A Prediction Model for Types of Treatment Indicated for Patients with Temporomandibular Disorders

Naichuan Su, PhD(c)
State Key Laboratory of Oral Disease
Department of Prosthodontics
West China Hospital of Stomatology
Sichuan University
Chengdu, PR China;
Department of Social Dentistry
Academic Centre for Dentistry
Amsterdam (ACTA)
University of Amsterdam and
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands

Corine M. Visscher, PT, PhD
Department of Oral Kinesiology

Arjen J. van Wijk, PhD
Department of Social Dentistry

Frank Lobbezoo, DDS, PhD
Department of Oral Health Sciences

Geert J.M.G. van der Heijden, PhD
Department of Social Dentistry

Academic Centre for Dentistry
Amsterdam (ACTA)
University of Amsterdam and
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands

Correspondence to:
Dr Naichuan Su
Department of Social Dentistry
Academic Centre for Dentistry
Amsterdam (ACTA)
University of Amsterdam and
Vrije Universiteit Amsterdam
Gustav Mahlerlaan 3004, 1081 LA
Amsterdam, the Netherlands
Email: n.su@acta.nl

Submitted October 12, 2017;
accepted March 14, 2018.

Aims: To identify potential predictors of types of treatment indicated for patients with temporomandibular disorders (TMD) and to develop, validate, and calibrate a prediction model for type of treatment. **Methods:** The derivation cohort at baseline was comprised of 356 adult patients with TMD. Patient and disease characteristics were recorded at baseline as potential predictors. Types of treatment indicated for TMD patients were the end points of the model, classified into no treatment, physical treatment only (including splint and/or physiotherapy), and combined physical and psychological treatment. Multinomial logistic regression analysis was used to develop the prediction model. The internal validation, calibration, discrimination, and external validation of the model were determined. For practical use, the prediction model was converted into score charts and line charts. The score of each included predictor was produced based on the shrunken regression coefficients. **Results:** Patient age, gender, anxiety, sleep bruxism, pain-related TMD, function-related TMD, stress, passive stretch of maximum mouth opening, and depression were significantly associated with the type of treatment indicated for TMD patients. The multinomial model showed reasonable calibration and good discrimination, with area under the curve values of 0.76 to 0.86. The internal validity of the model was good, with a shrinkage factor of 0.89. The external validity of the model was acceptable. **Conclusion:** Potential predictors in patient profiles for prediction of type of treatment indicated for TMD patients were identified. The internal validity, calibration, discrimination, and external validity of the model were acceptable. *J Oral Facial Pain Headache 2019;33:25–38. doi: 10.11607/ofph.2076*

Keywords: *decision-making, physical therapy modalities, prognosis, splints, temporomandibular joint disorders*

Temporomandibular disorders (TMD) is a collective term that encompasses several clinical problems related to the masticatory muscles, temporomandibular joints (TMJs), and associated structures[1] that manifest clinically as pain in the TMJ or masticatory muscles, sounds in the TMJ (including clicking, popping, and crepitus), headache limited to the temporal region, and otalgia and/or tinnitus in the absence of aural disease.[2] TMD are also conceptualized from a biopsychosocial perspective, which views pain and disability as a result of dynamic interactions among physical, psychological, behavioral, and social factors.[3] Nowadays, various reversible and conservative methods of physical treatment have been commonly adopted in TMD patient care, such as physiotherapy, splint therapy, and pain education.[2,4,5] Research evidence suggests that psychosocial factors also play an important role in treatment response and in the transition to chronicity.[3] It has been reported that psychological treatment, including cognitive behavioral therapy and relaxation techniques, can be effective in treating TMD.[6] In clinical practice, of course, not every TMD patient needs the same treatment, if any treatment at all, so pain and TMD specialists have emphasized the need to identify subgroups of patients who do not need treatment, who need only physical treatment including splint and/or physiotherapy (unitreatment), and who need psychological treatment in addition to physical treatment (multitreatment), and then to use this information for patient-tailored interventions.[3]

Traditionally, types of treatment indicated for TMD patients are based on the presentation of complaints from the patients and on observation, history taking, and clinical testing by the health care professional, as well as additional (laboratory) testing and imaging. Conventionally, expertise and consensus between professionals drives clinical routines in patient care; however, probably not all procedures routinely used in daily practice will contribute to the efficacy and accuracy of the differential diagnosis and subsequent types of treatment indicated for patients.

In current patient care for TMD patients, clinicians' decision-making on what type of treatment is indicated for a patient is affected by several factors, such as the clinicians' knowledge, professional traditions, clinical routines, clinical judgment, and differential diagnostic reasoning. It has been suggested that variation in treatment decisions stems essentially from two main sources: perceptual variation and judgmental variation.[7] Patient characteristics can also have an influence on which treatment is indicated. It has been shown that clinicians generally use patients' demographic characteristics as a starting point in their assessments.[8] Disease characteristics also affect which treatment is indicated. Different patient complaints, signs and symptoms, and psychosocial statuses can result in clinicians choosing different treatment options for comparable patients.

To gain insight into clinical decision-making for different treatment choices for TMD patients, contemporary predictive modeling methodologies can be applied. This can improve transparency, efficiency, consistency, and accuracy in clinicians' decision-making. Moreover, data collected while following conventional practice may provide a framework for identification of factors that weigh most strongly on treatment decision-making. Therefore, the aims of the present study were *(1)* to identify potential predictors in patient profiles for types of treatment indicated for TMD patients, including no treatment (NT), physical treatment only (PTO; including splint and/or physiotherapy), and combined physical and psychological treatment (CPPT); and (2) to develop, validate, and calibrate a model for prediction of type of treatment indicated for TMD patients based on patient profiles.

Materials and Methods

Participants

In the present cross-sectional study, consecutive TMD patients referred to the clinic of the Department of Oral Kinesiology of the Academic Centre for Dentistry Amsterdam (ACTA) between September 2013 and February 2016 were included in the study as the derivation cohort. The inclusion criteria were patients who entered the clinic for orofacial pain and/or functional disturbances, were 18 years of age or above, and had a complete diagnosis of TMD based on the Diagnostic Criteria for TMD (DC/TMD).[9]

This study concerns a Health Services Research project that, under the Medical Research Involving Human Subjects Act (WMO), is not considered medical scientific research (www.ccmo.nl/en/non-wmo-research; http://www.ccmo.nl/en/types-of-research). As such, ethics clearance from a Medical Ethics Research Board nor individual consent of patients are required. The clinicians documented patient data on health status, diagnosis, and treatment and have shared these data from their health care records with the research team. These data were anonymized to preclude the possibility of re-identification before being analyzed. The researchers involved have taken care to handle data (processing, cleaning, and analyzing) in a secure, anonymous, and privacy-protected manner and have taken due care that data cannot be traced back to individuals. During the data analyses, confidentiality was maintained by data coding.

Collection of Data

Potential Predictors. Potential predictors, including patient characteristics (gender and age), disease characteristics, and diagnosis based on DC/TMD, were collected and documented (Table 1). All potential predictors were identified based on consensus of the experts in TMD (F.L., C.V.) a priori. For inclusion of the potential predictors in the models, a hierarchical approach was adopted; ie, the variable with the most important contribution to the type of treatment indicated for TMD patients was followed, and the variable with the lowest burden to TMD patients was included first.

Diagnosis of TMD. The criteria for diagnoses of TMD were based on the DC/TMD.[9] The diagnoses can be classified as: myalgia; arthralgia; headache attributed to TMD; disc displacement; degenerative joint disease; and subluxation. Pain-related TMD was defined as the presence of myalgia and/or arthralgia, and non–pain-related TMD was defined as the absence of both myalgia and arthralgia. Function-related TMD was defined as the presence of disc displacement, degenerative joint disease, or subluxation, and non–function-related TMD was defined as the absence of disc displacement, degenerative joint disease, and subluxation. Furthermore, function-related TMD was classified into function-related TMD without treatment need and function-related TMD with treatment need. The former was defined in terms of patients who had one or more of the following diagnoses: disc displacement with reduction, disc displacement without reduction without limited mouth opening, or degenerative joint disease; the latter was

defined in terms of patients who had at least one of the following diagnoses: disc displacement with intermittent locking, disc displacement without reduction with limited opening, and subluxation. The criteria for coding of the TMD diagnoses are presented in Table 1.

TMD Pain Screener. The TMD pain screener was used to identify whether patients suffered from pain-related TMD conditions in the last 30 days.[10] The questionnaire consists of 6 items assessing how long the pain lasted in the jaw and temple areas, whether the patient had pain or stiffness in the jaw on awakening, and whether daily activities—including chewing hard or tough food, mouth opening, moving the jaw forward or to the side, jaw habits, and other activities—changed any pain in the jaw or temple areas. Item 1 is scored as 0 (no pain), 1 (pain comes and goes), or 2 (pain is always present); items 2 to 6 are scored as 0 (no pain) or 1 (pain). The answers to these 6 items are summed; a higher sum score indicates a higher risk of pain-related TMD (Table 1).[10]

Graded Chronic Pain Scale. The Graded Chronic Pain Scale (GCPS) was used to assess patients' orofacial chronic pain intensity and pain-related disability.[11] Orofacial chronic pain intensity was assessed using the Characteristic Pain Intensity (CPI) scale from the GCPS.[11] The CPI score can be between 0 and 100 and is assessed by calculating the mean of current facial pain intensity, as well as the worst and average facial pain intensities during the last 6 months. Orofacial chronic pain–related disability was measured with the disability points based on the GCPS.[11] Disability points are between 0 and 6 and are determined with the disability score (the mean ratings of how much facial pain has interfered with patients' daily activities, recreational, social, and family activities, and work ability during the last 6 months) and disability days (the total number of days that facial pain kept a patient from doing usual activities during the last 6 months). The GCPS categories, which range from 0 to 4, are based on both the CPI and pain-related disability. In the present study, Category 3 and Category 4 were combined, since both of these categories focus on the pain-related disability regardless of CPI (Table 1).

Oral Behaviors Checklist. The Oral Behaviors Checklist (OBC) is an instrument for determining the presence or awareness of oral parafunctional behaviors during sleep and during waking hours.[12] In the present study, patients' self-reported awake bruxism (clenching and grinding) and sleep bruxism (clenching and grinding) were based on the OBC. Patients were asked to report the frequency of awake bruxism and sleep bruxism over the past month. For sleep bruxism, the responses of each item are: 0 (none of the time), 1 (< 1 night per month), 2 (1–3 nights per month), 3 (1–3 nights per week), and 4 (4–7 nights per week). For awake bruxism, the responses of each item are: 0 (none of the time), 1 (a little of the time), 2 (some of the time), 3 (most of the time), and 4 (all of the time) (Table 1).[12]

Psychosocial Assessment. The 7-item Generalized Anxiety Disorder (GAD-7) was used to evaluate patients' anxiety during the last 2 weeks. For each item, the frequency of anxious problems that patients are bothered by in daily life can be rated as 0 (not at all), 1 (several days), 2 (more than half of the days), or 3 (nearly every day). The sum score of GAD-7 ranges from 0 to 21; a higher sum score indicates more severe anxiety (Table 1).[13]

The 15-item Patient Health Questionnaire (PHQ-15) was used to evaluate patients' nonspecific physical symptoms during the last 4 weeks. On each item, the extent to which patients were bothered by several somatic problems such as stomach pain, back pain, or headaches can be rated as 0 (bothered), 1 (bothered a little), or 2 (bothered a lot). The sum score of PHQ-15 ranges from 0 to 30; a higher sum score indicates more severe somatization (Table 1).[14]

The 9-item Patient Health Questionnaire (PHQ-9) was used to evaluate patients' depression during the last 2 weeks. On each item, the frequency of depressed problems in daily life can be rated as 0 (not at all), 1 (several days), 2 (more than half of the days), or 3 (nearly every day). The sum score of PHQ-9 ranges from 0 to 27; a higher sum score indicates more severe depression (Table 1).[15]

A 7-item questionnaire developed by van der Meulen et al was used to evaluate patients' psychological stress in daily life during the past 6 months.[16] One question focused on patients' self-reported overall stress experienced over the past month. On each item, the extent of stress can be rated as 0 (none), 1 (a little bit), 2 (somewhat), 3 (rather much), or 4 (very much). The sum score ranges from 0 to 28; a higher sum score indicates more stress in daily life (Table 1).[16]

The Epworth Sleepiness Scale (ESS) was used to assess patients' recent chronic daytime sleepiness in daily life. It includes 8 items that are related to how likely patients doze off in 8 different situations in daily life. The answer scale for each item is 0 (never doze), 1 (slight chance of dozing), 2 (moderate chance of dozing), or 3 (high chance of dozing). The sum score ranges from 0 to 24; a higher sum score indicates more severe chronic daytime sleepiness (Table 1).[17]

Chronicity of TMD Pain. Patients' chronicity of TMD pain was assessed by one question extracted from the DC/TMD Symptom Questionnaire: How many years or months ago did your pain in the jaw, temple, in the ear, or in front of the ear first begin? (Table 1).[9]

Table 1 Description of Diagnostic Coding and Bivariate Association Analyses Between Outcome Category and Potential Predictors in Derivation (n = 356) and Validation (n = 180) Cohorts

Predictors	Descriptions of coding
Gender	Female Male
Age[b]	Continuous (mean ± SD)
Pain-related TMD	No pain-related TMD Pain-related TMD
Function-related TMD	No function-related TMD Function-related TMD with treatment need
TMD pain screener	No pain-related TMD (score ≤ 3) Pain-related TMD (score > 3)
GCPS	No TMD pain in prior 6 months Low-intensity pain without disability High-intensity pain without disability Moderately/severely limiting
Clench or grind teeth when asleep, based on any information you may have (OBC)	None of the time Yes
Clench or grind teeth during waking hours (OBC)	None of the time Yes
GAD-7	No anxiety (score: 0-4) Mild anxiety (score: 5-9) Moderate/severe anxiety (score: 10-21)
PHQ-15	No somatization (score: 0-4) Low somatization (score: 5-9) Medium/high somatization (score: 10-30)
PHQ-9	No depression (score: 0-4) Mild depression (score: 5-9) Moderate-severe depression (score: 10-27)
Psychological stress (van der Meulen et al[16])	No stress (score: 0-3) Somewhat (score: 4-10) Quite/much/very much (score: 11-28)
ESS	Normal (score: 0-9) Sleepy or very sleepy (score: 10-24)
How many years or months ago did your pain in the jaw, temple, in the ear, or in front of the ear first begin?	No pain < 6 mo 6 mo to 2 y > 2 y
Have you undergone treatment for the complaint previously?	No Yes
Have you undergone psychological or psychiatric treatment previously?	No Yes
Passive stretch of maximal mouth opening	< 5 mm ≥ 5 mm

[a]*P* value based on bivariate association analyses between the outcome and each predictor.
[b]*P* value based on Kruskal-Wallis test.
[c]*P* value of the predictor was > .20 and therefore excluded from backward regression.
NT = no treatment; PTO = physical treatment only; CPPT = combined physical and psychological treatment; TMD= temporomandibular disorders; GCPS = Graded Chronic Pain Scale; PHQ = Patient Health Questionnaire; GAD = Generalized Anxiety Disorder; OBC = Oral Behaviors Checklist.

Previous Treatment. Patients were asked "Have you undergone treatment for the complaint previously?" and "Have you undergone psychological or psychiatric treatment previously?" The answers to both questions were classified into yes or no (Table 1).

Passive Stretch of Maximum Mouth Opening. The passive stretch of maximum mouth opening was defined as patient-assisted maximum mouth opening minus patient-unassisted maximum mouth opening (Table 1).[9]

Outcomes

The type of treatment that was indicated for each patient (NT, PTO, or CPPT) was extracted from the patient records by the first author (N.S.). Physical treatment included splint therapy and physiotherapy, and psychological treatment was recorded when the patient was referred to the multidisciplinary team's psychologist as part of the treatment. The initial treatments indicated for patients were proposed by the clinician who performed the intake examination based

Derivation cohort (n = 356)					Validation cohort (n = 180)				
No. of patients (%)	NT (n = 88)	PTO (n = 158)	CPPT (n = 110)	*P* value[a]	No. of patients (%)	NT (n = 36)	PTO (n = 83)	CPPT (n = 61)	*P* value[a]
279 (78)	60	128	91	.026	144 (80)	29	64	51	.626
77 (22)	28	30	19		36 (20)	7	19	10	
43.4 ± 14.8	47.8 ± 15.5	42.2 ± 14.3	41.5 ± 14.4	.005	44.9 ± 14.8	52.6 ± 16.1	44.1 ± 14.3	41.3 ± 13.4	.002
137 (39)	65	53	19	< .001	53 (29)	29	20	4	< .001
219 (61)	23	105	91		127 (71)	7	63	57	
268 (75)	78	111	79	.004	148 (82)	30	65	53	.405
88 (25)	10	47	31		32 (18)	6	18	8	
141 (39)	61	55	25	< .001	59 (33)	26	25	8	< .001
215 (61)	27	103	85		121 (67)	10	58	53	
74 (21)	33	33	8	< .001	31 (17)	15	13	3	< .001
88 (25)	20	47	21		47 (26)	9	24	14	
86 (24)	11	39	36		58 (32)	6	33	19	
108 (30)	24	39	45		44 (24)	6	13	26	
118 (33)	47	49	22	< .001	66 (37)	19	31	16	.032
238 (67)	41	109	88		114 (63)	17	52	45	
104 (29)	34	48	22	.015	54 (30)	21	21	12	< .001
252 (71)	54	110	88		126 (70)	15	62	49	
217 (61)	64	118	35	< .001	116 (64)	30	61	25	< .001
84 (24)	15	28	41		43 (24)	5	18	20	
55 (15)	9	12	34		21 (12)	1	4	16	
76 (21)	24	40	12	< .001	42 (23)	15	20	7	< .001
136 (38)	41	66	29		66 (37)	14	36	16	
144 (41)	23	52	69		72 (40)	7	27	38	
185 (52)	50	102	33	< .001	90 (50)	22	50	18	< .001
95 (27)	26	39	30		57 (32)	10	29	18	
76 (21)	12	17	47		33 (18)	4	4	25	
51 (14)	14	33	4	< .001	44 (24)	11	25	8	< .001
173 (49)	52	89	32		73 (41)	17	40	16	
132 (37)	22	36	74		63 (35)	8	18	37	
307 (87)	78	142	87	.031	154 (86)	33	69	52	.475
49 (13)	10	16	23		26 (14)	3	14	9	
69 (19)	31	30	8	< .001	28 (16)	12	13	3	.010
51 (15)	12	18	21		19 (11)	1	9	9	
104 (29)	17	50	37		48 (27)	9	24	15	
132 (37)	28	60	44		85 (47)	14	37	34	
215 (60)	58	93	64	.473[c]	92 (51)	17	46	29	.564
141 (40)	30	65	46		88 (49)	19	37	32	
207 (58)	55	108	44	< .001	116 (64)	30	54	32	.012
149 (42)	33	50	66		64 (36)	6	29	28	
279 (79)	84	112	83	< .001	144 (80)	33	62	49	.104
77 (21)	4	46	27		36 (20)	3	21	12	

on the patient's diagnosis, as well as on the patient and disease characteristics. For uncomplicated patients, the type of treatment indicated was discussed between the clinician who performed the intake examination and a senior consultant for a final decision. More complex patients were discussed in the multidisciplinary team (including senior consultants, dentists, physiotherapists, and a psychologist), and the staff represented in the multidisciplinary team made the final decision together. The final decision-making of the senior consultants and the multidisciplinary team was based on their expertise, experience, and knowledge, as well as on the patient's specific signs and symptoms in both the physical and psychological aspects.

Statistical Analyses

Screening of Potential Predictors and Modeling. First, the bivariate association of each potential predictor with the three-category outcome (NT, PTO,

and CPPT) was tested by using the chi-square test for categorical predictors and Kruskal-Wallis test for continuous predictors. Predictors with a bivariate *P* value ≤ .20 were selected for inclusion in the subsequent multivariate multinomial logistic regression analyses. To prevent untoward restriction in number of variables, the candidate predictors were included or excluded (*P* > .20) in the final model based on backward-selection procedures by using SPSS software 21.0 (IBM). In the multinomial logistic regression analyses, NT was regarded as the reference outcome category. If a multinomial regression model has very diverse outcome categories (that is, each outcome category is associated with different predictors), this may make the model involve too many predictors and too many regression coefficients estimated. In order to limit and optimize the number of regression coefficients in a multinomial model, some predictors were only considered for one outcome category (*P* < .20) by setting the coefficient of the other outcome category to 0 when performing the procedures below.[18]

Internal Validation. A model that has been developed from a dataset to which it fits easily can result in overoptimism when applied to similar future patients.[19,20] To harness against such overfitting (ie, to improve the internal validity of the model), the regression coefficients of the predictors in the model were multiplied by a shrinkage factor.[19,20] A shrinkage factor ranges from 0 to 1. Bootstrapping and cross-validation techniques are commonly used to assess the internal validity of a model and to produce a shrinkage factor. Both approaches are based on resampling techniques to validate a model by using random subsets.[21,22] The shrinkage factor based on resampling is more accurate and stable than the shrinkage factor based on a heuristic formula, especially for models with small sample size or large predictors[23]; however, bootstrapping and cross-validation techniques of multinomial logistic regression models cannot be performed in a standardized manner with statistical software packages.[18] So, a heuristic shrinkage factor for multinomial logistic regression coefficients was applied in the present study, and was calculated as:

$$(\text{model}X^2 - df) / \text{model}X^2$$

. . . where modelX^2 indicates the likelihood ratio of the fitted model (ie, the difference in −2log-likelihood between the model with and without predictors), and *df* indicates the degrees of freedom of the number of candidate predictors considered for the model.[18,24]

Calibration. Calibration is defined as the agreement between the predicted outcomes and observed outcomes.[25] Calibration is assessed by plotting the predicted individual probabilities against the observed actual probabilities for each outcome category. For assessment of calibration, study participants were separately grouped into deciles based on their predicted probability for treatment. The prevalence of the end point within each decile represents the observed probability. In the calibration plot, the actual and predicted probabilities were compared across the range of predicted risks, and the calibration of the multivariate models was evaluated using the Pearson goodness-of-fit statistic. If the *P* value of the Pearson goodness-of-fit statistic test was > .05, it indicated no or low evidence for lack of fit of the model.[26–28]

Discrimination. Discrimination is defined as the ability to differentiate between those with and those without the outcome event.[25] For the binary or ordinal logistic models, it is common to use the C statistic, or area under the receiver operating characteristic curve (AUC), as the single measure for discrimination.[18] However, for the multinomial logistic regression analyses with three unordered categories in the present study, discrimination was assessed with three AUC values, relating one outcome category to the other two outcome categories in each receiver operating characteristic curve (ROC) area.[18] An AUC of 0.70 to 0.80 indicates the discrimination of the prediction model is acceptable, while an AUC value of ≥ 0.8 indicates the discrimination of the prediction model is excellent to outstanding.[29]

Scoring System. A clinical prediction rule for the type of treatment indicated for TMD patients was developed to provide an estimate for individual patients of their absolute risk of a certain type of treatment indicated. For the final multivariate multinomial logistic regression model, patients with NT were regarded as the reference category of the outcome, so the probability (P) of receiving PTO and CPPT was predicted using the formulae[18] below:

$$P_{PTO} = \frac{\exp(LP_{PTO})}{1 + \exp(LP_{cppt}) + \exp(LP_{pto})}$$

$$P_{CPPT} = \frac{\exp(LP_{CPPT})}{1 + \exp(LPCPPT) + \exp(LPPTO)}$$

. . . where LP_{PTO} = linear predictor of PTO = $\beta 0_{PTO} + \beta 1_{PTO}X_1 + \ldots + \beta i_{PTO}X_i$; and LP_{CPPT} = linear predictor of CPPT = $\beta 0_{CPPT} + \beta 1_{CPPT}X_1 + \ldots + \beta i_{CPPT}X_i$.

As the sum of the predicted probabilities of each outcome category in any logistic model is 1, the probability of receiving NT can be predicted with the formula $P_{NT} = 1 - P_{PTO} - P_{CPPT}$.[18] Patients were allocated to the outcome category with the highest predicted probability.

To facilitate the calculation of the probabilities of NT, PTO, and CPPT in individual patients separately,

the multinomial regression model was converted to a score chart. The score of each included predictor in the score chart was produced by the shrunken regression coefficients being divided by the smallest regression coefficient of the predictors and subsequently rounded. Line charts were then developed, which helped to determine the predicted probability of NT, PTO, or CPPT.

All the statistical procedures mentioned above were based on the derivation cohort. The discrimination, calibration, and scoring system of the model were all assessed based on the shrunken regression coefficients.

External Validation. To assess the general applicability of the multinomial model, the shrunken multinomial model was validated in a new sample of TMD patients based on the DC/TMD. These patients received an intake examination at the clinic of the Department of Oral Kinesiology of the Academic Centre for Dentistry (ACTA) between February 2016 and January 2017. The new sample was regarded as the validation cohort. The inclusion criteria for patients were the same as for the derivation cohort. In the validation cohort, the predicted probability for type of treatment indicated for each TMD patient was calculated based on the developed multinomial model in the derivation cohort described above. The validity of the validated multinomial model was also assessed in aspects of calibration (calibration plots and Pearson goodness-of-fit statistic), discrimination (AUC), and internal validity (shrinkage factor). All the statistical analyses were performed with SPSS software 21.0 (IBM).

Sample Size Estimation. For multinomial logistic regression analyses, the prediction model is considered reliable if the number of events per variable (EPV) is ≥ 20.[30,31] The number of events indicates the number of patients in the smallest group among the outcome categories. The total number of variables is calculated as the number of the continuous predictors plus the number of categories (without the reference category) for categorical predictors included in the multivariate multinomial regression analysis.

Results

A total of 356 patients were enrolled in the derivation cohort; 78% were female and 22% were male. The mean age ± standard deviation (SD) was 44 ± 15 years for female patients and 40 ± 14 years for male patients. The distribution of potential predictors based on type of treatment indicated for TMD patients is presented in Table 1, along with the bivariate associations between the potential predictors and the types of treatment. All predictors except for "Have you undergone treatment for the complaint previously?" (P = .473) were significantly associated with the types of treatment indicated for TMD patients according to the chi-square or Kruskal-Wallis test. Therefore, 16 predictors were selected for possible inclusion in the multivariate multinomial logistic regression analyses using the backward-selection procedure (P > .20).

Table 2 presents the predictors included in the final model based on the multivariate multinomial logistic regression analysis. When PTO was compared to NT, age, sleep bruxism, pain-related TMD, function-related TMD, stress, stretch of assisted mouth opening, and depression were significantly associated with the outcome. When CPPT was compared to NT, age, anxiety, sleep bruxism, pain-related TMD, function-related TMD, gender, stress, and stretch of assisted mouth opening were significantly associated with the outcome.

The −2log-likelihoods of intercept only and of the final model were 752.736 and 54 0.178, respectively. The degrees of freedom of the number of candidate predictors in the model were 24. So, the shrinkage factor was 0.89.

Figure 1 shows the calibration plot of the model. Most plotted points in the three outcome categories were lying close to the diagonal line, which indicates that there was good fit between the predicted probability and actual probability of type of treatment indicated for TMD patients in the model. With a resulting P value for the Pearson goodness-of-fit test of .29, the multinomial model was shown to be a good fit. The AUC values for NT, PTO, and CPPT were 0.86 (95% CI: 0.82 to 0.90), 0.76 (95% CI: 0.71 to 0.81), and 0.83 (95% CI: 0.79 to 0.87), respectively, which showed very good discrimination of the model (Fig 2).

To enhance its clinical usefulness, the final multinomial regression model was transformed into a score chart based on the shrunken regression coefficients (Table 3). Based on the score chart, a doctor can calculate the sum scores for PTO and for CPPT separately based on the predictors. Then, a clinician can determine the corresponding predicted probability of PTO and CPPT based on their sum scores and the total sum score (the sum of both sum scores of PTO and CPPT) by using Fig 3. Finally, the corresponding predicted probability of NT can be calculated by 1 minus the corresponding predicted probability of PTO minus the corresponding predicted probability of CPPT.

The distribution of the predictors for the validation cohort is also presented in Table 1. The calibration, discrimination, and overall performance of the model based on the validation cohort were shown to be moderate to good and similar to those based on the derivation cohort (Figs 1 and 2, Table 4).

Table 2 Multivariate Multinomial Logistic Regression Analyses (*P* < .20 after Backward Selection) Based on Type of Treatment Indicated for TMD Patients (n = 356)

	Physical treatment only vs no treatment[a]			
Predictors	β (SE)	Shrunken β	OR (95% CI)	*P* value
Age	−0.036 (0.011)	−0.032	0.965 (0.943, 0.986)	.002
GAD-7				
No anxiety	Reference			
Mild anxiety	−0.113 (0.464)	−0.101	0.807 (0.360, 2.218)	.807b
Moderate/severe anxiety	0.040 (0.705)	0.036	1.041 (0.261, 4.146)	.954b
Sleep bruxism				
No	Reference			
Yes	1.178 (0.346)	1.048	3.248 (1.650, 6.392)	.001
Pain-related TMD				
No	Reference			
Yes	1.415 (0.334)	1.259	4.116 (2.140, 7.913)	< .001
Function-related TMD				
No	Reference			
Yes	1.075 (0.451)	0.957	2.929 (1.210, 7.091)	.017
Gender				
Male	Reference			
Female	0.434 (0.372)	0.386	1.543 (0.745, 3.195)	.243b
Stress				
High	Reference			
Low	−0.037 (0.447)	−0.033	0.963 (0.401, 2.313)	.934b
No stress	0.827 (0.610)	0.736	2.287 (0.691, 7.564)	.175
Stretch of assisted mouth opening				
< 5 mm	Reference			
≥ 5 mm	3.140 (0.826)	2.795	23.110 (4.575, 116.737)	< .001
PHQ-9				
Moderate-severe depression	Reference			
Mild depression	1.206 (0.608)	1.073	3.338 (1.014, 10.987)	.047
No depression	1.312 (0.621)	1.168	3.714 (1.100, 12.540)	.035
Intercept	−1.221 (0.899)			.175

[a] Reference category.
[b] Coefficients of these categories were regarded as 0 when calculating the predicted probability and developing the score chart.
GAD-7 = Generalized Anxiety Disorder; TMD= temporomandibular disorders; PHQ-9 = Patient Health Questionnaire (9 items); β = regression coefficient; SE = standard error; OR = odds ratio; CI = confidence interval.

Discussion

The present study has shown that when PTO was compared to NT, patients with younger age, sleep bruxism, pain-related TMD, function-related TMD, less severe stress, larger stretch of assisted mouth opening, and less severe depression were more likely to receive PTO. Furthermore, when CPPT was compared to NT, female patients with younger age, more severe anxiety, sleep bruxism, pain-related TMD, function-related TMD, more severe stress, and larger stretch of assisted mouth opening were more likely to receive CPPT.

As shown in Table 2, most predictors for PTO and CPPT (based on NT as the reference category) in the final model were the same, including age, sleep bruxism, pain-related TMD, function-related TMD, stress, and stretch of assisted mouth opening. However, gender was significantly associated with the outcome CPPT vs NT, which indicates that female patients had 2.3 times higher odds of receiving CPPT compared to male patients, while gender was not significantly associated with the outcome PTO vs NT. Similarly, anxiety was significantly associated with the outcome CPPT vs NT but not with the outcome PTO vs NT, while depression was significantly associated with the outcome PTO vs NT but not with the outcome CPPT vs NT. Female TMD patients with more severe anxiety were more likely to receive CPPT than NT, while TMD patients with less severe depression were more likely to receive PTO than NT. Furthermore, some predictors played opposite roles in predicting PTO and CPPT. For example, when PTO was compared to NT, patients with less severe stress were more likely to receive PTO; however, when CPPT was compared to NT, patients with more severe stress were more likely to receive CPPT. In effect, this means that TMD patients with higher stress were more likely to receive CPPT and TMD patients with lower stress were more likely to receive PTO. Moreover, the weights of predictors in predicting different types of treatment indicated for

	Combined physical + psychological treatment vs no treatment[a]			
	β (SE)	Shrunken β	OR (95% CI)	*P* value
	−0.037 (0.013)	−0.033	0.964 (0.940, 0.989)	.005
	Reference			
	0.893 (0.495)	0.795	2.442 (0.926, 6.439)	.071
	0.901 (0.701)	0.802	2.463 (0.618, 9.808)	.200
	Reference			
	1.431 (0.422)	1.274	4.182 (1.828, 9.567)	.001
	Reference			
	2.247 (0.399)	2.000	9.456 (4.328, 20.664)	< .001
	Reference			
	1.018 (0.496)	0.906	2.768 (1.047, 7.313)	.040
	Reference			
	0.824 (0.465)	0.733	2.278 (0.917, 5.682)	.076
	Reference			
	−0.164 (0.706)	−0.146	0.849 (0.213, 3.390)	.817b
	1.037 (0.791)	0.923	2.819 (0.599, 13.282)	.190
	Reference			
	2.295 (0.849)	2.043	9.924 (1.878, 52.446)	.007
	Reference			
	−0.318 (0.479)	−0.283	0.727 (0.284, 1.861)	.507b
	−0.274 (0.627)	−0.244	0.760 (0.222, 2.598)	.662b
	−2.074 (0.979)			.034

TMD patients were different. For example, pain-related TMD was positively associated with both outcomes PTO vs NT and CPPT vs NT; however, patients with pain-related TMD had 4.1 times higher odds of receiving PTO than those without pain-related TMD, while patients with pain-related TMD had 9.5 times higher odds of receiving CPPT than those without pain-related TMD. This indicates that TMD patients with pain-related TMD are very likely to receive certain types of treatment and are more likely to receive CPPT than PTO. Likewise, TMD patients with function-related TMD and larger stretch of assisted mouth opening are more likely to receive PTO than CPPT, while TMD patients with sleep bruxism are more likely to receive CPPT than PTO.

Based on the common and easily obtainable clinical variables mentioned above, the present study derived a multinomial model to predict the three types of treatment indicated for TMD patients (NT, PTO, and CPPT). To the authors' knowledge, this is the first prediction model presented to support decision-making in patient management, in patients with TMD in particular. The AUC values of the multinomial model in the derivation cohort ranged from 0.76 to 0.86, which indicates that the discriminative ability of the model is acceptable to excellent in clinical practice. For NT, the AUC value of 0.86 means that given 100 discordant pairs (ie, in which one was indicated to have NT and the other was indicated to have any of the other two treatments [PTO or CPPT]), the model could correctly discriminate 86% of them. For PTO, the AUC value of 0.76 meant that given 100 discordant pairs, the model could correctly discriminate 76% of them. For CPPT, the AUC value of 0.83 meant that given 100 discordant pairs, the model could correctly discriminate 83% of them. Furthermore, the calibration plots showed that the model was reasonably calibrated in general, and Pearson goodness-of-fit test showed that the model had good fit. Therefore, the performance of the prediction model in terms of discrimination and calibration in the present study was good. The external validation with respect to internal validity, discrimination, calibration, and overall performance of the model was acceptable based on the validation cohort, which indicates that the generalizability of the multinomial model is acceptable. This showed that the multinomial model can be applied well to new TMD patients in ACTA for predicting type of treatment indicated. For other clinics or hospitals, the rules for decision-making of treatment may not be the same, so the model needs to be tested and adjusted elsewhere to determine its generalizability. However, the present profiling of the model can provide clinicians with important clues toward the factors that may be taken into consideration when making patient-tailored treatment plans for TMD patients.

For clinicians, it is important to know whether the TMD patient should receive treatment, and if so, whether this should be physical treatment only or combined with psychological treatment. The present model provides information for clinicians on the patient profiles that relate to the type of treatment indicated for TMD patients. In addition, the model provides guidance for novices, junior clinicians, and solo clinicians when they make decisions on which treatment is indicated for a TMD patient. This may make their decision-making more accurate, transparent, and consistent. Also, this model may be helpful for clinicians to find their own potential errors of decision-making in TMD patient care. The model may thus help to optimize TMD patient care.

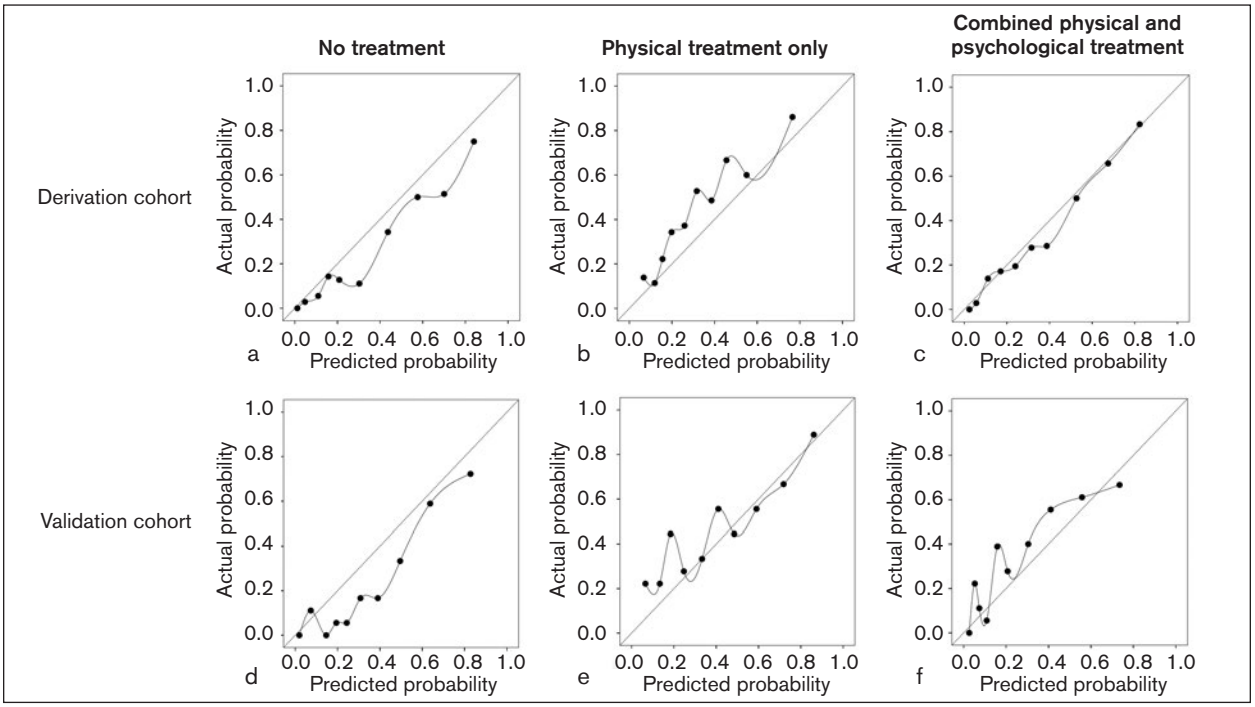


Fig 1 Calibration plots of the multinomial regression model for predicted and actual probability of the three outcome categories in TMD patients in **(a–c)** the derivation cohort (n = 356) and **(d–f)** the validation cohort (n = 180). **(a)** No treatment in the derivation cohort (number of patients with actual no treatment is 88, while that with predicted no treatment is 135). **(b)** Physical treatment only in the derivation cohort (number of patients with actual physical treatment only is 158, while that with predicted physical treatment only is 115). **(c)** Combined physical and psychological treatment in derivation cohort (number of patients with actual combined physical and psychological treatment is 110, while that with predicted combined physical and psychological treatment is 106). **(d)** No treatment in the validation cohort (number of patients with actual no treatment is 36, while that with predicted no treatment is 53). **(e)** Physical treatment only in the validation cohort (number of patients with actual physical treatment only is 83, while that with predicted physical treatment only is 79). **(f)** Combined physical and psychological treatment in the validation cohort (number of patients with actual combined physical and psychological treatment is 61, while that with predicted combined physical and psychological treatment is 48). The diagonal line is what would result if the predicted probability of the model was the same as the actual probability of the model so that the prediction is neither underestimated nor overestimated. The dots represent the deciles of the study members based on their predicted probability.

Table 3 Score Chart of the Multinomial Model for Prediction of Type of Treatment Indicated for TMD Patients (n = 356)

	Physical treatment only		Combined physical ± psychological treatment	
Predictors	Value	Score	Value	Score
Age	≥ 18 y	100	≥ 18 y	100
Sleep bruxism	No	0	No	0
	Yes	33	Yes	39
Pain-related TMD	No	0	No	0
	Yes	39	Yes	61
Function-related TMD	No	0	No	0
	Yes	30	Yes	27
Stress	Somewhat/quite/much/very much	0	No stress/somewhat	0
	No stress	23	Quite/much/very much	28
Stretch of assisted mouth opening	< 5 mm	0	< 5 mm	0
	≥ 5 mm	87	≥ 5 mm	62
PHQ-9	Moderate-severe depression	0		
	Mild depression	34		
	No depression	37		
GAD-7			No anxiety	0
			Mild/moderate/severe anxiety	24
Gender			Male	0
			Female	22
Sum score:				

Total sum score (St) = physical score (Sp) + combined score (Sc).
This score chart can be used to calculate the sum scores for physical treatment only and the sum scores for combined physiclal and psychological treatment.
The total sum score equals the sum score for physical treatment only + the sum score for combined physical and psychological treatment.
TMD = temporomandibular disorders; PHQ-9 = Patient Health Questionnaire (9 items); GAD-7 = Generalized Anxiety Disorder.

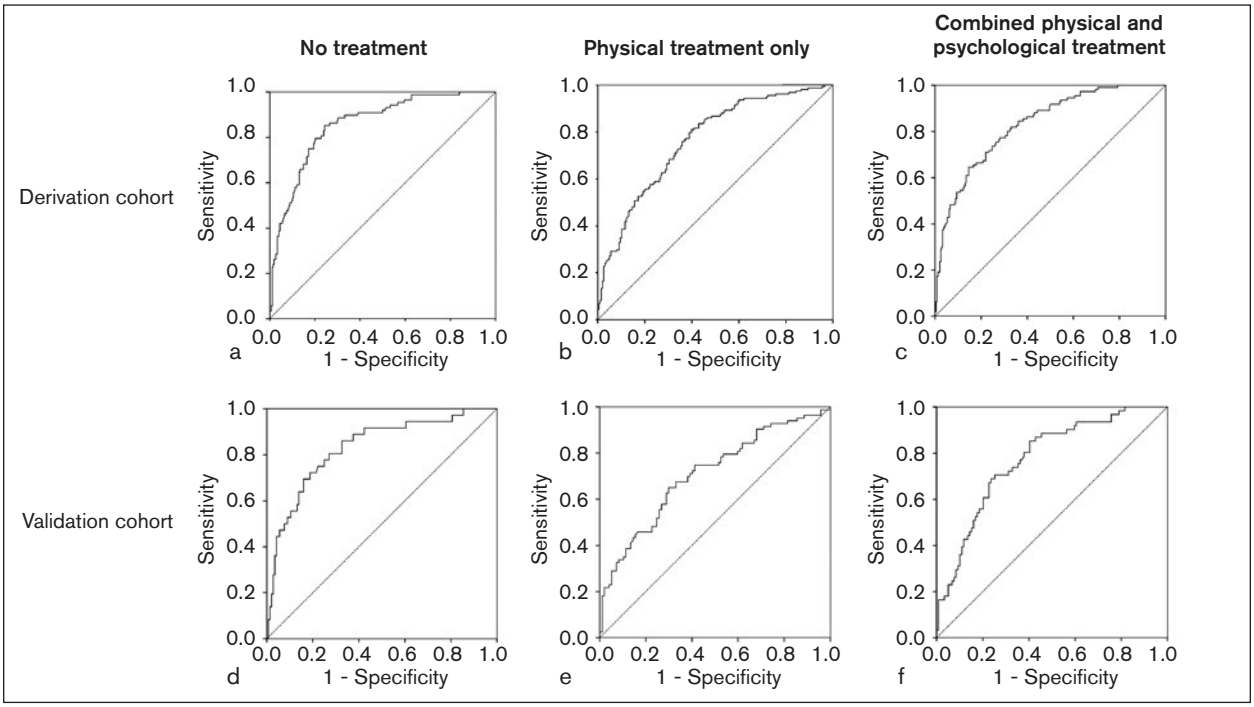


Fig 2 Discriminative ability of the multinomial regression model for prediction of the three outcome categories in TMD patients in **(a–c)** derivation cohort (n = 356) and **(d–f)** validation cohort (n = 180). **(a)** ROC areas of no treatment vs physical treatment only + combined physical and psychological treatment in derivation cohort with an AUC of 0.86 (95% CI: 0.82 to 0.90). **(b)** ROC areas of physical treatment only vs no treatment + combined physical and psychological treatment in derivation cohort with an AUC of 0.76 (95% CI: 0.71 to 0.81). **(c)** ROC areas of combined physical and psychological treatment vs no treatment + physical treatment only in the derivation cohort with an AUC of 0.83 (95% CI: 0.79 to 0.87). **(d)** ROC areas of no treatment vs physical treatment only + combined physical and psychological treatment in the validation cohort with an AUC of 0.83 (95% CI: 0.75 to 0.91). **(e)** ROC areas of physical treatment only vs no treatment + combined physical and psychological treatment in validation cohort with an AUC of 0.71 (95% CI: 0.63 to 0.78). **(f)** ROC areas of combined physical and psychological treatment vs no treatment + physical treatment only in the validation cohort with an AUC of 0.77 (95% CI: 0.70 to 0.84).

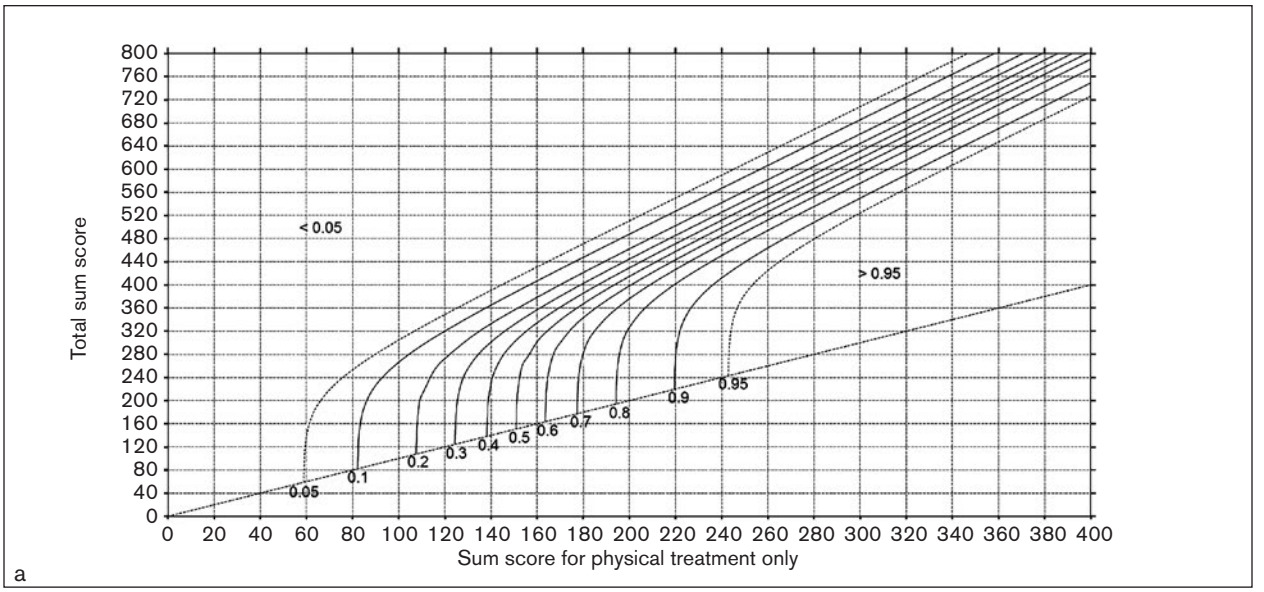
Table 4 Discrimination, Calibration, Internal Validity, and Performance of Derivation Cohort (n = 356) and Validation Cohort (n = 180) Models

	AUC (95% CI)			Pearson goodness-of-fit test		Shrinkage factor	Nagelkerke R^2
	NT	PTO	CPPT	Chi-square	*P* value		
Derivation cohort	0.86 (0.82, 0.90)	0.76 (0.71, 0.81)	0.83 (0.79, 0.87)	669.665	.288	0.89	0.51
Validation cohort	0.83 (0.75, 0.91)	0.71 (0.63, 0.78)	0.77 (0.70, 0.84)	270.248	.326		0.58

AUC = area under the curve; CI = confidence interval; NT = no treatment; PTO = physical treatment only; CPPT= combined physical and psychological treatment.

The present study simply tried to weigh the relative contributions of different predictors to the types of treatment indicated for TMD patients. The relative contribution of each predictor was expressed as a weight (included in the score chart [Table 3]) that can be used to calculate a possibility of each of the three outcome categories for an individual. Table 5 presents an example of how to use the score chart. In this example, a male patient with TMD came to the clinic to seek treatment. He was 40 years old with the presence of sleep bruxism, pain-related TMD, function-related TMD, stretch of assisted mouth opening of 8 mm, mild depression, mild anxiety, and no stress. So, this patient had a sum score for PTO of 306 and a sum score for CPPT of 274. The total sum score of this patient was 580. Then, based on the line charts (Fig 3), the predicted probability for PTO can be estimated to be around 84% and that for CPPT around 16%. The predicted probability for NT can be calculated by 1 – 84% – 16% = 0%. Therefore, the clinician can make a decision that PTO is indicated for this patient, because the predicted probability for PTO was the highest.

Sample size is typically a severe problem for multinomial regression models because one or more of the outcome categories often has very low



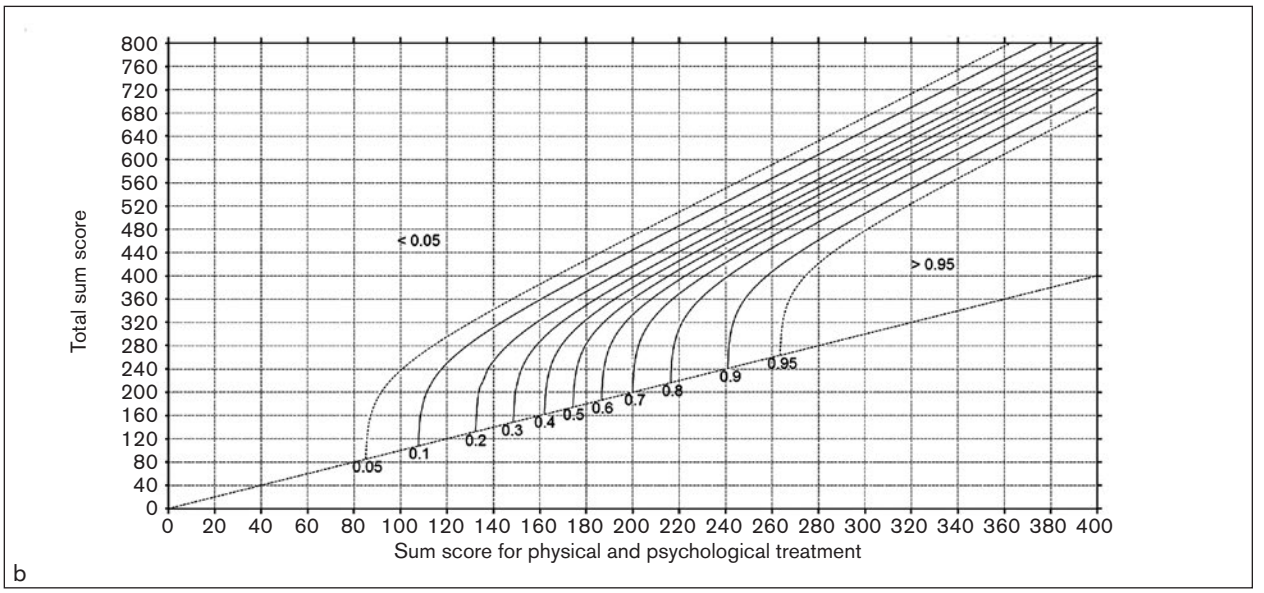


Fig 3 Line charts of the multinomial model for determining the predicted probability of **(a)** physical treatment only and **(b)** combined physical and psychological treatment. The sum scores for physical treatment only and for combined physical and psychological treatment are presented on the horizontal axes, and the total sum score on the vertical axes. The diagonal line represents the sum score for the outcome category that is the same as the total sum score (y = x). The cross point of a vertical line drawn from the x axis and a horizontal line drawn from the y axis shows the corresponding predicted probability of the outcome category. The corresponding predicted probability of no treatment can be calculated by 100% minus the predicted probability of physical treatment only minus the predicted probability of combined physical and psychological treatment.

prevalence, which necessitates the availability of large datasets in order to have a sufficient number of events per variable (EPV) for such categories.[32] For multinomial regression models, the power is determined by the number of patients in the smallest group.[18] For the dichotomous model, a rule of thumb is that EPV should be ≥ 10.[30,31] In the present multinomial model with three outcome categories, each predictor has been estimated twice and has two regression coefficients, so the EPV should be ≥ 20.[18] The present study did not meet this criterion because of the small sample size, which was a study limitation. Moreover, in the derivation of the model, the study deviated from using the conventional *P* value of .05 as the threshold for statistical significance. Instead, a less stringent threshold of .20 was used in both the bivariate chi-square tests (or Kruskal-Wallis tests) and the multivariate regression analyses in the selection and exclusion of potential predictors. This could avoid false negative findings in both modeling stages,

especially when the sample size is small, which could cause unjustified exclusion of independent predictors from the final model.[33] Also, the less stringent *P* value could to a large extent avoid a situation in which clinically important predictors with a low prevalence are excluded from the model due to the high *P* value. In addition, before modeling, the choice of potential predictors was decided by the consensus of the multidisciplinary team. Therefore, all the potential predictors in the present study are deemed clinically relevant, so it is impossible that predictors were included with no clinical relevance but a very low *P* value. After this initial selection of clinically relevant predictors, backward selection was used to select the potential predictors. So, theoretically, it is possible that predictors with too low prevalence of the event may not be included in the final model even though they are clinically important, which may be regarded as a limitation of the study. However, in the present study, the prevalence of all the predictors was larger than 10% (Table 1), indicating this possible limitation did not occur. In the future, hierarchical modeling procedures are recommended to further avoid this risk.

Due to the absence of follow-up data, this model can only be used to predict type of treatment indicated for TMD patients rather than treatment outcomes. Whether the treatment that was indicated for TMD patients selected from the present model truly resulted in an optimal treatment outcome in the follow-ups in the TMD patients still needs to be confirmed with follow-up data. Therefore, further studies should assess whether the type of treatment indicated for TMD patients in the present study are related to optimized treatment outcomes in the follow-ups.

Conclusions

Potential predictors in patient profiles for prediction of type of treatment indicated for TMD patients were identified. The multinomial regression model for type of treatment indicated for TMD patients (NT, PTO, or CPPT) was developed and validated, and the score chart and line charts may assist clinicians in decision-making regarding which type of treatment is indicated for an individual TMD patient.

Acknowledgments

This work was funded by the China Scholarship Council: http://www.csc.edu.cn/. The funders had no role in the study design, data collection and analyses, decision to publish, or preparation of the manuscript. The authors declare no potential conflicts of interest with respect to the authorship and the publication of this article.

Table 5 Example of How to Use the Score Chart

Variables	PTO	CPPT
Similar predictors		
Age (40 y)	60	60
Sleep bruxism (yes)	33	39
Pain-related TMD (yes)	39	61
Function-related TMD (yes)	30	28
Stress (no stress)	23	0
Stretch of assisted mouth opening (yes)	87	62
Different predictors		
PHQ-9 (mild depression)	34	–
GAD-7 (mild anxiety)	–	24
Gender (male)	–	0
Sum score	306	274
Total sum score	580	
Predicted probability	84%	16%
NT	0%	

PTO = physical treatment only; CPPT = combined physical and psychological treatment; TMD= temporomandibular disorders; PHQ-9 = Patient Health Questionnaire (9 items); GAD-7 = Generalized Anxiety Disorder; NT = no treatment.

References

1. Sommer I, Lavigne G, Ettlin DA. Review of self-reported instruments that measure sleep dysfunction in patients suffering from temporomandibular disorders and/or orofacial pain. Sleep Med 2015;16:27–38.
2. Durham J, Newton-John TR, Zakrzewska JM. Temporomandibular disorders. BMJ 2015;350:h1154.
3. Kotiranta U, Suvinen T, Forssell H. Tailored treatments in temporomandibular disorders: Where are we now? A systematic qualitative literature review. J Oral Facial Pain Headache 2014;28:28–37.
4. Naeije M, Te Veldhuis AH, Te Veldhuis EC, Visscher CM, Lobbezoo F. Disc displacement within the human temporomandibular joint: A systematic review of a 'noisy annoyance'. J Oral Rehabil 2013;40:139–158.
5. Durham J, Al-Baghdadi M, Baad-Hansen L, et al. Self-management programmes in temporomandibular disorders: Results from an international Delphi process. J Oral Rehabil 2016;43:929–936.
6. List T, Jensen RH. Temporomandibular disorders: Old ideas and new concepts. Cephalalgia 2017;37:692–704.
7. Kay E, Nuttall N. Clinical decision making—An art or a science? Part II: Making sense of treatment decisions. Br Dent J 1995;178:113–116.
8. Lutfey KE, Campbell SM, Renfrew MR, Marceau LD, Roland M, McKinlay JB. How are patient characteristics relevant for physicians' clinical decision making in diabetes? An analysis of qualitative results from a cross-national factorial experiment. Soc Sci Med 2008;67:1391–1399.
9. Schiffman E, Ohrbach R, Truelove E, et al. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for clinical and research applications: Recommendations of the International RDC/TMD Consortium Network and Orofacial Pain Special Interest Group. J Oral Facial Pain Headache 2014;28:6–27.
10. Gonzalez YM, Schiffman E, Gordon SM, et al. Development of a brief and effective temporomandibular disorder pain screening questionnaire: Reliability and validity. J Am Dent Assoc 2011;142:1183–1191.
11. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133–149.
12. Markiewicz MR, Ohrbach R, McCall WD Jr. Oral behaviors checklist: Reliability of performance in targeted waking-state behaviors. J Orofac Pain 2006;20:306–316.

13. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch Intern Med 2006;166:1092–1097.
14. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64:258–266.
15. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–613.
16. Van der Meulen MJ, Lobbezoo F, Aartman IH, Naeije M. Ethnic background as a factor in temporomandibular disorder complaints. J Orofac Pain 2009;23:38–46.
17. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1992;15:376–381.
18. Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KG. Polytomous logistic regression analysis could be applied more often in diagnostic research. J Clin Epidemiol 2008; 61:125–134.
19. Barrett TW, Martin AR, Storrow AB, et al. A clinical prediction model to estimate risk for 30-day adverse events in emergency department patients with symptomatic atrial fibrillation. Ann Emerg Med 2011;57:1–12.
20. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–781.
21. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med 1990;9:1303–1325.
22. Copas JB. Cross-validation shrinkage of regression predictors. J R Stat Soc Series B Stat Methodol 1987;45:175–183.
23. Preacher KJ. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008;40:879–891.
24. Steyerberg EW (ed). Validation of prediction models. In: Clinical Prediction Models: A Practice Approach to Development, Validation, and Updating. Berlin: Springer Science & Business Media, 2009:299–311.
25. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 2010;21:128–138.
26. Lee EC, Park SJ, Han SS, et al. Risk prediction of post-hepatectomy liver failure in patients with perihilar cholangiocarcinoma. J Gastroenterol Hepatol 2018;33:958–965.
27. Hicks KE, Zhao Y, Fallah N, et al. A simplified clinical prediction rule for prognosticating independent walking after spinal cord injury: A prospective study from a Canadian multicenter spinal cord injury registry. Spine J 2017;17:1383–1392.
28. Horn SD. Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale. Biometrics 1977; 33:237–247.
29. Hosmer DW, Lemeshow S (eds). Assessing the fit of the model. In: Applied Logistic Regression, ed 2. New York: John Wiley & Sons, 2000:143–202.
30. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373–1379.
31. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol 2016;76:175–182.
32. Van Hoorde K, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW, Van Calster B. Simple dichotomous updating methods improved the validity of polytomous prediction models. J Clin Epidemiol 2013;66:1158–1165.
33. Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, Hadaegh F, Steyerberg EW, Khalili D. A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. J Clin Epidemiol 2016;71:76–85.