'commitment to even handedness' by the researchers. In this study, this was achieved using patients at different stages of their treatment. Reflexivity involves the researcher being aware of the potential to influence the response to questions by their behaviour and the type of questions asked. All the questions asked were non-leading. Repeated topics were used to produce a list of 16 items. The information formed the basis of an initial questionnaire that was then tested on 26 patients. Thirteen of these patients had not yet commenced surgical-orthodontic treatment, and 13 had completed treatment and were attending a review appointment. The majority of the questions were 'open-ended' to allow patients to generate a number of responses. This produced a list of 16 questions aimed at reflecting the psychological impact of dento-facial deformity.

Item reduction

Item reduction evaluated patient's responses to the initial questionnaire. Comments occurring with a high frequency were included in the questionnaire, whereas low-frequency responses were removed.

Items which were selected by ≤11.5% of the respondents (<3 out of 26) were excluded. This

figure was decided by the team of investigators as it was considered more inclusive than previously reported item reduction rates of 20%, ensuring broader item inclusion. This produced a list of statements which were used in the initial questionnaire.

Pilot study

The pilot questionnaire was handed out at appointment visits in the orthodontic department at Leeds Dental Institute, as well as posted to patients whose details were on the surgical-orthodontic patient database. All patients were supplied with a patient information leaflet, a consent form to sign and a stamped addressed envelope to return the questionnaire. Patients who did not respond were contacted via telephone and asked whether they had received a questionnaire or whether they wanted a new one posted out. A total of 110 questionnaires were distributed (50 pre-surgical and 60 post-surgical patients) and 75 were returned (39 pre-surgical and 36 post-surgical patients). Two forms of the questionnaire were developed: A pre-surgical and post-surgical questionnaire with grammatical alterations for the two separate patient groups.

Analysis of the initial questionnaire

Readability and reliability

Readability was tested using the Flesch scoring system (13). Reading ease and Flesch–Kincaid grade level tests were applied to the pre- and post-surgical questionnaires using Microsoft Word (version Microsoft Office XP Professional).

For the initial study, only the test–retest method was used and was carried out by telephone 'interviews' on a random sample of 30 patients (15 pre-operative and 15 post-operative patients). The patients were contacted, by telephone, 2 weeks following the completion of the initial questionnaire to reduce the 'practice' effect (14). Questions from the questionnaire were repeated verbatim and responses recorded. A kappa statistic was calculated using SPSS (statistics package, version 14, for Windows, SPSS Inc., Chicago, IL, USA, 1989–2006).

Questionnaire revision

During the process of questionnaire revision, all final decisions were made following discussion at regular meetings with the author, one clinical psychologist, one clinical psychiatrist and a consultant orthodontist.

Questionnaire testing

Sample size calculation

As this study attempted to test its reliability, validity and responsiveness, it still has to be considered as an improved 'pilot' project. For this reason, and following statistical advice, a sample size calculation was not required and the research team aimed to obtain the largest possible sample over the data collection period, which lasted from May 2008 to May 2009.

Inclusion criteria

To reduce potential bias when testing the questionnaire, the following changes were made to the inclusion criteria:

- Patients with dento-facial deformity, who had already started orthognathic treatment in the Orthodontic Departments at Leeds Dental Institute or Seacroft Hospital in Leeds and St. Luke's Hospital in Bradford. For this study, the start of orthognathic treatment was defined as 'the date of placing the first active fixed appliance'.
- Post-operative patients more than three months following definitive orthognathic surgery.

Questionnaire distribution

Potential participants were selected from clinic day lists prior to their actual appointment and were sent a patient information sheet. Questionnaires were collected, where possible, prior to the clinical appointment to improve the response rate. To ensure the data collection process was anonymous, an independent healthcare worker received all completed questionnaires. When a questionnaire was given out, this was recorded, along with the patient's hospital number, on a separate data sheet. This process

ensured anonymity of the actual questionnaire but also enabled the overall response rate to be calculated. All subjects who refused to participate were recorded.

Testing the reliability

The reliability of the questionnaire was tested by both the test–retest method and the internal consistency method. Patients were asked to complete the same questionnaire within a 4 to 6-week period in the clinical setting. This sample was dictated by patients who returned within this time period as part of the normal treatment process.

Testing the validity

Rasch analysis was used to test the validity of the different sections of the questionnaire. Rasch analysis fundamentally assesses a questionnaire for unidimensionality (is the questionnaire measuring the same construct/subject area e.g. well-being), and if the measure is not measuring the same construct/subject, then it will not fit the statistical model. It can be seen that it is an excellent way to test for metric (ability to measure) validity and unidimensionality (15). It is probability-based and allows the raw scores to be converted to scores that can be used in statistical tests. If the correct processes have been followed in questionnaire construction (see part 1), and the questionnaire has content validity (questions are a well-balanced sample of the subject area to be measured), the Rasch analysis should show whether it also has construct validity (how well the test is measuring the underlying concept it purports to measure). In other words, if the questions are all thought to be about well-being, and analysis proves unidimensionality, then it is highly likely that the questions are about well-being. Construct validity can be tested using a similar measure (questionnaire), with a prediction assumption that related variables will correlate (convergent validity) and dissimilar variables will not (divergent). This type of validity is the one most popularly tested for in the literature (6, 16, 17). Despite this, if a new questionnaire is unique, the use of this test is limited, as a high correlation would suggest that the 'same questions are being asked' and therefore the questionnaire would not be unique after all. Rasch analysis can be seen to be the most appropriate and statistically robust method to assess validity when the questionnaire has scaled responses and is reported to be testing one construct/subject area (18).

Testing the responsiveness

Within the data collection time period, all patients who had completed a pre-operative questionnaire and were greater than three months following surgery completed a post-operative questionnaire. This resulted in a longitudinal sample. The research team aimed to obtain the largest possible sample over the data collection period.

Statistical analysis

The ages of the sample were calculated using Microsoft Excel for Mac 2004. Rasch analysis was completed using RUMM 2020 (2003) for Windows. All other statistical analysis was completed using SPSS (statistics package, version 15, for Windows, SPSS Inc., Chicago, IL, USA, 1989–2009) for Windows.

Results

Questionnaire development

26 of the 27 patients approached to take part in the semi-structured interviews agreed, giving a response rate of 96%.

The initial questionnaire study resulted in an overall response rate of 68% (78% pre-surgical and 60% post-surgical). The gender distribution of the sample was approximately ⅔ (68%) female and ⅓ (32%) male. The mean age of the patients in the initial study was 26.1 years (16.0 – 47.0, SD 8.1). The ethnic distribution of the sample is shown in Table 1 with the majority of the sample being white.

Table 1. **Ethnic distribution of the pilot and test samples**

	Pre-operative pilot (%)	Pre-operative test (%)	Post-operative pilot (%)	Post-operative test (%)	Total pilot (%)	Total test (%)
White	33 (84.6)	87 (79.1)	31 (86.1)	61 (82.4)	64 (85.3)	148 (80.4)
Black/African	1 (2.6)	1 (0.9)	0	1 (1.4)	1 (1.3)	2 (1.1)
Black/other	0	1 (0.9)	0	0	0	1 (0.5)
Indian	1 (2.6)	12 (10.9)	3 (8.3)	5 (6.8)	4 (5.3)	17 (9.2)
Pakistani	2 (5.1)	8 (7.3)	2 (5.6)	7 (9.5)	4 (5.3)	15 (8.2)
Bangladeshi	0	0	0	0	0	0
Chinese	2 (5.1)	1 (0.9)	0	0	2 (2.7)	1 (0.5)
Other	0	0	0	0	0	0

The Flesch readability score for the pre-operative questionnaire was 82.4 and 82.2 for the post-operative questionnaire. Both questionnaires had Flesch–Kincaid grade level of 3.2. This categorizes the questionnaires as having 'good' readability with an 'easy' level of difficulty and indicates that 86% of the UK population would understand them (16).

The overall reliability of the questionnaires was found to be 'very good' (19) with a kappa score of 0.81 (0.60–1). The response rates to the closed questions in both questionnaires ranged from 91 to 100%, with the open questions ranging from 24 to 45%. Looking in more detail, 77% of the sample had been 'teased about their appearance', but only 17% admitted that they 'still got teased'. Of the people who still got teased around ⅔ said that they 'coped with it' and ⅓ said it made them feel 'depressed, upset or angry'. The responses to the question 'How is your life different / how do you think your life will be different following treatment?' are shown in Table 2 and highlight that the aspect of confidence is an important factor to orthognathic patients.

Questionnaire revision

After discussing the relevant literature, results from the initial study and aims of the questionnaire, two 'final' revised questionnaires (pre-operative, supplementary file 1, and post-operative, supplementary file 2) were completed fulfilling the following criteria:

- To be clear and have structured sections involving a scaled response, allowing statistical analysis to be performed.
- The structured sections should

1. enable accurate screening of patients in need of psychiatric assistance
2. be amenable to a scoring system to allow ease of use in the clinical setting and aid future research
3. attempt to explore different psychological aspects of the surgical-orthodontic 'journey'
4. enable repeated administration to stable patients to yield similar results (i.e. good reliability)
5. be sensitive enough to reflect small, but clinically important, changes (i.e. good responsiveness)
6. be valid
7. be relatively short and simple to administer

Table 2. **Responses to the question 'How is your life different/how do you think your life will be different following treatment?' during semi-structured interviews**

Response	Pre-op (%)	Post-op (%)
'I would have more confidence'	53.8	63.9
'I would have been teased/ bullied less'	12.8	5.6
'I would be looked on the same as others'	2.6	13.9
'Life would be no different'	15.4	8.3
'Other'	2.6	8.3

With permission, a few specific and particularly relevant questions from the only validated condition-specific questionnaire, exploring psychological domains (5, 6), were integrated in to the new questionnaire. In particular, questions relating to avoiding social situations were included. The robustly tested Hospital and Anxiety Scale (HADS) (20) was included alongside the newly developed questionnaire, as it was felt that this would act as a useful screening tool, to aid in the diagnosis of patients with severe anxiety or depression. These patients could then be offered extra psychological support via a psychological referral prior to commencing orthognathic treatment. The resulting final questionnaires had the following sections.

- HADS
- Well-being
- Expectations

Questionnaire testing

Sample demographics

110 patients completed the pre-operative questionnaire and 2 people refused giving a response rate of 98.2%, whilst 72 patients completed the post-operative questionnaire, and 1 refused, giving a response rate of 98.7%. The overall response rate was 98.4%. The reliability sample consisted of 23 patients who completed the same questionnaire within a 4 to 6-week period, with 10 completing pre-operative questionnaires and 13 completing post-operative questionnaires. The longitudinal reliability sample consisted of 23 patients.

The gender distribution of the sample was similar to the pilot study with approximately 2/3 (70%) being female and 1/3 (30%) being male. The mean age of the pre-operative sample was 25.4 years with a range of 16.1 - 56.4 years (standard deviation of 9.4 years). The mean average of the post-operative sample was 27.4 years with a range of 17.4 – 45.7 years (standard deviation of 7.5 years). The ethnicity of the sample is shown in Table 1. The majority of the sample (80.4%) was made up of white patients, with 17.4% being Indian/Pakistani patients.

Reliability

Test–retest

Both the well-being and the expectation sections showed very good levels of agreement, illustrating acceptable reliability for the measures (Table 3). A correlation of >0.80 indicates that the scale is highly reliable (21). The ICC for the anxiety section of HADS was comparable to both sections of the orthognathic condition-specific questionnaire, although the reliability of the depression section was poor.

Internal consistency

The Cronbach alpha values of all the sections of both the pre-operative and the post-operative questionnaire show correlations that were acceptable (Table 4). The internal consistency of both constructs of the HADS was found to be less than acceptable.

Validity

Hospital anxiety and depression scale

As the HADS was the same in both the pre- and post-operative questionnaire, the Rasch analysis includes the combined data from both samples.

The chi-square probability (Table 4) indicates that for both the anxiety and the depression sections, the data did not fit the Rasch statistical model, as the lack of fit was significant ($p < 0.001$). This means that the questions were not leading to responses that would be expected from patients at different grouped levels of mental state, that is, very anxious/depressed, less anxious/depressed and not anxious/depressed.

Table 3. **Test-rest reliability of the sections of the 'final' orthognathic item-specific questionnaire (n = 23)**

'Section' of the questionnaire	Intra-class correlation coefficient (ICC)	95% confidence interval	
		Lower bound	Upper bound
Well-being	0.857	0.688	0.939
Expectations	0.861	0.676	0.941
HADS - anxiety	0.829	0.603	0.927
HADS - depression	0.232	−0.785	0.673

Table 4. **Internal consistency of the sections of the 'final' orthognathic item-specific questionnaire (n = 23)**

Questionnaire	Section of the questionnaire	Cronbach alpha
Pre-operative	Well being	0.895
Pre-operative	Expectations	0.882
Post-operative	Well being	0.827
Post-operative	Expectations	0.804
Pre and post-operative	HADS-anxiety	0.625
Pre and post-operative	HADS-depression	0.670

The power of test-of-fit is a representation of the person separation index and is indicative of the power of the questionnaire to discriminate between the respondents. This indicates that all-questions may have been well answered in a similar way, and the anxiety/depression questions did not lead to discrimination between the different orthognathic patients that have answered the questions. It may be that the anxiety questions were not 'sensitive' enough for the orthognathic population. The poor fit to the Rasch model meant that further Rasch analysis was not beneficial.

Well-being

Pre-operative

Table 5 shows that the chi-square probability was 0.235, indicating that the data fit the model and the excellent power of test-of-fit shows that the questionnaire discriminates between respondents. Rasch analysis allows each individual question to be analysed. For each question, a chi-squared fit statistic and a fit residual value, which approximates to a z-standardized normal distribution, are produced. Any question with a fit residual +/- 2.5 is a cause for concern and is a source of misfit to the model. The Rasch analysis will give a category probability curve for each question (Fig. 2). The probability curve 0 corresponds to the 'strongly agree' response, 1 to 'agree', 2 to 'disagree' and 3 to 'strongly disagree'. The *x*-axis is person location or logits and corresponds to how the question is being answered by the respondents. Ideally, there should be a separate peak for each response category, as seen in Fig. 2, which shows that all response categories are valid and none are redundant. Thresholds that do not increase monotonically across the rating scale are considered disordered. Each category should have a distinct peak in the probability curve. Questions 'I find it difficult making friends' and 'I avoid shopping when it's busy' were found to have disordered thresholds. An example of a disordered threshold is shown in Fig. 3. Each question, through the Rasch analysis, is given a logit score based on a z-distribution anchored on 0. Half the questions will have positive logit scores and half will have negative ones. The logits from the 'positive subset of questions' were then compared to the 'negative subset'. A paired, two-tailed student's *t*-test was used with a 5% significance level. $p = 0.957$ indicating no difference between the two subsets and therefore unidimensionality.

Post-operative

Chi-square probability was >0.05 (Table 5), indicating that the data fit the model and the

Table 5. **Summary statistics of the Rasch analysis of the 'final' orthognathic item-specific questionnaire**

'Section' of the questionnaire	Overall fit (item-trait interaction- chi squared prob.)	Power of test-of-fit (person separation index)	Test for unidimensionailty (paired *t*-test)
HADS - anxiety	0.000	Poor (0.141)	n/a
HADS - depression	0.000	Poor (−0.561)	n/a
Well-being – pre-operative	0.235	Excellent (0.920)	0.957 - uni
Well-being – post-operative	0.053	Excellent (0.902)	0.813 - uni
Expectations – pre-operative	0.001	Excellent (0.870)	n/a
Expectations – post-operative	0.090	Excellent (0.890)	0.807 - uni

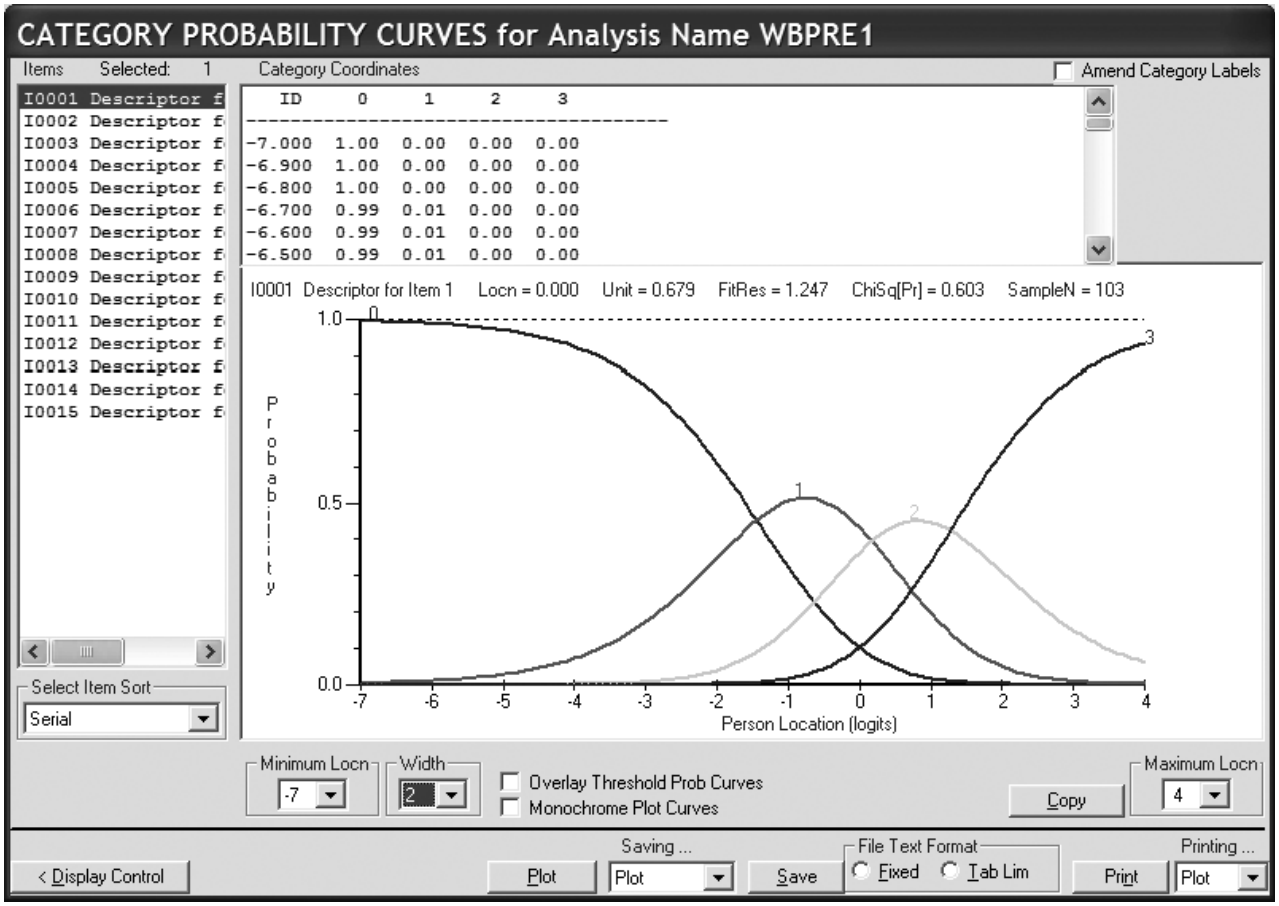

Fig. 2. Category probability curve question 1(well being), 'I get teased about my facial appearance', for the pre-operative 'final' orthognathic item-specific questionnaire.

excellent power of test-of-fit shows that the construct discriminates between respondents. Only questions 'I find it difficult making friends' and 'my facial appearance limits my ability to do the things I want to do' were found to have disordered thresholds. Table 5 indicates that the construct showed unidimensionality.

Expectations

Pre-operative

Table 5 indicates that the questions did not fit the Rasch model and therefore further Rasch analysis was not indicated.

Post-operative

Table 5 indicates that the data fit the model satisfactorily. Only the question 'I look similar to other people' was found to have a disordered threshold. The questions were found to be unidimensional.

Responsiveness

To obtain an accurate logit score for individual patients, there must be a satisfactory fit to the Rasch model. As the well-being section was the only one to fulfil this criterion for the pre- and post-operative samples, it is the only section where the responsiveness can be tested. The fact that the logit scores are based on a normal distribution means a paired student *t*-test was appropriate. As a result of numerous questions unanswered, one patient was removed from the responsiveness analysis giving a sample of 22. $p = 0.001$ indicating a responsive questionnaire. The mean post-operative logit score (2.16) was higher than the pre-operative score (0.58) indi-

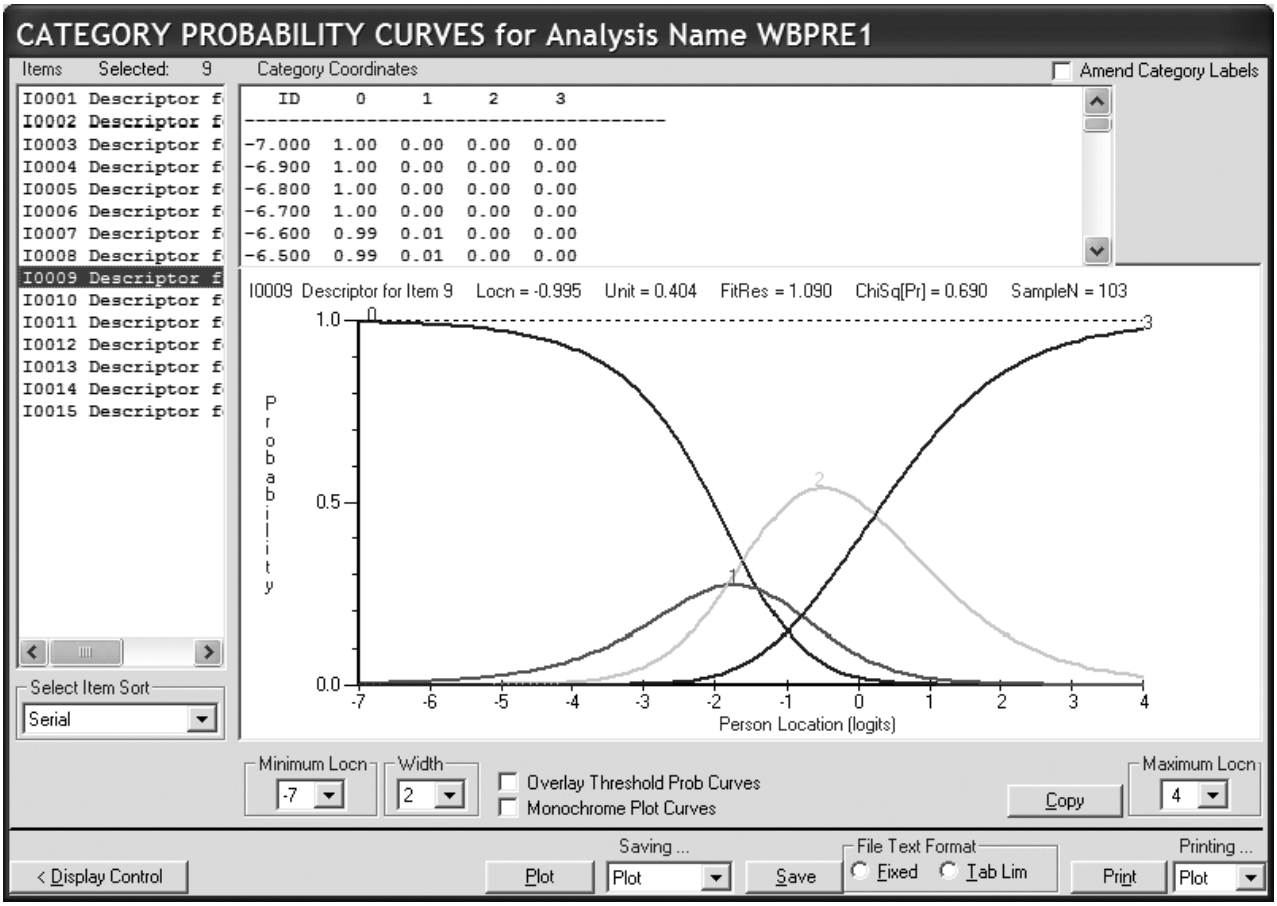

Fig. 3. Category probability curve for question 9(well-being), 'I find it difficult making friends', for the pre-operative 'final' orthognathic item-specific questionnaire.

cating a statistically significant improvement in well-being.

Discussion

The aim of the first part of this study was to develop a questionnaire to assess the psychological aspects important to patients undergoing orthognathic treatment. Although at first glance any questionnaire can be seen to be subjective, it is the process of development that will dictate the final validity of the questionnaire. The development of this questionnaire follows closely the robust methods used in the only other condition-specific orthognathic questionnaires (5, 6). The true prevalence of body dysmorphic disorder (BDD) in the orthognathic population has not been assessed to date. From the current literature, it seems that mental health professionals use structured interviews to make the diagnosis of BDD. The ability to suspect BDD and refer accordingly is key to avoiding undesirable treatment results. In view of the interview-based diagnostic tools currently used to diagnose BDD, it would not be appropriate to attempt to diagnose BDD with the use of an orthognathic psychological questionnaire. For this reason, it was decided that BDD would not be tested for with these questionnaires.

A questionnaire is a measurement tool (14), that is relatively inexpensive, familiar and acceptable to most people. It is well documented that open-ended questions lead to a decreased response rate (14), as was seen in this pilot study. It was for this reason that in the revised questionnaire, there were no open-ended questions. The initial questionnaires were found to have an 'easy' level of readability (FRE of 82.4 and 82.2 for pre- and post-operative

questionnaires respectively). The only other condition-specific questionnaire tested for readability was found to have a FRE of 61.442 indicating the initial questionnaire compares favourably.

The results indicate that as tested by Rasch analysis, the well-being section of this newly developed questionnaire was found to be reliable, valid and responsive. This means that it is appropriate for use longitudinally in orthognathic patients. Both the expectation section and the HADS were not found to stand up to the same rigorous testing.

Approximately, 70% of the sample was made up of females, which compares closely to previous studies (6, 22). It also accurately reflects the commonly found demographics of patients seeking orthognathic treatment (23). Although there was a mix of different races in both the pre-operative and the post-operative samples, the majority of the samples were made up of white (Caucasian) patients (80.4%). It must be remembered that the results of this study are applicable to orthognathic populations that are a match to the demographics of the population studied.

Hospital anxiety and depression scale

A review (20), involving 747 studies, indicated the HADS to demonstrate validity and found it to perform well in assessing symptom severity and 'caseness' (i.e. its ability to determine anxiety or depression) of anxiety disorders and depression in both primary care psychiatric patients and in the general population. In this study, the HADS was found to be unsatisfactory when subjected to rigorous reliability and validity testing and is in contrast to previous thinking (20). The analysis used shows that the scale employed by the HADS does not stand up to metric testing for this population, and so individual scores do not have true statistical meaning.

Well-being and expectations

Both sections of this orthognathic condition-specific psychological questionnaire show very good levels of agreement, illustrating acceptable reliability for the measures. It compares favourably with the only other condition-specific orthognathic questionnaire (6), which had an ICC range of 0.78–0.93. A correlation of >0.80 indicates that the scale is highly reliable (21). The Cronbach alpha values (0.804–0.895) of all the sections of both the pre-operative and the post-operative questionnaires show acceptable correlations, which compare satisfactorily with the only other condition-specific questionnaire (5).

The two questions that showed disordered thresholds in the well-being section were 'I find it difficult making friends' and 'I avoid shopping when it's busy'. It has been stated that recoding for disordered thresholds is problematic when only a minimal number of questions are involved (18). As the overall fit was good, rescoring these two questions was not appropriate.

For the post-operative sample, once more, 'I find it difficult making friends' showed a disordered threshold leading to the possibility of either rescoring the question or deleting it altogether in the future. A further question, 'my facial appearance limits my ability to do the things I want to do' also showed a disordered threshold. There appears to be no consensus in the literature with regard to the appropriate course of action when a question has a disordered threshold or is not fitting the model (15, 18, 24). As the Rasch analysis is based on probability, it is a question of degree in the context of the questionnaire being tested.

The pre-operative sample data in the expectations section led to a poor fit to the Rasch model. The overall picture suggests that the pre-operative subjects are 'expecting' a great deal from surgery and, consequently, are responding in a similar way. This will lead to a poor fit to the model as the questions were not leading to a range of responses. It is highly probable that because expectations of the sample were on the whole likely to be similar (high), there was no scale in the responses and so no discrimination in responses.

The well-being section of this orthognathic condition-specific questionnaire was found to be responsive. The small sample size (n = 22) used to test the responsiveness of the questionnaire

means that inferences with regard to the changes following surgery are illustrated with this in mind. The well-being of patients has been shown to improve following surgery by the use of this reliable, valid instrument.

In comparison to the available literature, the methods used in questionnaire development, reliability testing, validity testing and responsiveness testing were completed to a high standard (6, 18, 22, 25, 26). The sample size of 184 compares favourably to previous studies (1, 7, 22, 26, 27) and might suggest a potential for increased reliability. However, it must be remembered that this is a cross-sectional sample and not a true longitudinal sample. The response rate of 98.4% compares very favourably with the previous studies (1, 5, 7, 22, 27) and is a result of the method used. The response rate seen in this study reduces the possibility of non-response bias.

There are some areas of bias, however, which must be highlighted. The patients were interviewed in order for item selection to take place. There is a risk of bias related to the type of patient who is willing to devote time to discussing their feelings. These patients tend to be either those who are very happy and want to give 'something back' or those that are unhappy and see it as a platform to discuss their concerns (28). Although this is a potential source of bias, the other processes that were followed, in questionnaire development, should eradicate 'extreme' items. The fact that there were two orthodontists and two clinical psychologists should mean that the final questionnaire should have a balance of different questions. This process would be more robust with a larger group of professionals. However, resources did not permit this. There is always an element of subjectivity in the wording and topic of the questions although again the processes followed reduced the bias in this respect.

Ideally, the patients selected for testing the reliability of the final questionnaire should have come from randomization to reduce bias. The only way that this would be possible would be to conduct a telephone or postal questionnaire. The change from a self-administrated questionnaire to a telephone questionnaire is likely to bring in significant bias and risks jeopardizing the validity of this sample. The use of a postal questionnaire is likely to increase non-response bias and again affect the validity of the sample. Recalling a patient for an appointment purely to complete the questionnaire is unlikely to be appreciated by the patient, and so the method used was thought to be the most appropriate. It is for this reason that the recall period was 4–6 weeks and not a defined time. One drawback of the RUMM 2020 software is that in questions where the patient has not given an answer, it is counted as missing data. This leads to a further decrease in the sample size for the questions unanswered. The longitudinal sample was dictated by the number of operations completed during the data collection period with the knowledge that the patient must be at least three months post-surgery. This limited the data collection period for this sample to only nine months, which led to a sample size of only 23 subjects, with one being discounted.

Clinical implications

In the current climate where outcome measures for time consuming and costly treatment are vital, the use of the well-being section would applicable. This robustly tested and sensitive tool would be a great addition to the armamentarium of outcome measures for orthognathic surgery. The demographics of this sample compare favourably to previous studies(6, 22)and therefore the external validity of this questionnaire is enhanced. It is possible, in the future, to relate the 'raw scores' from the well-being section to the logits score for each individual in order to try and indicate whether there is a linear range of scores. This would mean that 'raw scores', obtained easily in a clinical setting, would have both statistical and clinical meaning. This technique has been used previously (18, 24) but is beyond the scope of this report.

The HADS was originally placed in the questionnaire to provide a potential 'screening tool' so that clinicians would be able to 'flag-up' those patients who may need further psychological input. This study has shown that

for this orthognathic population, the HADS lacks both reliability and metric validity. It is unlikely that a single questionnaire will ever replace a clinician's full history over of a multitude of appointments, before the decision to refer to a clinical psychiatrist. In terms of well-being, however, a reliable, valid and responsive questionnaire can now be used in a longitudinal manner.

Conclusions

- Validity testing in the form of Rasch analysis indicated that the HAD anxiety scale, HAD depression scale and the expectation section were not valid for this orthognathic population.
- The 15-point well-being section of this condition-specific questionnaire has been found to be reliable, valid and responsive for the orthognathic population.

Clinical relevance

There is currently a shortage of condition-specific, robustly tested, psychological orthognathic questionnaires. This study presents the development of a new psychological questionnaire that will enable the clinician to obtain a valid outcome measure that can be used in a longitudinal manner. This questionnaire sets out to enable the clinician to quantify well-being changes and aid audit, research and cost justification for current treatments. It is also hoped that this questionnaire will be a useful aid in the decision to make any psychological/psychiatric referrals as part of the treatment pathway.

References

1. Kiyak HA, West RA, Hohl T, Mcneill RW. The psychological impact of orthognathic surgery: a 9-month follow-up. *Am J Orthod* 1982;81:404–12.
2. Cunningham SJ. The psychology of facial appearance. *Dent Update* 1999;26:438–43.
3. Proffit WR, Fields HW, Sarver DM. *Contemporary Orthodontics*. St. Louis, Missouri: Mosby Elsevier; 2007.
4. Hunt OT, Johnston CD, Hepper PG, Burden DJ. The psychosocial impact of orthognathic surgery: a systematic review. *Am J Orthod Dentofacial Orthop* 2001;120:490–7.
5. Cunningham SJ, Garratt AM, Hunt NP. Development of a condition-specific quality of life measure for patients with dentofacial deformity: I. Reliability of the instrument. *Community Dent Oral Epidemiol* 2000;28:195–201.
6. Cunningham SJ, Garratt AM, Hunt NP. Development of a condition-specific quality of life measure for patients with dentofacial deformity: II. Validity and responsiveness testing. *Community Dent Oral Epidemiol* 2002;30:81–90.
7. Kiyak HA, Hohl T, West RA, Mcneill RW. Psychologic changes in orthognathic surgery patients: a 24-month follow up. *J Oral Maxillofac Surg* 1984;42:506–12.
8. Motegi E, Hatch JP, Rugh JD, Yamaguchi H. Health-related quality of life and psychosocial function 5 years after orthognathic surgery. *Am J Orthod Dentofacial Orthop* 2003;124:138–43.
9. Rispoli A, Acocella A, Pavone I, Tedesco A, Giacomelli E, Ortiz L et al. Psychoemotional assessment changes in patients treated with orthognathic surgery: pre- and post-surgery report. *World J Orthod* 2004;5:48–53.
10. Lee S, Mcgrath C, Samman N. Impact of orthognathic surgery on quality of life. *J Oral Maxillofac Surg* 2008;66:1194–9.
11. Stewart TD, Sexton J. Depression: a possible complication of orthognathic surgery. *J Oral Maxillofac Surg* 1987;45:847–51.
12. Cunningham SJ, Crean SJ, Hunt NP, Harris M. Preparation, perceptions, and problems: a long-term follow-up study of orthognathic surgery. *Int J Adult Orthodon Orthognath Surg* 1996;11:41–7.
13. Flesch R. A new readability yardstick. *J Appl Psychol* 1948;38:221–33.
14. Oppenheim AN. *Questionnaire Design, Interviewing and Attitude Measurement*. London: Continuum; 1992.
15. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med* 2003;35:105–15.
16. Herrmann C. International experiences with the hospital anxiety and depression scale–a review of validation data and clinical results. *J Psychosom Res* 1997;42:17–41.
17. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck depression inventoy: 25 years of evaluation. *Clin Psychol Rev* 1988;8:77–100.
18. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46:1–18.
19. Landis JR, Kock GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
20. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002;52:69–77.
21. Bowling A. *Research Methods in Health. Investigating Health and Health Services*. New York: Open University Press; 2002.
22. Ryan FS, Shute J, Cunningham SJ. A qualitative study of orthognathic

patients' perceptions of referral to a mental health professional: part 2–a questionnaire study. *J Orthod* 2009;36:93–102.
23. Kiyak HA, Mnneill RW, West RA, Hohl T, Bucher F, Sherrick P. Predicting psychologic responses to orthognathic surgery. *J Oral Maxillofac Surg* 1982;40:150–5.
24. Wolfe F, Kong SX. Rasch analysis of the Western Ontario MacMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. *Ann Rheum Dis* 1999;58:563–8.
25. Sayers MS, Newton JT. Patients' expectations of orthodontic treatment: part 1 – development of a questionnaire. *J Orthod* 2006;33:258–69.
26. Travess HC, Newton JT, Sandy JR, Williams AC. The development of a patient-centred measure of the process and outcome of combined orthodontic and orthognathic treatment. *J Orthod* 2004;31:220–34.
27. Cunningham SJ, Hunt NP, Feinman C. Perceptions of outcome following orthognathic surgery. *Br J Oral Maxillofac Surg* 1996;34:210–3.
28. Mays N, Pope C. Assessing quality in qualitative research. *Br Med J* 2000;320:50–2.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary file 1. Surgical-orthodontic patient questionnaire (pre-op quesyionnaire).
Supplementary file 2. Surgical-orthodontic patient questionnaire (post-op questionnaire).