Received: 5 May 2022 | Revised: 25 July 2022 | Accepted: 6 August 2022

DOI: 10.1111/ocr.12600

RESEARCH ARTICLE

Orthodontics & Craniofacial Research WILEY

Subjective and objective analysis of orthodontic expert consensus on the assessment of orthodontic treatment outcomes

Huanhuan Chen[1,2] | Guangying Song[1,2] | Weiran Li[1,2] | Ruoping Jiang[1,2] | Xiaoyun Zhang[1,2] | Si Chen[1,2] | Gui Chen[1,2] | Siqi Liu[3] | Fanfan Dai[4] | Fei Teng[1,2] | Bing Han[1,2] | Tianmin Xu[1,2]

[1]Department of Orthodontics, Cranial-Facial Growth and Development Center, Peking University School and Hospital of Stomatology, Beijing, China

[2]National Center of Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Laboratory for Digital and Material Technology of Stomatology & Beijing Key Laboratory for Digital Stomatology & Research Center of Engineering and Technology for Computerized Dentistry Ministry of Health, Beijing, China

[3]First Clinical Division, Peking University

Abstract

Objective: The objective of the study was to explore and validate the consensus of orthodontic experts on the assessment of orthodontic treatment outcomes based on subjective and objective analysis.

Materials and Methods: The research consisted of two parts: the exploration and verification of expert consensus. First, a sample of 108 cases randomly selected from six dental schools in China were evaluated by 69 orthodontic experts and measured by researchers based on post-treatment study casts and lateral cephalograms, respectively. Then, through statistical analysis, the objective indicators significantly

Abbreviations: AB/NP(°), Anterior inferior angle between line AB (A point-B point) and line NP (nasion-pogonion); **ANB(°)**, Posterior inferior angle between line NA (nasion-A point) and line NB (nasion-B point); **AsUL-BsLL(°)**, Posterior inferior angle between line AsUL (A point of soft tissue-upper labrale) and line BsLL (B point of soft tissue-lower labrale); **AsUL-FH(°)**, Anterior superior angle of line AsUL (A point of soft tissue-upper labrale) and Frankfort plane; **Bs-B (FH)(mm)**, Perpendicular distance from the supramental to the line perpendicular to Frankfort plane through the most posterior point of mentolabial sulcus; **BsLL-FH(°)**, Anterior inferior angle between line BsLL (B point of soft tissue-lower labrale) and Frankfort plane; **Chin Thickness(mm)**, Distance between gnathion and gnathion of soft tissue; **Cm-Sn-UL(°)**, Anterior superior angle of line Cm-Sn (columella-subnasal) and line Sn-UL (subnasal-upper labrale); **FH/NP(°)**, Anterior inferior angle between Frankfort plane and line NP (nasion-pogonion); **FH/OP(°)**, Anterior inferior angle between Frankfort plane and occlusal plane; **GoGn/SN(°)**, Anterior inferior angle between line SN (sella-nasion) and line GoGn (Gonion-Gnathion); **G'-Sn-Pos(°)**, Anterior inferior superior angle of line G'-Sn (glabella-subnasal) and line Sn-Pos (subnasal-pogonion of soft tissue); **L1/AP(°)**, Anterior inferior angle between the line through long axis of lower central incisor and line AP (A point-pogonion); **L1/MP(°)**, Posterior superior angle between the line through long axis of lower central incisor and mandibular plane; **L1/NB(°)**, Anterior superior angle between the line through long axis of lower central incisor and line NB (nasion-B point); **L1-NB(mm)**, Perpendicular distance from L1 (incision inferius) to line NB (nasion-B point); **LL-Bs-Pos(°)**, Anterior inferior angle between line BsLL (B point of soft tissue-lower labrale) and line Bs-Pos (B point of soft tissue-pogonion of soft tissue); **LL-EP(mm)**, Perpendicular distance from the lower labrale to the aesthetic-line (pronasale-pogonion of soft tissue); **LL-H(mm)**, Perpendicular distance from the lower labrale to the H-line (upper labrale-pogonion of soft tissue); **MP/FH(°)**, Anterior inferior angle between mandibular plane and Frankfort plane; **MP/SN(°)**, Anterior inferior angle between mandibular plane and line SN (sella-nasion).; **NA/AP(°)**, Anterior inferior angle between line NA (nasion-A point) and line AP (A point-pogonion); **Ns-N(FH)(mm)**, Perpendicular distance from nasion to the line perpendicular to the Frankfort plane through the soft tissue; **Ns-Prn-Pos(°)**, Anterior superior angle of line Ns-Prn (nasion of soft tissue-pronasale) and line Prn-Pos (pronasale-pogonion of soft tissue); **Pg-NB(mm)**, Perpendicular distance from pogonion to line NB (nasion-B point); **SE(mm)**, Distance between Sella and the foot point from the most posterior point of the condyle to line SN (sella-nasion); **SN/OP(°)**, Anterior inferior angle between line SN (sella-nasion) and occlusal plane; **SNA(°)**, Posterior inferior angle between line SN (sella-nasion) and line NA (nasion-A point); **Sn-A(FH)(mm)**, Perpendicular distance from subspinale to the line perpendicular to Frankfort plane through the subnasale; **SNB(°)**, Posterior inferior angle between line SN (sella-nasion) and line NB (nasion-B point); **SND(°)**, Posterior inferior angle between line SN (sella-nasion) and line ND (nasion-central point of midline junction of mandible); **Sn-Prn(FH)(mm)**, Perpendicular distance from the pronasale to the line perpendicular to Frankfort plane through the subnasale; **S-Ns-Bs(°)**, Posterior inferior angle between line S-Ns (sella-nasion of soft tissue) and line Ns-Bs (nasion of soft tissue-B point of soft tissue); **S-Ns-Sn(°)**, Posterior inferior angle between line S-Ns (sella-nasion of soft tissue) and line Ns-Sn (nasion of soft tissue-subnasal); **Sn-stoms(mm)**, Distance between subnasale and stomion superius; **Stomi-Mes(mm)**, Distance between stomion inferius and menton of soft tissue; **U1/L1(°)**, Posterior inferior angle between the line through long axis of upper central incisor and the line through long axis of lower central incisor; **U1/NA(°)**, Anterior inferior angle between the line through long axis of upper central incisor and line NA (nasion-A point); **U1/PP(°)**, Posterior inferior angle between the line through long axis of upper central incisor and palatal plane; **U1/SN(°)**, Anterior inferior angle between the line through long axis of upper central incisor and line SN (sella-nasion); **U1-NA(mm)**, Perpendicular distance from U1 (incision inferius) to line NA (nasion-A point); **UL-EP(mm)**, Perpendicular distance from the upper labrale to the aesthetic-line (pronaslae-pogonion of soft tissue); **Y-axis(°)**, Anterior inferior angle between line S-Gn (sella-gnathion) to Frankfort plane; **Z-Angle(°)**, Anterior inferior angle between line Pos-UL (pogonion of soft tissue-upper labrale) and Frankfort plane.

Huanhuan Chen and Guangying Song: Co-first author.

School and Hospital of Stomatology, Beijing, China

[4]Second Clinical Division, Peking University School and Hospital of Stomatology, Beijing, China

Correspondence
Bing Han and Tianmin Xu, Department of Orthodontics, Peking University School and Hospital of Stomatology, 22 Zhongguancun South Avenue, Haidian District, Beijing 100081, China.
Email: kqbinghan@bjmu.edu.cn and tmxuortho@163.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 82071172, 51972005 and 82001082; National Natural Science Foundation of China, Grant/Award Number: 82071172, 51972005 and 82001082

correlated with experts' subjective evaluations were selected, their weights were determined, and the critical values of satisfactory, acceptable and unacceptable grades were screened. Subsequently, another sample of 72 cases were evaluated by another 36 orthodontic experts, and the subjective evaluation results were compared with the objective measurement results.

Results: There were six model indicators and seven cephalometric indicators being significantly correlated with the experts' subjective evaluations, including occlusal contact, overjet, midline, interproximal contact, alignment, occlusal relationship, L1/NB, ANB, SN/OP, U1/SN, LL-EP, Cm-Sn-UL and Ns-Prn-Pos, with a cumulative R^2 of 0.704. In the verification part, the correlation coefficient between the 36 experts' subjective scores and objective regression scores was 0.716 ($P < .001$); the correlation coefficient between the 36 experts' subjective grades and objective grades was 0.757 ($P < .001$).

Conclusions: Orthodontic experts had good consistency in the subjective evaluation of the combined records of post-treatment study casts and lateral cephalograms. The objective indicators selected from subjective and objective analysis had good reliability and validity and could further improve the existing occlusal indices.

KEYWORDS
expert consensus, objective indicators, orthodontic treatment outcomes, subjective evaluation

1 | INTRODUCTION

The orthodontic field has been committed to exploring objective assessment methods for orthodontic treatment outcomes since the 1960s, including the Peer Assessment Rating (PAR) index, American Board of Orthodontics objective grading system (ABO-OGS), index of complexity, outcome and need (ICON)[1–5] etc. The mode of establishing occlusal indices usually include three steps.[3,6] The first step is to establish the professional standard, which stems from the subjective evaluation of orthodontic treatment outcomes by experienced orthodontic experts, and the experts' subjective evaluations need to pass the consistency test. Second, the objective indicators and their weights should be agreed upon orthodontic experts or by subjective and objective analysis. The third step is to determine the critical value and clinical significance of each grade of objective measurement results. After decades of research and verification, many attempts have been made to develop methods for measuring dental aesthetics and establish standards that represent socially acceptable dental appearance.[7–10]

However, although these assessment methods serve their respective purposes well, most of them ignored the evaluation of profile harmony, which limits their application range. In addition, the original intention of some indices was not for professional orthodontic evaluation, and the wide coverage of the judges' occupations would affect the consistency of the subjective evaluation. Furthermore, the statistical process for establishing these indices is not sufficiently clear and lacks cross validation, and the methods may be too strict to exhibit general improvements after treatment, which is not conducive to simple clinical application.[9] Therefore, the purpose of this study was to explore and validate the consensus of orthodontic experts on the assessment of post-treatment dental aesthetic and profile harmony in orthodontic patients.

In orthodontic diagnosis and treatment, study casts (SC), lateral cephalograms (LX) and facial photographs (PH) are all significant clinical records.[11,12] In previous work, our research group conducted a correlation analysis of experts' subjective evaluations of SC, LX and PH records.[13,14] It was found that the SC record was the most significant predictive component in the evaluation of combined records, while the inclusion of LX and PH records also contributed to a more comprehensive evaluation. Since PH records are difficult to keep consistent and lack measurement calibration,[15,16] thus in this study, we intend to select the relevant items from SC and LX records as objective indicators to assess the orthodontic treatment outcomes.

2 | MATERIALS AND METHODS

2.1 | Source of samples

The research samples were divided into two parts in this study. The experimental sample was meant to explore an expert consensus, and the validation sample was meant to validate the reliability and validity of the expert consensus.

2.1.1 | Experimental sample

The experimental sample came from six major orthodontic treatment centres in different parts of China, including the Peking University School of Stomatology (PKUSS), the West China College of Stomatology at Sichuan University, the School of Stomatology at the Fourth Military Medical University, the Beijing Stomatological Hospital and School of Stomatology at the Capital Medical University, the Stomatological Hospital at Nanjing Medical University and the Hospital of Stomatology at Wuhan University. A total of 2383 cases completed from 2005 to 2008 were collected from the listed orthodontics departments. Using Angle's classification of molar relationship as stratification factor, 18 cases (6 Class I, 6 Class II and 6 Class III cases) were randomly selected from each dental school for a total of 108 cases. 69 experts were invited to judge the quality of orthodontic treatment outcomes of the given cases according to the combined SC + LX records.

2.1.2 | Validation sample

The validation sample of 72 cases were also randomly selected with Angle's classification of molar relationship as stratification factor, ensuring 24 cases of each Angle's classification. The difference was that the 72 validation samples consisted of 48 new cases and 24 overlapping cases. That was because the panel of 36 experts evaluating the validation sample were different from the panel of 69 experts evaluating the experimental sample, so in order to determine whether there was a difference in the two panels of experts' subjective evaluations, 24 cases of 108 experimental sample were repeatedly included to form the final validation sample. The remaining 48 new cases were randomly selected from 1323 new cases which were completed from 2014 to 2015 in PKUSS.

2.2 | Inclusion criteria

All cases of experimental sample and validation sample possessed complete clinical records, including charts, pre-treatment and post-treatment SCs, LXs and PHs. For all LXs, there was a calibrating ruler to ensure that the magnification of the X-ray films could be unified, and the LXs were printed in a unified format using the rulers.

The final experimental sample had an age distribution from 10 to 27 years, and there were 30 males and 78 females. The final validation sample had an age distribution from 10 to 32 years, and there were 19 males and 53 females. All selected cases were of Han nationality and free of craniofacial anomaly or syndrome. This study was approved by the ethics committee of PKUSS, and all subjects signed informed consent forms.

2.3 | Selection of experts

All invited experts participating in the exploration or verification research came from different dental schools or hospitals within China. They all had received professional postgraduate training in orthodontics or had the qualification of graduate tutor and had been engaged in professional orthodontic clinical work for more than 10 years, with an academic rank of associate professor or above.

2.4 | Subjective evaluation method

2.4.1 | Experimental sample

The 108 experimental cases were randomly divided into nine groups, 12 cases in each group with four cases of each Angle's classification. The 12 cases in each group were ranked from the most satisfactory to least by 69 experts according to the quality of treatment outcomes of the combined SC + LX records. In order to simplify experimental operation and guarantee the consistency of evaluations, the ranking process was divided into three steps: (a) 4 cases which were most satisfactory from the 12 cases were selected and put into 'good' subgroup, then 4 cases which were least satisfactory were selected and put into 'poor' subgroup and the remaining 4 cases belonged to 'medium' subgroup; (b) the 4 cases in 'good', 'medium' and 'poor' subgroups were ranked from best to worst respectively; (c) the cross-subgroup adjustment was allowed; for example, if the treatment outcome of the fourth case of the 'good' subgroup was worse than that of the first case of the 'medium' subgroup, the positions could be exchanged. After ranking, the 12 ranked cases in each group were given a grading evaluation of satisfactory, acceptable and unacceptable levels according to the quality of treatment outcomes from best to least by 69 experts and scored as 1, 2 and 3 points, respectively.

2.4.2 | Validation sample

The same grouping mode was also applied to the 72 validation cases. The ranking and grading method were the same as those in the evaluation of 108 experimental cases.

The consistency of experts' subjective evaluations would determine whether it could be used as the professional standard to screen out relevant objective indicators that were significantly correlated with experts' subjective evaluations of orthodontic treatment outcomes.

2.5 | Objective measurement method

2.5.1 | Model indicators

The PAR index and ABO-OGS were both applied for measurement of the experimental and validation samples. The PAR index comprised of five indicators, including alignment, occlusal relationship, overbite, overjet and midline. The ABO-OGS comprised of seven indicators, including alignment, marginal ridge height, buccolingual inclination, occlusal relationship, occlusal contact, overjet and

interproximal contact. Six orthodontic postgraduates in PKUSS were randomly divided into two groups and each group use one index to measure the post-treatment SCs. Before starting the measurement, the pilot study of examining measurers' reliability and validity were conducted. Then, the agreement on the measurement standards of all indicators was reached, and the calibration rules were listed.

2.5.2 | Cephalometric indicators

Before carrying out digital cephalometric measurements, three orthodontic postgraduates in PKUSS all had completed the cephalometry course, and the calibration rules of measuring 43 soft and hard tissue identification indicators were listed. All measured linear distances were corrected for the magnification using the calibrating ruler. The mean value and standard deviation of all cephalometric indicators referred to the research of Yu et al.[17] The actual measured value of all cephalometric indicators were standardized, that is Z-score = |(measured value - mean value)/standard deviation|.

2.6 | Statistical analysis

All statistical analyses were performed with SPSS software (version 20.0; SPSS). Intraclass correlation coefficients (ICCs) were computed to evaluate the intraexaminer and interexaminer reliabilities of the measurers. The Kendall's tau-c analysis were used to analyse rank and grade variables. The Pearson correlation analysis and multiple linear regression analysis were used to analyse the subjective and objective relationship between grade variables and metric variables. Receiver operating characteristic (ROC) curves were used to determine the critical values of satisfactory, acceptable and unacceptable grades. The whole research scheme was present as the flow chart below (Figure 1).

3 | RESULTS

3.1 | Consistency analysis of experts and measurers of 108 experimental sample

According to the previous studies of our research group, the panel of 69 orthodontic experts was moderately consistent in their evaluation of the post-treatment SC + LX records.[13] The ICCs among the three measurement groups who measured the PAR index, ABO-OGS and cephalometric indicators were 0.72, 0.74 and 0.97, respectively. The ICCs of the measurers themselves ranged from 0.77 to 0.81.

3.2 | Selection of objective indicators

In this study, the experts' subjective evaluation scores of 108 experimental sample were analysed by Pearson correlation with five indicators of the PAR index, seven indicators of ABO-OGS, and 43 cephalometric indicators, and the correlation coefficient and *P* value of all objective indicators were shown below (Table 1). In regard to the PAR index, the subjective evaluation showed a high correlation with overjet and midline, an average correlation with alignment and occlusal relationship, and no statistical correlation with overbite. In regard to ABO-OGS, the subjective evaluation showed a high correlation with occlusal contact, overjet, interproximal contact and marginal ridge height, an average correlation with alignment and occlusal relationship, and no statistical correlation with buccolingual inclination. In regard to the cephalometric indicators, the hard tissue

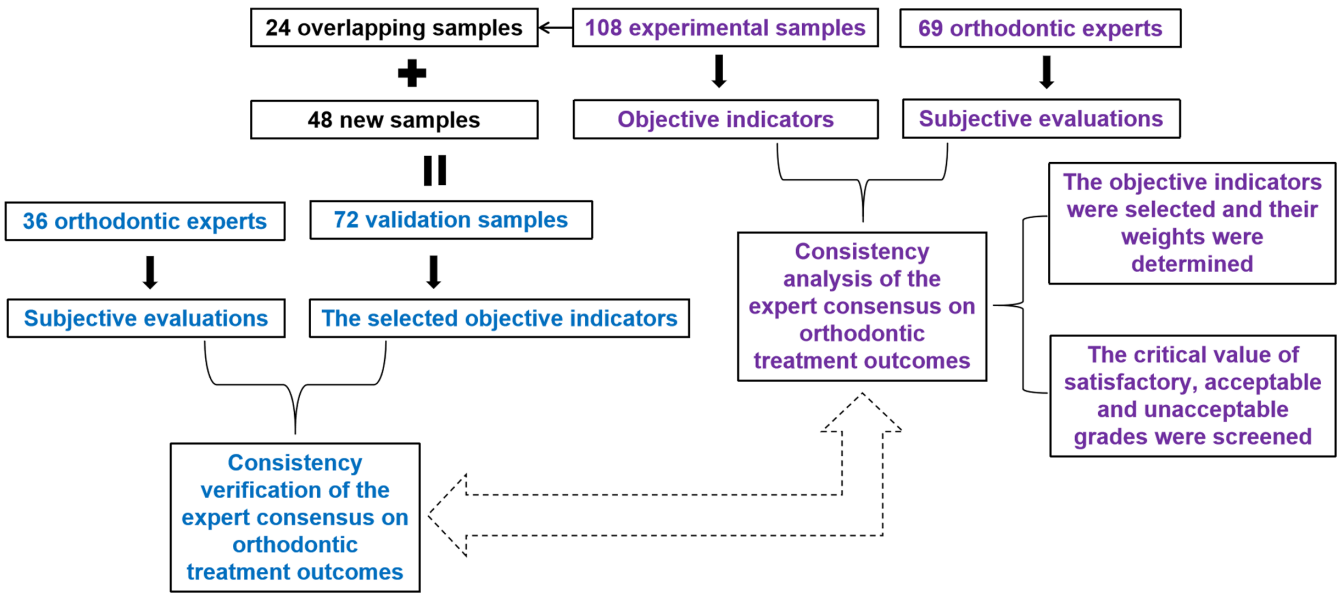

FIGURE 1 Flow chart of research scheme. (Purple). The exploration of expert consensus. (Blue) The verification of the reliability and validity of the expert consensus

indicators with a high correlation with the subjective evaluation were L1/NB, L1/MP, ANB, AB/NP, L1-NB, NA/PA, SNB, SND, U1/L1 and SN/OP; the indicators with an average correlation were FH/OP, U1/SN, FH/NP, L1/AP, U1/NA and U1/PP, and the remaining 7 indicators had no statistical correlation with the subjective evaluation. Among the soft tissue indicators, the indicators with a high correlation with the subjective evaluation were LL/EP, LL/H, StomiMes and Z-angle; the indicators with an average correlation were Cm-Sn-UL, Bs-B (FH), AsUL-FH, Ns-Prn-Pos, UL-EP, G'-Sn-Pos and Sn-stoms; and the remaining 9 indicators had no statistical correlation with the subjective evaluation.

By comparing the values of the correlation coefficients and considering the statistical collinearity and clinical significance, six model indicators and seven cephalometric indicators significantly correlated with the experts' subjective evaluation was finally selected, which were occlusal contact, overjet, midline, interproximal contact, alignment, occlusal relationship, L1/NB, ANB, SN/OP, U1/SN, LL-EP, Cm-Sn-UL and Ns-Prn-Pos.

3.3 | Determination of the weight of objective indicators and the critical value of satisfactory, acceptable and unacceptable grades

Multiple linear regression analysis was used to determine the weight of the selected objective indicators. The dependent variable was the experts' subjective evaluation scores of the combined post-treatment SC + LX records, and the independent variables were the 13 objective indicators selected above, including occlusal contact, overjet, midline, interproximal contact, alignment, occlusal relationship, L1/NB, ANB, SN/OP, U1/SN, LL-EP, Cm-Sn-UL and Ns-Prn-Pos. The results of multiple linear regression analysis between the experts' subjective evaluation and 13 objective indicators were shown below, with a cumulative R^2 of 0.704 (Table 2). The relative weight of each indicator was determined by comparing the standardized coefficients. The SC record in the evaluation of treatment outcomes was significantly greater than that of LX record, approximately twice that of LX record.

The objective regression scores obtained from the objective assessment criteria would not be able to explain the corresponding clinical significance without threshold setting; then one purpose of this study was to develop an objective assessment method with clinical discriminative significance. Since the panel of 69 orthodontic experts were moderately consistent in their evaluations, whose Kendall coefficient of concordance for group ranking and grading scores were 0.58 and 0.52 ($P < .05$),[13] it is reasonable to consider the grading evaluation results given by more than half of experts as the final subjective grading results of the given case. Therefore, the principles for determining the subjective critical values of satisfactory, acceptable and unacceptable grades were as follows, the averaged subjective grading scores of a case with more than half of experts grading as satisfactory was taken as the critical value for distinguishing between the satisfactory and acceptable grades, which was 1.57; and the averaged subjective grading scores of a case with more than half of experts grading as acceptable was taken as the critical value for distinguishing between the acceptable and unacceptable grades, which was 2.42. Subsequently, the objective critical values of satisfactory, acceptable and unacceptable grades with the highest diagnostic sensitivity, specificity and accuracy were screened by ROC curve analysis and shown below (Table 3). The objective critical values of satisfactory/acceptable and acceptable/unacceptable were 1.56 and 2.41, respectively. Thus the treatment outcomes of cases whose objective regression scores were lower than 1.56 were satisfactory; those with objective regression scores between 1.56 and 2.41 were acceptable cases; those with objective regression scores greater than 2.41 were unacceptable cases.

3.4 | Consistency analysis of experts and measurers of 72 validation samples

The consistency of 36 experts' subjective evaluations for the 72 validation samples was above average. The Kendall coefficient of concordance for group ranking and grading scores were 0.53 and 0.54 ($P < .05$), respectively. For the 24 overlapping cases, which were selected from 108 experimental sample, the ICCs of consistency analysis between two panels of experts ranged from 0.44 to 0.72. The ICCs of the objective measurement of model and cephalometric indicators were 0.81 and 0.97, respectively. The ICCs of the measurers themselves ranged from 0.77 to 0.81.

3.5 | Verification of the expert consensus on the assessment methods

Since the panel of 36 orthodontic experts were moderately consistent in their evaluations, the averaged subjective evaluation of 36 experts for the 72 validation samples were taken as the professional standard to verify the validity of 13 objective indicators selected above and the validity of satisfactory/acceptable and acceptable/unacceptable critical values determined above. The consistency between the average scores of 36 experts' subjective evaluation and the objective regression scores was analysed by Pearson correlation analysis, and the correlation coefficient was 0.716 ($P < .001$); the consistency between the 36 experts' subjective grades and objective grades was analysed by Kendall's tau-c analysis, and the correlation coefficient was 0.757 ($P < .001$). The coincidence rate of satisfactory/acceptable and acceptable/unacceptable cut-off of regression equation by ROC curve analysis were 80.7% and 86.5%, respectively.

4 | DISCUSSION

Previous studies have found that orthodontic experts with years of clinical experience are able to provide the most authoritative

TABLE 1 Pearson correlation between subjective evaluation scores and objective indicators

Model indicators	Pearson *r*	*P* value	Cephalometric indicators-hard tissue	Pearson *r*	*P* value	Cephalometric indicators-soft tissue	Pearson *r*	*P* value
PAR index			L1/NB (°)	.446	<.01	LL-EP (mm)	.436	<.01
Overjet	.504	<.001	L1/MP (°)	.396	<.01	LL-H (mm)	.340	<.01
Midline	.390	<.001	ANB (°)	.390	<.01	Stomi-Mes (mm)	.298	<.01
Alignment	.278	.004	AB/NP (°)	.353	<.01	Z-angle (°)	.264	<.01
Occlusal relationship	.226	.018	L1-NB (mm)	.336	<.01	Cm-Sn-UL (°)	.245	<.05
Overbite	.091	.350	NA/AP (°)	.321	<.01	Bs-B (FH) (mm)	.244	<.05
ABO-OGS			SNB (°)	.306	<.01	AsUL-FH (°)	.235	<.05
Occlusal contact	.641	<.001	SND (°)	.300	<.01	Ns-Prn-Pos (°)	.231	<.05
Overjet	.429	<.001	U1/L1 (°)	.288	<.01	UL-EP (mm)	.212	<.05
Interproximal contact	.374	<.001	SN/OP (°)	.271	<.01	G'-Sn-Pos (°)	.208	<.05
Marginal ridge height	.370	<.001	FH/OP (°)	.234	<.05	Sn-Stoms (mm)	.202	<.05
Alignment	.311	.001	U1/SN (°)	.230	<.05	Chin Thickness (mm)	.160	.099
Occlusal relationship	.205	.034	FH/NP (°)	.219	<.05	BsLL-FH (°)	.149	.124
Buccolingual inclination	.049	.613	L1/AP (°)	.204	<.05	AsUL-BsLL (°)	.143	.139
			U1/NA (°)	.203	<.05	LL-Bs-Pos (°)	.133	.169
			U1/PP (°)	.201	<.05	Ns-N (FH) (mm)	.047	.626
			GoGn/SN (°)	.185	.055	S-Ns-Sn (°)	.044	.652
			U1-NA (mm)	.146	.133	S-Ns-Bs (°)	.034	.729
			Pg-NB (mm)	.101	.296	Sn-Prn (FH) (mm)	.017	.864
			MP/FH (°)	.068	.485	Sn-A (FH) (mm)	.010	.916
			Y-axis (°)	.064	.508			
			SE (mm)	.047	.629			
			SNA (°)	.036	.712			

standard in the evaluation of orthodontic treatment outcomes.[3,18,19] In this study, orthodontic experts with at least 10 years of orthodontic experience were selected to assess the orthodontic treatment outcomes in a qualitative manner. Since the reliability and validity should be considered when evaluating the indicators of any indices, the consistency of experts' subjective evaluation (reliability) and consistency between experts' subjective evaluation and objective measurement results (validity) need being paid most attention.

In this study, a subjective evaluation modality combining ranking and grading was designed to guarantee the consistency of experts' subjective evaluation. The previous subjective evaluation modality mainly included grade scoring method and rank scoring method.[20] However, judges making grading evaluation tend to score at the medium level and could not distinguish the subtle differences of various complex cases, whereas ranking evaluation makes the results of multiple cases approximate the normal distribution and reflect the differences in more detail, but it may be difficult and cumbersome for experts to rank the multiple cases at one time. In this study, the whole sample was divided into several groups and every 12 cases of each group were ranked first and then graded. Since the SC record evaluated in this study belonged to three-dimensional data, whose dentition alignment and occlusal relationship should be observed from three directions of horizontal, vertical and sagittal side, it was difficult for experts to rank 12 cases at one time. Therefore, the ranking method for every 12 cases included subgrouping, intra-subgroup ranking and cross-subgroup adjusting, which was convenient for experts to compare the subtle differences in the given cases and avoided the difficulty of cumbersome ranking if there were too many cases in each subgroup, or cumbersome subgroup-comparing if there were too many subgroups. After group ranking, experts were asked to provide satisfactory, acceptable and unacceptable grades for the 12 cases of each group. By means of Kendall's tau-c analysis, it was comforting to find that the panel of 69 orthodontic experts were moderately consistent in their ranking and grading evaluations, showing good reliability. Then the averaged experts' evaluation could be taken as the professional standard.

Modern orthodontic treatment emphasizes not only the importance of dental aesthetic but also profile harmony after orthodontic treatment.[21] In addition to the positive evaluation of dentition characteristics, the evaluation of profile harmony is also significant, whose involvement will greatly affect the evaluation results. For example, bimaxillary protrusion cases with and without teeth

TABLE 2 Multiple linear regression between subjective evaluation scores and objective indicators

Variables	Unstandardized coefficients		Standardized coefficients	*t* value	*P* value
	B	Standard error			
Constant	0.921	0.105		8.766	0
Occlusal contact	0.056	0.010	0.399	5.416	0
Overjet	0.296	0.102	0.197	2.891	.005
Midline	0.241	0.115	0.135	2.086	.040
Interproximal contact	0.146	0.039	0.226	3.721	0
Alignment	0.035	0.013	0.157	2.621	.010
Occlusal relationship	0.035	0.013	0.166	2.758	.007
L1/NB	0.107	0.027	0.278	3.898	0
ANB	0.006	0.032	0.014	0.189	.851
SN/OP	0.007	0.035	0.012	0.203	.840
U1/SN	0.092	0.030	0.201	3.108	.002
LL-EP	0.033	0.033	0.065	1.006	.317
Cm-Sn-UL	0.032	0.050	0.039	0.642	.523
Ns-Prn-Pos	0.033	0.043	0.045	0.759	.450

TABLE 3 Objective cut-off, sensitivity, specificity, accuracy and AUC obtained by ROC curve analysis

Grading evaluation	Cut-off	Sensitivity	Specificity	Accuracy	AUC
Satisfactory/acceptable	1.56	0.909	0.988	0.983	0.957
Acceptable/unacceptable	2.41	0.969	0.987	0.996	0.995

extractions were shown below (Figure 2). The study casts for these cases show good dental alignment and occlusal relationship, thus both of them would receive excellent results when using the PAR index or ABO-OGS to evaluate the treatment outcomes. However, after evaluating the profile harmony from lateral cephalograms, the subjective assessment of the non-extraction case was poorer in comparison to the extraction case (Table 4), due to the protrusion of upper and lower lips, the proclination and protrusion of upper and lower incisors and the disharmony of nose–lip–chin relationship. Besides, a more comprehensive assessment method would be more conducive to guide orthodontists in accurate record analysis and diagnostic design. Therefore, in addition to the PAR index and ABO-OGS,[22–24] an aesthetic evaluation of the profile harmony was proposed in this study. The LX records were included in that they covered a large number of morphological characteristics of the bone, dentition and soft tissue and the cephalometric methods could be standardized and quantified. The PH records were not included because the evaluation of PHs could be affected by the sex, age, skin colour and eye characteristics of patients, which were difficult to quantify and control.

Evaluation of the validity of an index involves enquiring into whether the index measures what it claims to measure.[25–29] Various studies have confirmed the validity of objective measurement components of the PAR index and ABO-OGS,[1–3,22,23] thus the model indicators of the PAR index and ABO-OGS were used to measure the features of SC record in this study. Since the ICCs of intraexaminers and interexaminers were all above 0.7, the measurement results of three times of objective indicators were averaged as the final measurement results. Statistical analysis showed that objective indicators such as occlusal contact, overjet, midline and interproximal contact were highly correlated with the experts' evaluation. Although the indicator of marginal ridge height in ABO-OGS showed a high correlation with the experts' subjective evaluation, since its high correlation with occlusal contact that was also high correlated with the experts' subjective evaluation, it was not selected. By the same token, although alignment and occlusal relationship were not the most important indicators in the correlation analysis, they were still selected because they could independently reflect information that could not be represented by other indicators. With respect to the LX record, since various anthropometric values have a certain normal range and the greater the deviation from the normal range, the more serious the deformity or the worse the curative effect, we converted the cephalometric data into standard scores to quantify the degree of deviation from normal. According to the correlation analysis, the inclination and protrusion of the lower incisors, the protrusion of the lower lip and the sagittal and vertical positional relationships of the maxilla and mandible were the cephalometric indicators of most concern for experts. Considering the statistical collinearity and the profile characteristics represented by cephalometric indicators, the indicators related to profile harmony were selected, such as the inclination and protrusion of upper incisors, the nasolabial coordination and mentolabial coordination, which was consistent with a study by Oh et al.[30] showing that profile angle, chin prominence, lower lip prominence, Z-angle, NA/PA, MP/SN, ANB, LL-E, L1/AP, B-line to lower lip and L1/MP were significantly correlated with the subjective evaluation. As a result, the 13 objective

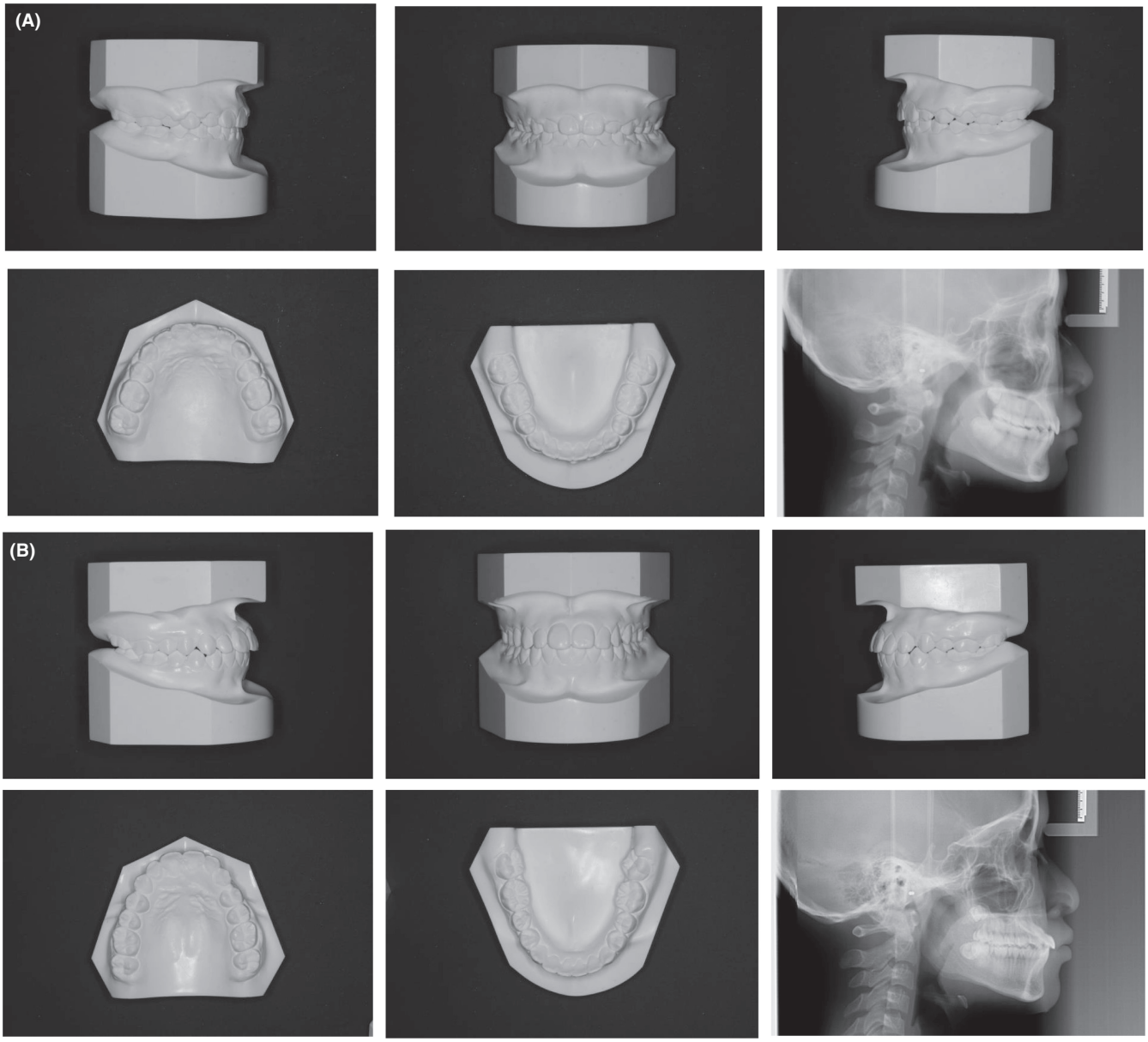

FIGURE 2 Bimaxillary protrusion cases with (A) and without (B) teeth extractions

TABLE 4 Subjective evaluation of the bimaxillary protrusion cases with and without teeth extractions

Grade	Times of experts' grading evaluation	
	Case with teeth extractions	Case without teeth extractions
Satisfactory	47	20
Acceptable	20	31
Unacceptable	2	18
Average score	1.35	1.97
Grade	Satisfactory	Acceptable

indicators of occlusal contact, overjet, midline, interproximal contact, alignment, occlusal relationship, L1/NB, ANB, SN/OP, U1/SN, LL-EP, Cm-Sn-UL and Ns-Prn-Pos were ultimately selected, and the regression equation of the 13 objective indicators yielded a cumulative R^2 of 0.704 in multiple linear regression analysis, showing great validity. Simultaneously, the standardized coefficients of each indicator in regression analysis were used to determine their weights in the objective assessment. Although there were more cephalometric indicators being included than model indicators, the weight of SC record in the objective evaluation was greater than LX record, which was consistent with the research conclusion of Song et al.[13]

In the verification part of this study, the average evaluation results of the combined post-treatment SC+LX records of the 24 overlapping cases by 36 experts were moderately correlated with the average evaluation results of the 69 experts in the exploration part, which implied that the two panels of experts were of a similar professional perception. In addition, the consistency of 36 experts' subjective evaluations of the validation sample was moderate,

showing good reliability and that the average evaluation of the 36 experts could be taken as the professional standard. As a result, the correlation coefficient between the average scores of 36 experts' subjective evaluation and the objective regression scores was 0.716 ($P < .001$), showing the good consistency between subjective and objective evaluation; the correlation coefficient between the 36 experts' subjective grade and objective grade was 0.757 ($P < .001$), showing the good validity of the satisfactory/acceptable and acceptable/unacceptable critical values. Thus, the expert consensus on the assessment methods in this study can be used to qualitatively and quantitatively measure the dental aesthetic and profile harmony and represent a professionally acceptable appearance with satisfactory discrimination.

5 | CONCLUSIONS

Orthodontic experts had good consistency in the subjective evaluation of the combined records of post-treatment study casts and lateral cephalograms. The objective indicators significantly correlated with experts' subjective evaluations were selected on the basis of correlation analysis and regression analysis between subjective and objective evaluation, with good reliability and validity. The expert consensus on the assessment of orthodontic treatment outcomes included six model indicators and seven cephalometric indicators, thus could qualitatively and quantitatively measure the dental aesthetic and profile harmony and further improve the existing occlusal indices.

AUTHOR CONTRIBUTIONS

Huanhuan Chen and Guangying Song contributed to data curation, formal analysis, investigation and writing original draft. Weiran Li, Ruoping Jiang, Xiaoyun Zhang, Si Chen, Gui Chen, Siqi Liu, Fanfan Dai and Fei Teng contributed to conceptualization and project guidance. Bing Han and Tianmin Xu contributed to conceptualization, funding acquisition and critically revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work.

ACKNOWLEDGEMENTS

We thank all the patients, 69 orthodontic experts evaluating the 108 experimental samples, and 36 orthodontic experts evaluating the 72 validation samples. We also thank the West China School of Stomatology at Sichuan University, the School of Stomatology at the Fourth Military Medical University, the Beijing Stomatological Hospital and School of Stomatology at Capital Medical University, the Stomatological Hospital and College of Nanjing Medical University and the Hospital of Stomatology at Wuhan University. This study would not have been possible without their participation. This work was supported by the National Natural Science Foundation of China (No. 82001082, No. 51972005 and No. 82071172).

CONFLICTS OF INTEREST

All authors have completed and submitted the ICMJE Form for disclosure of potential conflicts of interest, and none were reported.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Bing Han https://orcid.org/0000-0003-4584-1092

REFERENCES

1. Hamdan AM, Rock WP. An appraisal of the Peer Assessment Rating (PAR) Index and a suggested new weighting system. *Eur J Orthod*. 1999;21(2):181-192.
2. Casko JSs, Vaden JL, Kokich VG, Damone J, James RD, Cangialosi TJ. Objective grading system for dental casts and panoramic radiographs. American Board of Orthodontics. *Am J Orthod Dentofacial Orthop*. 1998;114(5):589-599.
3. Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD. The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod*. 1992;14(2):125-139.
4. Daniels C, Richmond S. The development of the index of complexity, outcome and need (ICON). *J Orthod*. 2000;27(2):149-162.
5. Liao ZY, Jian F, Long H, Lu Y, Wang Y, Yang Z. Validity assessment and determination of the cutoff value for the Index of Complexity, Outcome and Need among 12-13 year-olds in Southern Chinese. *Int J Oral Sci*. 2012;4(2):88-93.
6. Eismann D. A method of evaluating the efficiency of orthodontic treatment. *Trans Eur Orthod Soc*. 1974;4(1):223-232.
7. Onyeaso CO, Begole EA. Relationship between index of complexity, outcome and need, dental aesthetic index, peer assessment rating index, and American Board of Orthodontics objective grading system. *Am J Orthod Dentofacial Orthop*. 2007;131(2):248-252.
8. Fox NA, Chapple JR. Measuring failure of orthodontic treatment: a comparison of outcome indicators. *J Orthod*. 2004;31(4):319-322.
9. Pinskaya YB, Hsieh TJ, Roberts WE, Hartsfield JK. Comprehensive clinical evaluation as an outcome assessment for a graduate orthodontics program. *Am J Orthod Dentofacial Orthop*. 2004;126(5):533-543.
10. Savastano NJ Jr, Firestone AR, Beck FM, Vig KW. Validation of the complexity and treatment outcome components of the index of complexity, outcome, and need (ICON). *Am J Orthod Dentofacial Orthop*. 2003;124(3):244-248.
11. Kragt L, Dhamo B, Wolvius EB, Ongkosuwito EM. The impact of malocclusions on oral health-related quality of life in children-a systematic review and meta-analysis. *Clin Oral Investig*. 2016;20(8):1881-1894.
12. Baan F, de Waard O, Bruggink R, Xi T, Ongkosuwito EM, Maal TJJ. Virtual setup in orthodontics: planning and evaluation. *Clin Oral Investig*. 2020;24(7):2385-2393.
13. Song GY, Zhao ZH, Ding Y, Bai YX, Wang L, He H. Reliability assessment and correlation analysis of evaluating orthodontic treatment outcome in Chinese patients. *Int J Oral Sci*. 2014;6(1):50-55.
14. Song GY, Li WR, Geng Z, Xu TM. Discussion on the consistency of subjective evaluation of orthodontic efficacy satisfaction. *J Peking Univ*. 2014;44(1):103-107.
15. Kiekens RM, Kuijpers-Jagtman AM, van't Hof MA, van't Hof BE, Straatman H, Maltha JC. Facial esthetics in adolescents and its relationship to "ideal" ratios and angles. *Am J Orthod Dentofacial Orthop*. 2008;133(2):188 e1-8.

16. Yu X, Liu B, Pei Y, Xu T. Evaluation of facial attractiveness for patients with malocclusion: a machine-learning technique employing Procrustes. *Angle Orthod*. 2014;84(3):410-416.
17. Xiaohui Y, JinFang H. Cephalometric analysis of soft tissue profile. *Foreign Med*. 1987;2:93-97.
18. Xiao-Ting L, Tang Y, Huang XL, Wan H, Chen YX. Factors influencing subjective orthodontic treatment need and culture-related differences among Chinese natives and foreign inhabitants. *Int J Oral Sci*. 2010;2(3):149-157.
19. DeGuzman L, Bahiraei D, Vig KW, Vig PS, Weyant RJ, O'Brien K. The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. *Am J Orthod Dentofacial Orthop*. 1995;107(2):172-176.
20. Song GY, Zhao ZH, Ding Y, Bai YX, Wang L, He H. Study on subjective evaluation of orthodontic efficacy satisfaction by 69 orthodontists. *Chin J Stomatol*. 2012;47(3):134-138.
21. Andrews LF. The 6-elements orthodontic philosophy: Treatment goals, classification, and rules for treating. *Am J Orthod Dentofacial Orthop*. 2015;148(6):883-887.
22. Song GY, Baumrind S, Zhao ZH, Ding Y, Bai YX, Wang L. Validation of the American Board of Orthodontics Objective Grading System for assessing the treatment outcomes of Chinese patients. *Am J Orthod Dentofacial Orthop*. 2013;144(3):391-397.
23. Chaison ET, Liu X, Tuncay OC. The quality of treatment in the adult orthodontic patient as judged by orthodontists and measured by the Objective Grading System. *Am J Orthod Dentofacial Orthop*. 2011;139(4):S69-S75.
24. Deguchi T, Honjo T, Fukunaga T, Miyawaki S, Roberts WE, Takano-Yamamoto T. Clinical assessment of orthodontic outcomes with the peer assessment rating, discrepancy index, objective grading system, and comprehensive clinical assessment. *Am J Orthod Dentofacial Orthop*. 2005;127(4):434-443.
25. Hong M, Kook YA, Kim MK, Lee JI, Kim HG, Baek SH. The Improvement and Completion of Outcome index: a new assessment system for quality of orthodontic treatment. *Korean J Orthod*. 2016;46(4):199-211.
26. Louwerse TJ, Aartman IH, Kramer GJ, Prahl-Andersen B. The reliability and validity of the Index of Complexity, Outcome and Need for determining treatment need in Dutch orthodontic practice. *Eur J Orthod*. 2006;28(1):58-64.
27. Liu S, Oh H, Chambers DW, Baumrind S, Xu T. Validity of the American Board of Orthodontics Discrepancy Index and the Peer Assessment Rating Index for comprehensive evaluation of malocclusion severity. *Orthod Craniofac Res*. 2017;20(3):140-145.
28. Costa RN, Abreu MH, Magalhães CS, Moreira AN. Validity of two occlusal indices for determining orthodontic treatment needs of patients treated in a public university in Belo Horizonte, Minas Gerais State, Brazil. *Cad Saude Publica*. 2011;27(3):581-590.
29. Younis JW, Vig KW, Rinchuse DJ, Weyant RJ. A validation study of three indexes of orthodontic treatment need in the United States. *Community Dent Oral Epidemiol*. 1997;25(5):358-362.
30. Oh HS, Korn EL, Zhang X, Liu Y, Xu T, Boyd R. Correlations between cephalometric and photographic measurements of facial attractiveness in Chinese and US patients after orthodontic treatment. *Am J Orthod Dentofacial Orthop*. 2009;136(6):762-763.

How to cite this article: Chen H, Song G, Li W, et al. Subjective and objective analysis of orthodontic expert consensus on the assessment of orthodontic treatment outcomes. *Orthod Craniofac Res*. 2023;26:197-206. doi: 10.1111/ocr.12600