Kai Becker

K. Becker[a], H. A. Jakstat[b], M. O. Ahlers[c]

Quality improvement of functional diagnostics in dentistry through computer-aided diagnosis: a randomized controlled trial

Abstract

Functional diagnostic examinations such as clinical functional analysis and manual structural analysis (´orthopedic tests´) allow the dentist to establish a structured diagnosis. Previously, the process of correlating findings with the appropriate diagnoses was guided by human thought processes alone. The experimental diagnostic randomized controlled trial (RCT) in this study investigated whether computer-aided diagnosis (CADx) of temporomandibular disorders (TMD) offers quality advantages over traditional diagnosis (TRAD).

Subjects and methods: Thirty-nine 5th-year dental students (examiners) at a university in Hamburg, Germany, received joint training in the diagnosis of TMD by clinical functional analysis and manual structural analysis (´orthopedic tests´). This study is based on anonymized data from 10 patients who were consecutively recruited at a specialized TMJ treatment center. The examiners were randomly allocated to two groups. Each examiner established a structured diagnosis through a traditional diagnostic method and by computer-aided diagnosis (CMDfact 4 functional diagnostics software) of five cases, each using the AB/BA crossover design. The diagnoses established by each individual examiner were then compared with the corresponding reference diagnoses (gold standard) and with those of the other examiners.

Results: Cohen's kappa coefficient analysis showed that median agreement with the reference diagnoses was significantly higher ($P < 0.001$) with computer assistance (median 0.692) than without it (0.553). Fleiss' kappa showed that the median interexaminer consistency of diagnoses was significantly higher ($P < 0.001$) with computer assistance (0.497) than with traditional diagnostic methods alone (0.271). Likewise, the number of false-positive and false-negative diagnoses was significantly lower with computer assistance ($P < 0.001$).

Conclusions: This study determined that dentists who are less experienced and not specialized in dental functional diagnostics achieve a significantly better and more consistent diagnostic quality with computer assistance by means of the system used in this study. Therefore, it seems advisable to extend computer-aided diagnostics to further functional examination techniques (condylar position analysis and jaw motion analysis).

Keywords: *computer-aided diagnostics, temporomandibular disorders (TMD), clinical functional analysis, manual structural analysis, orthopedic tests, randomized controlled trial, diagnostic classification, CMDfact*

a Kai Becker, Doctoral Candidate, Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf (UKE), Germany

b Prof. Dr. med. dent. Holger A. Jakstat, Department of Prosthetic Dentistry and Dental Materials, Center for Dental and Oral Medicine, University of Leipzig, Germany

c PD Dr. med. dent. M. Oliver Ahlers, Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf (UKE); and CMD-Center Hamburg-Eppendorf, Hamburg, Germany

Introduction

Computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems have been successfully implemented in many areas of medicine.[1,2] The range of applications for these technologies has grown tremendously in recent decades and now includes the diagnostic imaging of malignomas, among other things. Image processing systems used in dermatology, for example, analyze dark skin lesions and establish diagnoses by means of computer-aided image analysis.[3-5] Computer-aided triage systems that provide diagnostic mapping based on computed tomography (CT) angiograms are available for emergency management in cardiology.[6] A new trend is to merge findings from different diagnostic sources into one system to support general practitioners in establishing differential diagnoses based on this information.[7] These systems cannot replace the clinician, however.

Digital technology has been successfully adopted in many areas of dentistry. It is used, for example, in computer-aided impression-taking;[8] in the fabrication of restorations at chairside[9] and in the dental laboratory;[10] for digital tooth shade determination;[11] and for various formats of digital imaging, including intraoral dental radiographs,[12] panoramic tomography,[13] and cone beam computed tomography (CBCT).[14] These different techniques are now coalescing in navigated dental implantology with subsequent restorative dentistry.[15] Electronic systems for digitally acquired examination findings have also been introduced in the fields of periodontology,[16] endometry,[17] caries detection technology,[18] and instrumental motion tracking to provide a basis for individual articulator programming.[19,20] However, all of the above applications are merely systems for the electronic acquisition of examination findings – none of them perform computer-aided diagnosis.

In the case of craniomandibular disorders (CMD) and temporomandibular disorders (TMD), however, a computer-aided diagnosis system is already available. The diagnosis of CMD or TMD is generally established in a given sequence of examinations ('diagnostic cascade').

- First, TMD screening is performed to determine whether the patient's symptoms might be attributable to temporomandibular disorders (TMD). Special software is already available for the computer-aided analysis of TMD screening tests.[21]
- If the screening results are positive, clinical functional analysis is generally carried out as the next step.[22] Manual structural analysis (´orthopedic tests´) may also be performed as an additional examination, if necessary.[23]

The initial diagnosis is established based on the results of these examinations. Adequate differentiation of the global diagnosis of 'temporomandibular disorder' (TMD) or 'craniomandibular dysfunction' (CMD) is needed to adequately describe and treat the entities.[24]

Various diagnostic criteria for the classification of TMD or CMD have been published.[25-28] The diagnostic classification used in Germany was developed by specialists from several dental schools.[29] It was first presented at the annual meeting of the German Association for Functional Diagnostics and Therapy (DGFDT) in 2002, and has now been adopted as the official diagnostic classification.[30] Detailed classification criteria have been published for this diagnostic classification.[29]

In practice, this means that the dentist must first gather and document all relevant findings and then analyze them within the framework of a given diagnostic system to establish the correct diagnosis or diagnoses and make relevant treatment decisions. This baseline data lends itself to input into a computer-aided diagnostic system.[31] This is precisely why software-based systems for the documentation and interpretation of the findings from clinical functional analysis and manual structural analysis (´orthopedic tests´) have been developed and introduced in the past.[32-34] Software for the semi-automatic generation of examination reports based on this data is also available.[35] However, studies on the effect of using such computer-aided diagnostic systems on the quality of TMD diagnoses are still lacking.

Therefore, the aim of the present study was to develop and utilize a replicable study design to measure this effect. After establishing reference diagnoses based on anonymized patient records and findings from clinical examinations, the data were analyzed to determine whether examiners achieved better agreement with reference diagnoses with computer-aided diagnosis than with traditional diagnostic methods.

Materials and methods

Study design

The study was designed as a randomized controlled trial (RCT) with an AB/BA crossover design and standardized instruction of examiners, who used anonymized clinical patient data.

Prior to the investigation, the test conditions were evaluated in a pilot study in which 20 dentists diagnosed three patients each. Based on the results of this pilot study, test

parameters were set such as the number of examiners, the number of patients, and the time needed to establish the diagnosis.

Examiners

Forty dental students from the dental school at the University Hospital Hamburg-Eppendorf served as the examiners; all 40 were at the end of their final year in undergraduate dental education. They had maximal clinical experience as undergraduate students, but until then had no clinical experience in diagnosing TMD or CMD. Therefore, the examiners had not developed a preference (bias) for either traditional or computer-aided diagnosis.

In the framework of a 'Dies academicus' on 16 June 2017, all examiners received joint training in the interpretation of the findings from clinical functional analysis and manual structural analysis (´orthopedic tests´). This included the attribution of the respective findings to the diagnostic subgroups of the classification of TMD or CMD, respectively. The training also covered the proper documentation and the use of the computer-aided analysis tools.

Patients

The study included 10 consecutively recruited patients seen at a center that was specialized in the diagnosis and treatment of TMD or CMD, respectively.

- *Inclusion criteria:* Differentiated diagnosis of TMD/CMD established by clinical functional analysis and manual structural analysis (´orthopedic tests´).
- *Exclusion criteria:* Differential diagnoses consistent with symptoms mainly suggestive of non-dysfunctional causes (eg, tumor or neuralgia), as well as patients suffering from dysfunctional pain and TMD/CMD, in which the pain disorder overlies the typical TMD-like clinical picture.

The differentiation and exclusion of cases was performed based on the prior identification of patients with ≥ 3 Grade III disability according to the Grading Chronic Pain, German Version (GCS),[36] the German equivalent of the Graded Chronic Pain Scale (GCPS).[37] Within the scope of the functional diagnostic examinations, all patients completed the standard Chronic Pain Questionnaire designed for this purpose (dentaConcept Verlag, Hamburg).[38]

An anonymized patient chart was compiled for each of the 10 patients, and each examiner received a printout of the anonymized patient chart, which contained the following items:

- General and specific medical history.
- Intraoral photographs (high-resolution images printed on excellent-quality photographic paper).
- Panoramic radiographs (high-resolution images printed on excellent-quality photographic paper).
- Patient's clinical functional analysis report[39] (dentaConcept Verlag; Fig 1a).
- Patient's manual structural analysis report[40] (dentaConcept Verlag; Fig 1b).
- Printout of *possible* diagnoses within the TMD/CMD spectrum, as per the DGFDT diagnostic classification.

Randomization

Of the 40 examiners, 39 participated until the end of the study and submitted a consent form to participate in the anonymous collection and use of the data; only these 39 examiners participated in the subsequent diagnostic assessments and were included in the analysis.

The 39 examiners were randomly assigned to one of the following test groups:

- TRAD/CADx (traditional diagnosis first, followed by computer-aided diagnosis).
- CADx/TRAD (computer-aided diagnosis first, followed by traditional diagnosis).

Briefly, each examiner received a sealed randomization envelope. Each envelope contained a slip of paper specifying the examiner's ID number and test group (not legible from outside the envelope). To ensure the complete anonymity of data collection, each ID number was only disclosed to the respective individual examiner. After consultation with a specialized scientific advisory service (Clinical Trial Center North, Hamburg), ethics approval by the competent medical authority was not necessary, provided that complete anonymity of the patient data was maintained.

Of the 39 examiners recruited, 19 were assigned to one test group and 20 to the other.

Clinical Functional Analysis (4)

dentaConcept

Patient1 Patient01 — Patient
3.1.2017 — Examination date
1 — Patient ID
1.1.1987 — Date of birth

CMD-Centrum Hamburg-Eppendorf
Priv.-Doz. Dr. M. Oliver Ahlers
Falkenried 88 (CiM, Hs. C)
20251 Hamburg

Practice/office/clinic seal

Purpose This form documents the findings obtained by Clinical Functional Analysis. As a holistic approach, the form also includes sections for the screening on orthopaedic co-factors and the assessment of possible co-factors stress, anxiety and depression. These, however, constitute separate diagnostic procedures.

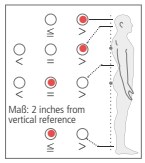
Development Dr. M.O. Ahlers, School of Dental Medicine, Universitätsklinikum Hamburg-Eppendorf; Prof. Dr. H.A. Jakstat, School of Dental Medicine, Universität Leipzig, Germany 1996-2003

Stress / DASS

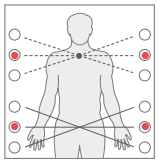
Score from **Stress Load Evaluation** 101 points
The stress load is…
≥ 200 ○ extreme ≥ 100 ◉ high < 100 ○ normal
OR
Scores from **DASS**
Depression points Anxiety points Stress points
Limit(s) surpassed on… ≥ 10 ☐ ≥ 6 ☐ ≥ 10 ☐

Screening for Orthopaedic Co-Factors

Posture lateral



Posture frontal



Function of CS
Rotation is limited in…

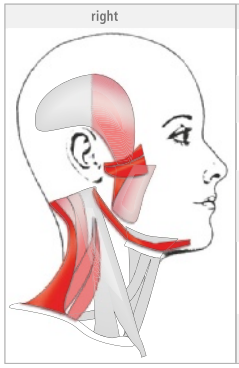
Inclination ○ L ◉ none ○ R
Elevation ○ L ◉ none ○ R
Reclination ○ L ○ none ○ R
☐ Discomfort ☐ Pain

☐ Change in occlusion upon postural correction
☐ Increased ligament elasticity

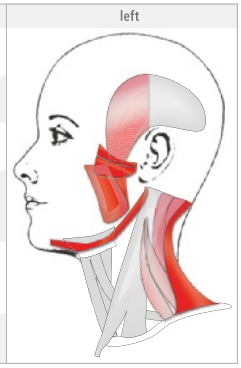
Parafunctions

☒ Indentations in tongue
☒ Alterations of buccal mucosa
☒ Severe abrasion … auf allen Zähnen
☐ Wedge-shaped defects
☐ Gingival recessions
☐ McCall's festoons
☐ Stillman's clefts
☐ Mobility of teeth

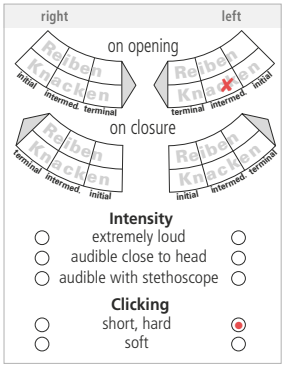
Palpation



		Discomfort	Pain (2marks)		
☐	☒	superficial masseter m.	☒	☒
☒	☒	profound masseter m.	☒	☒
☐	☐	TMJ lateral	☐	☐
☐	☐	TMJ dorsal	☐	☐
☐	☒	anterior temporalis m.	☒	☐
☐	☐	int./post. temporalis m.	☐	☐
☐	☒	subocc. muscles of neck	☒	☐
☒	☒	trapezius m.	☒	☒
☐	☐	sternodeidomastoideus m.	☐	☐
☐	☐	infrahyoid m.	☐	☐
☐	☐	suprahyoid m.	☐	☐
☐	☐	medial pterygoid m.	☐	☐
☒	☒	post. body of digastric m.	☒	☒
☒	☒	lateral pterygoid m.	☒	☒



TMD-Sounds

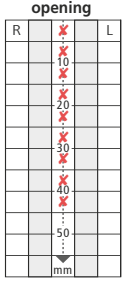


Intensity
○ extremely loud ○
○ audible close to head ○
○ audible with stethoscope ○

Clicking
○ short, hard ◉
○ soft ○

Motion

opening



Active opening 46 mm ☐ Discomfort ☐ Pain
Passive opening 47 mm ☐ Discomfort ☐ Pain
Location of **Pain** ☐ right ☐ left
Endfeel ○ soft ○ hard
Overjet (horiz.) 2 mm
Overbite (vert.) 2 mm

Clinical Evaluation of Mobility and Occlusion

Protrusion 6 mm ☐ Discomfort ☐ Pain
Laterotrusion R 15 mm ☐ Discomfort ☐ Pain
Laterotrusion L 14 mm ☐ Discomfort ☒ Pain
Slide: **centric relation** into intercuspal position 0,5 mm ☐ Discomfort ☐ Pain

if range > 0: ○ sagittal
Slide motion is ◉ deviating to R
○ deviating to L

Occlusion

Intercuspal position
◉ stable ○ unstable
☐ interincisal contact
☐ intermolar non-occlusion
☐ equilibrated with occlusal splint

Unphysiological occlusion contacts
☐ intercusp. pos. ……
☐ centric occl. ……
☐ protrusion ……
☐ laterotrusion ▸R ……
☐ laterotrusion ▸L ……

Reaction Tests

Krogh-Poulsen's provocation test
○ pos. ○ neg. …… applied test

Gerber's resilience test

right	Shimstock held with tin-foil	left	Shimstock held with tin-foil
0,0 left mm		0,0 right mm	
◉ compression		◉ compression	
○ neutral		○ neutral	
○ distraction		○ distraction	

☐ not applicable due to infraocclusion

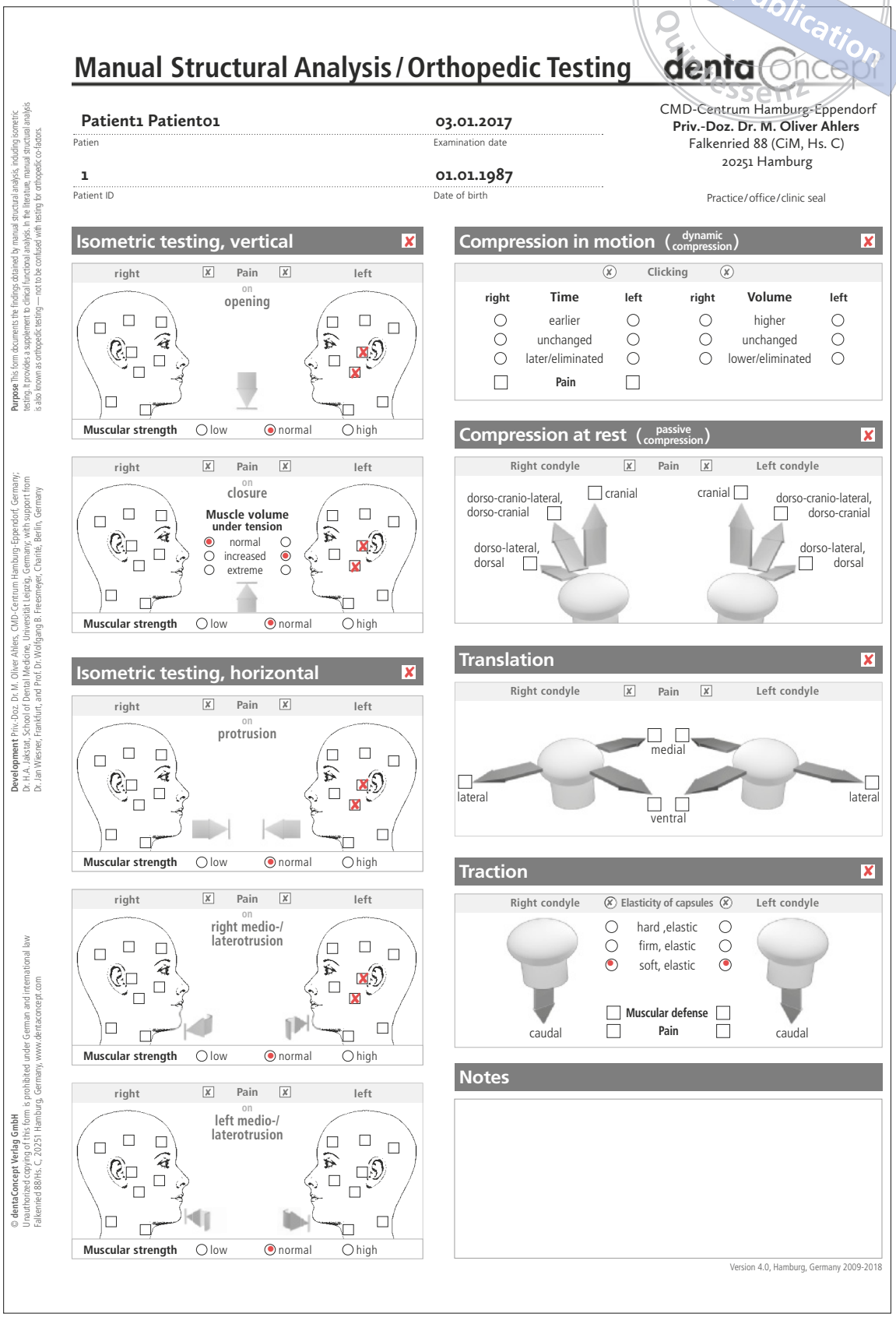
Notes

Notes are printed on a separate sheet.

Version 4.0, Hamburg, Germany 1996-2018

Fig 1a Findings from the clinical functional analysis for patient case 1 (form equivalent to the software module CMDstatus, ©dentaConcept Verlag, www.dentaconcept.de).

Please note: As an holistic approach, the form also includes sections for the screening of orthopedic co-factors as well as the assessment of the co-factors stress, anxiety and depression. These, however, constitute separate diagnostic procedures.



Manual Structural Analysis / Orthopedic Testing

denta Concept

Patient1 Patient01 — Patien
03.01.2017 — Examination date
1 — Patient ID
01.01.1987 — Date of birth

CMD-Centrum Hamburg-Eppendorf
Priv.-Doz. Dr. M. Oliver Ahlers
Falkenried 88 (CiM, Hs. C)
20251 Hamburg

Practice/office/clinic seal

Purpose This form documents the findings obtained by manual structural analysis, including isometric testing. It provides a supplement to clinical functional analysis. In the literature, manual structural analysis is also known as orthopedic testing — not to be confused with testing for orthopedic co-factors.

Development Priv.-Doz. Dr. M. Oliver Ahlers, CMD-Centrum Hamburg-Eppendorf, Germany; Dr. H.A. Jakstat, School of Dental Medicine, Universität Leipzig, Germany; with support from Dr. Jan Wiesner, Frankfurt, and Prof. Dr. Wolfgang B. Freesmeyer, Charité, Berlin, Germany

Isometric testing, vertical

right — Pain — left

on opening

Muscular strength: low / normal / high

on closure

Muscle volume under tension: normal / increased / extreme

Muscular strength: low / normal / high

Isometric testing, horizontal

on protrusion

Muscular strength: low / normal / high

on right medio-/laterotrusion

Muscular strength: low / normal / high

on left medio-/laterotrusion

Muscular strength: low / normal / high

Compression in motion (dynamic compression)

Clicking

right	Time	left	right	Volume	left
○	earlier	○	○	higher	○
○	unchanged	○	○	unchanged	○
○	later/eliminated	○	○	lower/eliminated	○
□	Pain	□			

Compression at rest (passive compression)

Right condyle — Pain — Left condyle

dorso-cranio-lateral, dorso-cranial; cranial; dorso-lateral, dorsal

Translation

Right condyle — Pain — Left condyle

medial; lateral; lateral; ventral

Traction

Right condyle — Elasticity of capsules — Left condyle

hard ,elastic / firm, elastic / soft, elastic

Muscular defense / Pain

caudal; caudal

Notes

Version 4.0, Hamburg, Germany 2009-2018

Fig 1b
Findings from the manual strutural analysis (´orthopedic tests´) for patient case 1 (form equivalent to the software module CMDstatus, ©dentaConcept Verlag, www.dentaconcept.de).

Please note: The ´orthopedic tests´ are not to be confused with the testing for orthopedic co-factors.

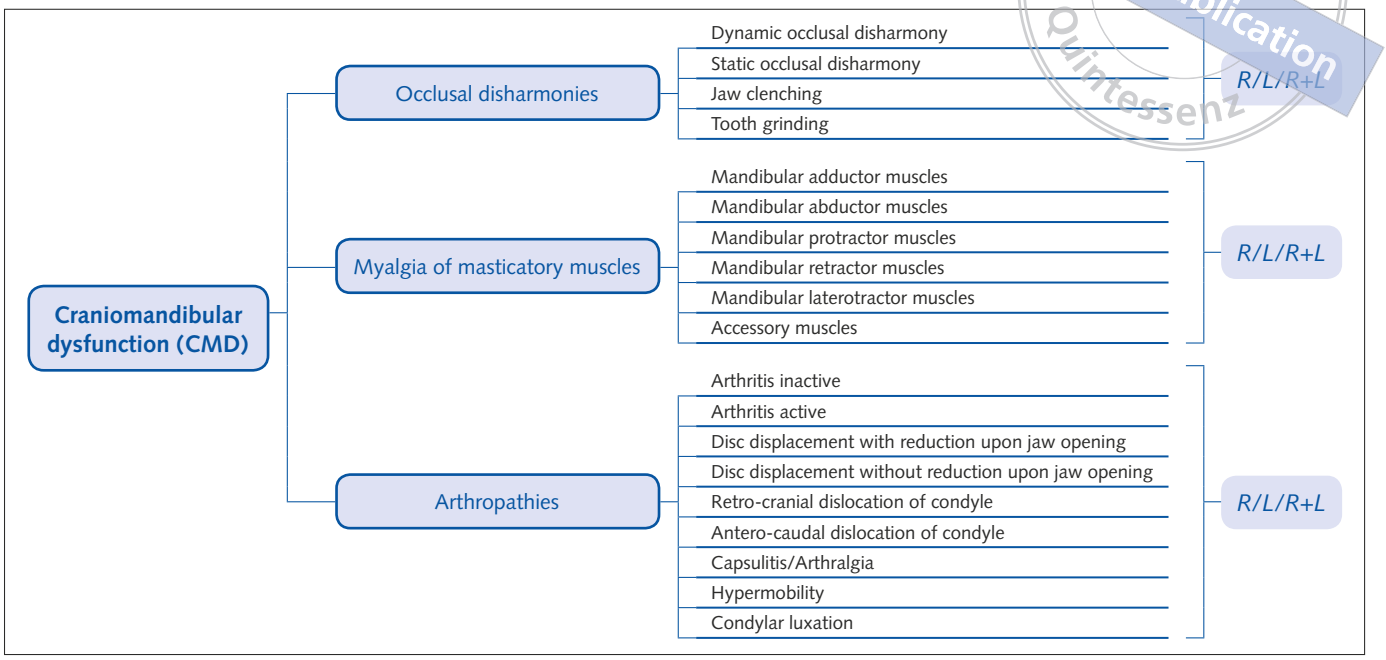


Fig 1c Struture of possible dental CMD diagnoses based on the diagnostic classification of the German Society of Craniomandibular Function and Disorders (DGFDT).[29]

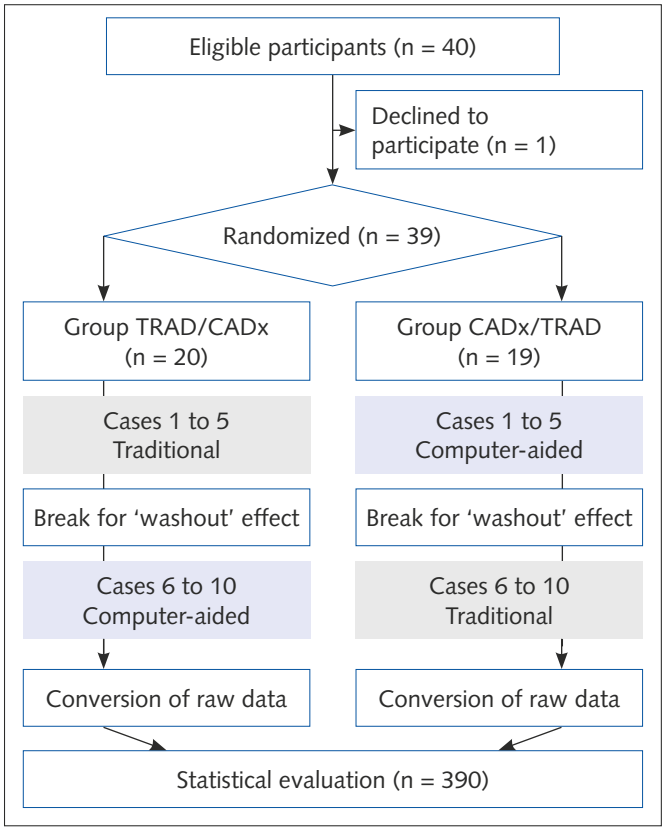


Fig 2 Flowchart illustrating the AB/BA crossover design with 39 participants in two groups.

Data collection

During the data collection phase, all examiners diagnosed the same 10 patient cases, starting from case 1 to case 10 in ascending order. Each examiner had a time limit of 15 min to diagnose each case. None of the examiners required the full 15 min for the assessment.

- Cases 1 to 5 were evaluated in the first phase of the study by the TRAD/CADx group based on the traditional diagnosis assessments, and the CADx/TRAD group by means of computer-aided diagnostic assessments.
- After completing the first phase, the examiners took a 20-min 'washout' break to reduce the odds of carrying over experience from Phase 1 to Phase 2.
- In the second assessment phase immediately after the break, the TRAD/CADx group examined the data based on the computer-aided diagnostics, and the CADx/TRAD group performed the traditional diagnostic assessments.

Consolidated Standards of Reporting Trials (CONSORT) requires the clear and transparent reporting of randomized clinical trials.[41] A flow diagram consistent with the CONSORT standards[42] illustrates the design of this study (Fig 2).

A second flow diagram illustrates the flow of analysis for patient case 1 (Fig 3):

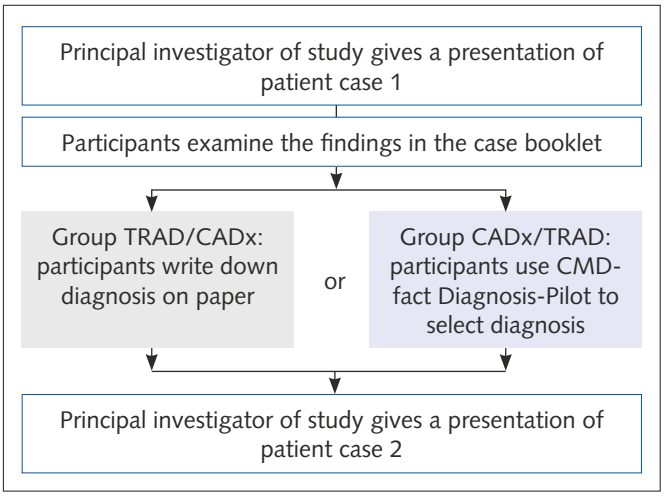
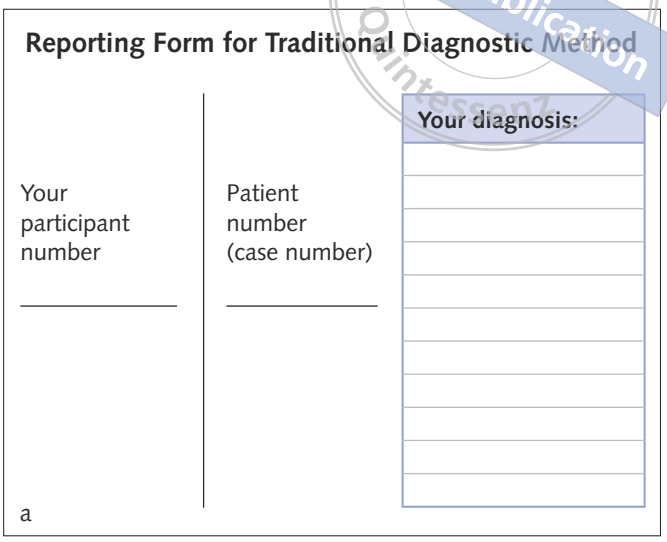


Fig 3 Flowchart illustrating the study process for patient case 1.



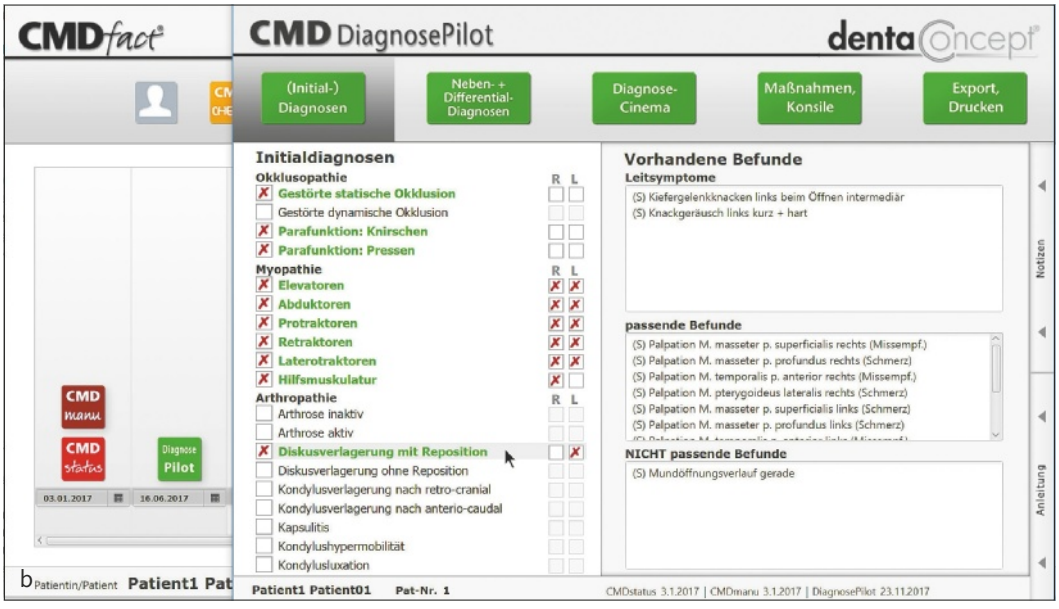
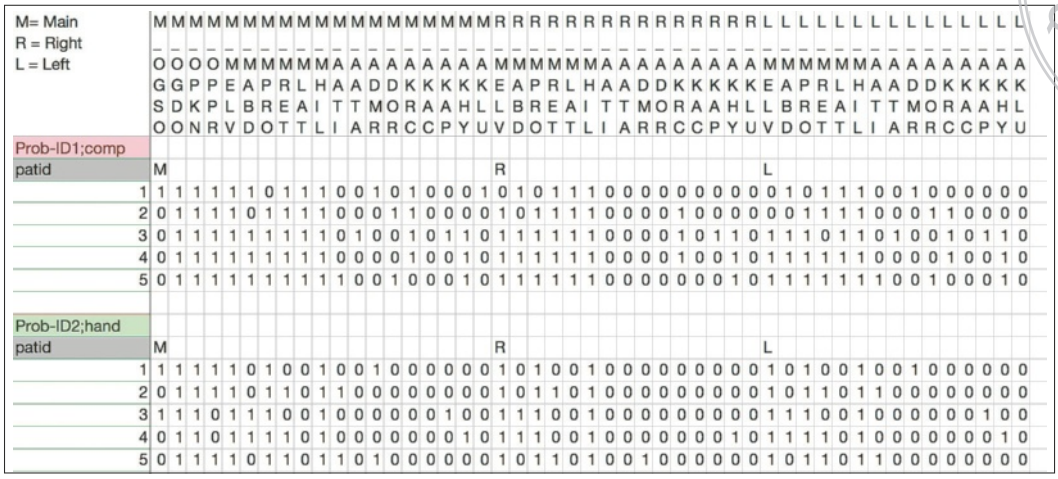


Fig 4a (above right) and 4b (right)
a) Reporting form provided for traditional assessment and documentation of diagnosis.
b) Exemplary view of the CMDfact Diagnosis-Pilot software (German version) used for the computer-aided diagnosis.

- First, one of the principal investigators (the senior author) gave a presentation-based description of findings on the respective case to be assessed, including the patient's history, photographs, and radiographs. Identical information and findings were included in the illustrated patient chart, which was provided to the examiners.
- After the presentation, each examiner independently assessed the findings.
  - For each case, the examiners established the diagnosis and documented the findings depending on their group assignment by either the traditional method (TRAD) or the computer-aided diagnosis (CADx). For the traditional method, they filled out a printed reporting form by hand (Fig 4a), while for the computer-aided method they used a digital reporting form generated by the CMDfact Diagnosis-Pilot software, which they filled out by clicking on the appropriate diagnoses (Fig 4b). The software module's blank reporting form is similar to the printed form used for the traditional method, in which the clinician establishes a diagnosis or diagnoses by mental thought processes based on the available findings, then writes them down by hand on an index card or types the information onto a digital patient chart.



M= Main R = Right L = Left	MMMMMMMMMMMMMMMMMMMMRRRRRRRRRRRRRRRRRRRRLLLLLLLLLLLLLLLLLLLL
	OOOOMMMMMMMAAAAAAAAAAAAAAAAAAAAMMMMMMMMAAAAAAAAAAAAAAAAAAAA
	GGPPEAPRLHAADDKKKKKKEAPRLHAADDKKKKKKEAPRLHAADDKKKKKKEAPRLHAADDKKKKKK
	SDKPLBREAITTMORAAHLLBREAITTMORAAHLLBREAITTMORAAHLLBREAITTMORAAHL
	OONRVDOTTLIARRCCPYUVDOTTLIARRCCPYUVDOTTLIARRCCPYUVDOTTLIARRCCPYU
Prob-ID1;comp	
patid	M R L
1	1 1 1 1 1 1 0 1 1 1 0 0 1 0 1 0 0 0 1 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 0 1 0 0 0 0 0
2	0 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 0 0 0 1 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 0
3	0 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 0 1 1 0 1 1 1 1 1 1 1 0 0 0 0 1 0 1 1 0 1 1 1 0 1 1 0 1 0 0 1 0 1 1 0
4	0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 0 0 1 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 1 0 1 1 1 1 1 1 0 0 0 0 1 0 0 1 0
5	0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 0 0 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 1 0 0 1 0 0 0 1 0
Prob-ID2;hand	
patid	M R L
1	1 1 1 1 1 1 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 1 0 0 0 0 0 0
2	0 1 1 1 1 1 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 1 1 0 0 0 0 0 0 0 0
3	1 1 1 1 0 1 1 1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 0 0 0 0 0 0 1 0 0
4	0 1 1 1 0 1 1 1 1 0 1 0 0 0 0 0 0 0 0 1 0 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 0 1 0 0 0 0 0 0 0 1 0
5	0 1 1 1 1 1 0 1 1 1 0 1 1 1 0 1 0 0 0 0 0 0 0 1 0 1 1 1 0 1 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 1 1 1 0 1 1 0 0 0 0 0 0 0 0

Fig 5 Exemplary view of raw study data in binary code used for the comparison of traditional and computer-aided evaluation.

  - In this study, computer-aided diagnoses were established using the CMDfact 4 software suite and its CMDstatus, CMDmanu, and CMDfact Diagnosis-Pilot modules (dentaConcept Verlag). The patients' findings from the clinical functional analysis and manual structural analysis (´orthopedic tests´) were entered into the appropriate software modules (CMDstatus or CMDmanu, respectively) prior to the study. The CMDfact Diagnosis-Pilot module designates the digitally stored findings to 'matching' diagnoses consistent with the DGFDT diagnostic classification, and also performs grading and weighting. In so doing, it classifies a finding as either a 'principal matching' symptom that is characteristic of the diagnosis, as a 'solely matching' symptom that at least fits the diagnosis, or as a 'possibly conflicting' symptom that is contradictory to the respective diagnosis.
- The possible diagnoses are based on the DGFDT diagnostic classification, under which the global diagnosis of TMD/CMD is divided into three subgroups and 19 initial diagnoses (Fig 1c). Occlusal disharmonies ('occlusopathies') is a subgroup of four diagnoses always associated with bilateral jaw involvement. On the other hand, myalgia of masticatory muscles ('myopathies') and 'arthropathies' are subgroups comprising six and nine diagnoses, respectively; these diagnoses must be differentiated as either 'right,' 'left' or 'bilateral,' because in clinical treatment it is important to know whether joint complaints are associated with the involvement of the right and/or left temporomandibular joint (TMJ).

Reference diagnoses

Before the start of data collection, the principal investigators submitted reference diagnoses for each of the 10 patients enrolled in the study. The reference diagnoses were established using the same patient records that were subsequently available to the examiners. Each reference diagnosis was converted into binary code individually, using the same method by which the data collected from the examiners were converted later for analytical purposes (see 'Analysis'). This enabled a later comparison of agreement between the examiners' diagnoses and the reference diagnoses.

Analysis

First, all examiner diagnoses for each patient case were uniformly compiled in one table in which every possible diagnosis was listed. Code numbers 1 and 0 were used to specify whether the examiner had made the diagnosis in question: '1' = diagnosis established; '0' = diagnosis not confirmed.

In the next step, the diagnostic assessments collected from all 39 examiners for all 10 patients (five by traditional diagnosis and five by computer-aided diagnosis) were converted into uniform, 49-digit strings. The first 19 digits represented the 19 possible diagnoses, and the next two sets of 15 digits each specified the affected side (right or left) in the case of diagnoses requiring lateral differentiation (eg, 0110111101000000010(R)111101000000010(L)110001000000010). Figure 5 shows an excerpt of the data for cases 1 to 5, as assessed by examiners 1 and 2. Overall, the examiners produced a total of 195 traditional diagnoses and 195 computer-aided diagnoses, yielding a total of 390 uniform strings, which were included in the statistical analysis.

Statistical analysis

The objective of the statistical analysis was to test the following parameters:

1. *Agreement of diagnoses by examiners with reference diagnoses:* Cohen's Kappa coefficient[43] was used to measure interexaminer reliability in terms of the agreement of diagnoses by examiners with the reference diagnoses by the experienced clinicians (who, in this case, served as the 'second examiners'). This coefficient of agreement adjusts pure percentage agreement by the expected random agreement between only two possible responses ('Yes' or 'No').[44] We then examined the means of the two groups for statistically significant differences ($P < 0.05$).
2. *Agreement of diagnoses between examiners:* Fleiss' kappa coefficient was used to analyze the consistency within the two groups: traditional diagnosis vs computer-aided diagnosis. The higher the Fleiss' kappa, the more likely the probability of agreement between diagnoses by different examiners within the same group.[45]
3. *Number of false-positive and false-negative diagnoses by test group compared with the reference diagnoses:* In the third part of the statistical analysis, we analyzed the number of examiner diagnoses beyond those specified in the reference diagnoses, and the number of missing diagnoses from those specified in the reference diagnoses in each group. We thereby calculated the number of false-positive and false-negative diagnoses (compared with the reference diagnoses) for each examiner individually, and the percentage of false-positives and false-negatives relative to the total number of 'correct' diagnoses. Box plots of the results were subsequently generated for all four subgroups.

The Shapiro-Wilk test was used for the assessment of normality. The t-test was used to search for significant differences in normally distributed data. The Mann-Whitney U test was used to analyze data that were not normally distributed. Statistical analyses were performed using the SigmaStat 4.0 program package, and graphs were created using SigmaPlot 13 (Systat Software GmbH, Erkrath, Germany).

Null and working hypotheses

This study was based on the following *null hypothesis*:

There is *no difference* ('null' or 'zero difference') between computer-aided diagnoses and traditional diagnoses in terms of:

- Agreement of examiner diagnoses with reference diagnoses.
- Agreement of diagnoses between examiners.
- Number of false-positive and false-negative diagnoses.

The *working hypothesis*, on the other hand, assumes that the use of computer assistance will result in a greater consistency of examiner diagnoses with reference diagnoses and between examiners, but that the ease of clicking on response items in the graphical user interface might result in a higher number of false-positive diagnoses.

Results

1. Agreement of diagnoses by examiners with reference diagnoses: The agreement of diagnoses by the examiners with the reference diagnoses was determined by calculating Cohen's kappa (κ) for each case and examiner. Median kappa values were then determined for each diagnostic method (traditional vs computer-aided diagnosis) across all examiners and cases (n = 10). Mathematically, a kappa value of 0 corresponds to random agreement, values ranging from κ = 0.61 to 0.80 indicate a high probability of agreement, and a value of κ = 1 represents perfect agreement.[46]

Since the Shapiro-Wilk test showed non-normal distribution ($P < 0.05$), the Mann-Whitney U test was used to analyze the data for statistical significance. The median agreement values were κ = 0.553 for traditional diagnosis, and κ = 0.692 for computer-aided diagnosis (Fig 6a). The probability that this difference was due to chance was less than 0.01% ($P < 0.001$).

The individual case results are graphically depicted in box and whisker plots, which show that traditional diagnosis resulted in a much wider scatter range of κ values than computer-aided diagnosis (Fig 6b). This is also consistent with the observed difference in standard deviation for traditional diagnosis (0.20) compared with computer-aided diagnosis (0.16).

2. Agreement of diagnoses between examiners: The degree of agreement of diagnoses between examiners, as determined using Fleiss' kappa (κ), is a measure of consistency of the diagnoses established by the examiners.

In this case, the Shapiro-Wilk test of normality was positive ($P = 0.053$). Application of the two-tailed t-test yielded a Fleiss' kappa value of κ = 0.271 for traditional diagnosis compared with κ = 0.497 for computer-aided diagnosis (Fig 7). The probability that the mean-value difference of κ = 0.266

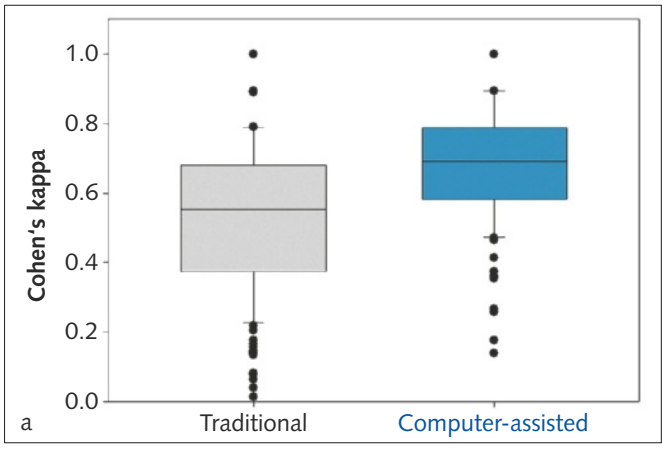
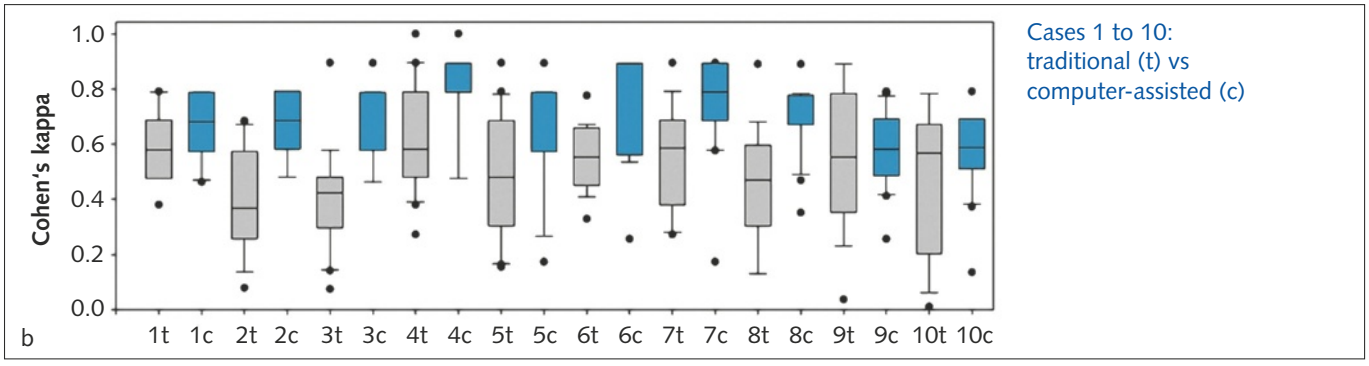


Fig 6a and b a) Diagnostic agreement of probands vs sample solution. Mean values over 10 cases. b) Diagnostic agreement of probands vs sample solution. Values for each of the 10 cases: traditional (t) and computer-aided (c).



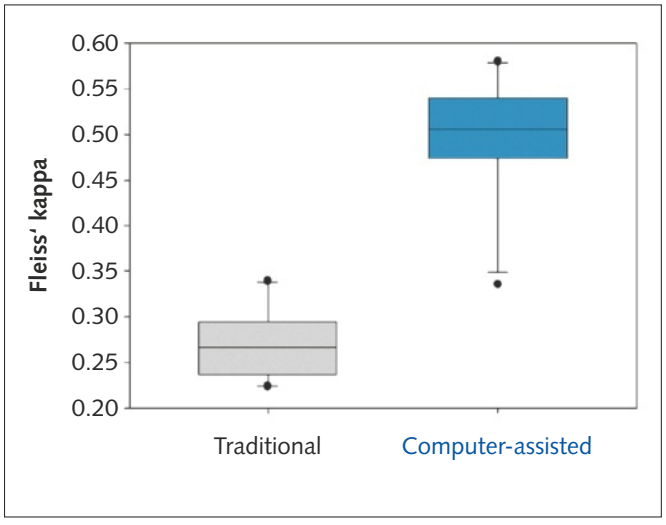


Fig 7 Diagnostic agreement between 39 probands. Mean values over 10 cases.

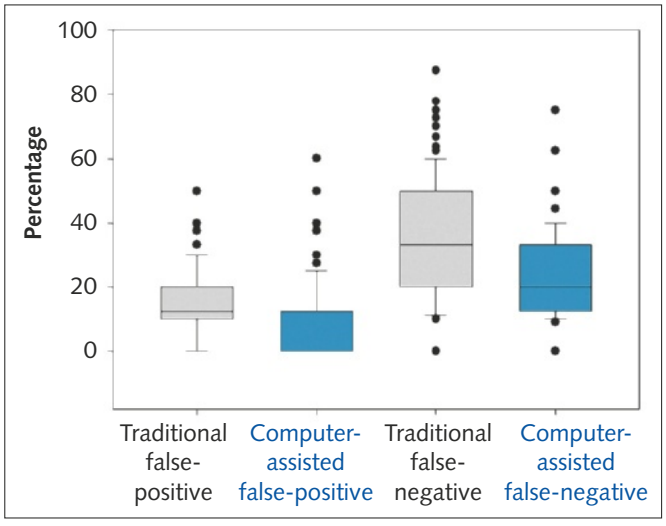


Fig 8 Percentage of false-positive and false-negative diagnoses. Mean values over 10 cases.

was due to a random distribution was less than 0.01% (two-tailed t-test; $P < 0.001$).

As shown in the graph (Fig 7), the sizes of the boxes in the plots for the two groups are roughly equivalent. However, the calculated standard deviation for traditional diagnosis was 0.067, which is about twice as large as that for computer-aided diagnosis.

3. Number of false-positive and false-negative diagnoses: Finally, we analyzed the number of false-positive and false-negative diagnoses for each examiner regarding computer-aided versus traditional diagnosis. This was done by comparing the number of diagnoses specified in the reference diagnoses for each case with the diagnoses established by the examiners in the two different groups. The median percentage of false-positive diagnoses was 20% for computer-aided diagnosis and 33.3% for traditional diagnosis. The rate of false-negative diagnoses was 0% with computer-aided diagnosis compared with 12.5% with traditional diagnosis.

The false-positive and false-negative diagnoses failed the Shapiro-Wilk test for normal distribution ($P < 0.05$). Consequently, agreement between the two distributions was examined using the Mann-Whitney U test, which showed that the differences in median values between the two groups were larger than those for random differences. The differences were statistically significant (false-positive diagnoses: $P < 0.01$; false-negative diagnoses: $P < 0.01$).

As shown in the graph (Fig 8), the *false-positive* diagnoses in both traditional and computer-aided diagnosis fell within a relatively narrow scatter range. On the other hand, the scatter ranges for *false-negative* diagnoses were wider, especially in the case of traditional diagnosis.

Overall, computer-aided diagnosis was associated with a smaller proportion of false-positive and false-negative diagnoses in terms of both median and scatter range.

Discussion

Discussion of study design

In this study design, examiners were selected so as to ensure that they all had the same uniform base of prior knowledge. This measure served to prevent distortion of the results and bias due to knowledge inequalities. As a result of the randomization method, this study design comes close to the RCT principle.

Regarding *blinding*, it was technically impossible to blind the examiners to the method because they had to know whether to establish the diagnosis by the traditional or the computer-aided method when performing the assessments. However, computer analysis of the data was performed in order to keep bias from entering into the analysis of the results of the diagnostic assessments.

The study population reflects clinical reality because it consists of consecutively recruited *real patients*.

The *diagnostic classification* used in this study was chosen because it is suggested by the DGFDT, one of the world's largest specialist dental associations in this field. This DGFDT system was also suitable for our research question because it meets the basic requirements for such applications: It is a published, replicable, and clear system for linking findings to diagnoses. Thus, the diagnostic classification was also appropriate for the research question in this study – apart from being suggested for the task by the applicable scientific association anyway. In principle, the choice of the diagnostic classification is of no significance to the subject investigated in this study, as long as a system to allocate findings to diagnoses or sub-diagnoses is published and thus provides appropriate instructions for establishing reproducible diagnoses.

The *low level of experience of the examiners* in the diagnosis and treatment of TMD or CMD, respectively, is a potential limitation of the study design. However, this prerequisite was the only conceivable way to ensure that the examiners had a uniform knowledge base without bias attributable to one-sided experience.

The *AB/BA crossover design* was used in this study to prevent distortion of the results due to learning effects. This ensured that the learning effects in one group (eg, the traditional diagnosis group) were compensated for as far as possible by the simultaneous learning success of the other group using the other diagnostic method.

The *number of patients* (n = 10) was relatively small at first view. This was not restricted by the number of patients available. Instead, the amount of time required to train the examiners and to review the extensive findings with them, as well as the time required for the examiners to perform the diagnostic assessments, did not allow for a larger number of patients. However, even with this small number of patients, the difference between the test groups was so distinct that the number of patients was sufficient to answer the research question in the end.

Discussion of the results

Regarding the findings, it is striking that significant differences between the two diagnostic methods were found in *all* of the analyses performed. Specifically, computer-aided diagnosis resulted in clearly better results in terms of agreement of the diagnoses of the individual examiners with the reference diagnoses (gold standard) and with the diagnoses of other examiners. Hence, the findings demonstrate the superior quality of computer-aided diagnosis as well as provide evi-

dence of the consistency of diagnoses established using the chosen diagnostic classification.

Future trials could use the same study design to check whether this also applies to the same extent to other diagnostic classifications that are commonly used in research, eg, the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD),[26] and its successor, the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD).[27]

Surprisingly, the results of the study did not confirm the authors' assumption that computer-aided diagnosis might lead to more false-positive diagnoses; to the contrary, the number of false-positive diagnoses was even smaller in the computer-aided diagnosis group. Therefore, the option to simply click on a diagnosis *does not lead to more false-positive diagnoses.*

Regarding the overall *agreement with the reference diagnoses*, a considerable difference also remained in the case of the superior group of diagnoses established using computer assistance. This might be due to the lack of clinical experience of the examiners. The authors believe that closer agreement with the reference diagnoses may be achieved by repeating the study with more experienced examiners. Furthermore, the study design required the examiners to make a diagnosis for each side separately. This is very demanding as, especially in myopathy, different levels of symptom intensity (discomfort or real pain) lead to the respective diagnosis. Also, results from two different kinds of clinical investigations together served as the basis for establishing the diagnosis, which can be confusing. Nonetheless, the results demonstrate that even with this complication, the inexperienced examiners made far more accurate diagnoses when using computer assistance.

Clinical perspective

In clinical practice, however, dentists face the challenge of processing a much larger set of data, as findings from later diagnostic studies such as instrumental functional analysis (eg, condylar position analysis and instrumental jaw motion analysis) and perhaps also magnetic resonance imaging (MRI) are added to the data pool as needed over time. A compatible concept for the standardized assessment of instrumental analyses of jaw movements[47] is already available and recognized.[48] Therefore, a computer-aided system that integrates these results can be expected to have even greater diagnostic benefit.

Conclusions

The system for computer-aided diagnosis in dental functional analysis investigated in the present study facilitates the establishment of correct diagnoses, even by dentists who are not specialized in this field.

Since the established diagnoses are the basis for deciding which type of functional therapy to perform, it can be expected that better and more consistent diagnoses will lead to better treatment outcomes.

The present findings suggest that the quality of diagnoses established with computer assistance is far superior. In view of these findings, it would be desirable to extend the existing system to include the findings of instrumental functional diagnostics, such as instrumental motion analysis of the mandible, condylar position analysis, and diagnostic imaging of the TMJs.

Disclosure statement

The first author of the study was involved in checking preliminary versions of the software used in the study previously, when he was a student assistant. The second and third authors are the developers who designed the software used in this study and have interests in development via author contracts.

Acknowledgments

The authors would like to thank the 40 participating students from the dental school as well as the following dentists who participated in the preliminary trial performed to confirm the suitability of the study design: Ms. Linda Bodart, Dr. Anette Exner-Holzheidt, Dr. Rolf Flier, Nikolas Ganß, Dr. Ute Gresch, Ms. Mona Herzberg, Dr. Manfred Kaufmann, Dr. Peter Krauß, Dr. Jörg Länge, Mr. Lutz Neumann, Dr. Norbert Odenthal, Ms. Wencke von Oheimb, Dr. Sebastian Quaas, Dr. Katharina Reiter, Dr. Silke Rinder, Ms. Wiebke Scheibel, Dr. Wolfgang Schlerf, Dr. Thomas Schlösser, Dr. Eicke Schönmuth, Dr. Ralph Sucker, Ms. Claudia Volz, and Dr. Karin Wedekind.

References

1. Castaneda C, Nalley K, Mannion C, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma 2015;5:4.
2. Berner ES (ed). Clinical Decision Support Systems: Theory and Practice, ed 3. Springer, 2016.
3. Hauschild A, Chen SC, Weichenthal M, et al. To excise or not: impact of MelaFind on German dermatologists' decisions to biopsy atypical lesions. Journal of the German Society of Dermatology 2014;12:606–616.

4. Wollina U, Burroni M, Torricelli R, et al. Digital dermoscopy in clinical practise: a three-centre analysis. Skin Res Technol 2007;13:133–142.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–118.
6. Goldenberg R, Eilot D, Begelman G, Walach E, Ben-Ishai E, Peled N. Computer-aided simple triage (CAST) for coronary CT angiography (CCTA). Int J Comput Assist Radiol Surg 2012;7:819–827.
7. Henderson EJ, Rubin GP. The utility of an online diagnostic decision support system (Isabel) in general practice: a process evaluation. JRSM Short Rep 2013;4:31.
8. Zimmermann M, Mehl A, Mörmann WH, Reich S. Intraoral scanning systems – a current overview. Int J Comput Dent 2015;18:101–129.
9. Zaruba M, Mehl A. Chairside systems: a current review. Int J Comput Dent 2017;20:123–149.
10. Grösser J, Sachs C, Stadelmann M, Schweiger J, Güthe JF, Beuer F. Computer-aided fabrication of a zirconia 14-unit removable dental prosthesis: a technical report. Int J Comput Dent 2014;17:307–316.
11. Kalantari MH, Ghoraishian SA, Mohaghegh M. Evaluation of accuracy of shade selection using two spectrophotometer systems: Vita Easyshade and Degudent Shadepilot. Euro J Dent 2017;11:196–200.
12. Bansal GJ. Digital radiography. A comparison with modern conventional imaging. Postgrad Med J 2006;82:425–428.
13. Molander B, Gröndahl HG, Ekestubbe A. Quality of film-based and digital panoramic radiography. Dentomaxillofac Radiol 2004;33:32–36.
14. (ARö) AfR. s2k-Leitlinie Dentale digitale Volumentomographie: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, 2013 [updated 5.9.2013]. Available from: http://www.dgzmk.de/uploads/tx_szdgzmkdocuments/083-005l_S2k_Dentale_Volumentomographie_2013-10.pdf. Accessed 28 December 2017.
15. Herklotz I, Beuer F, Kunz A, Hildebrand D, Happe A. Navigation in implantology. Int J Comput Dent 2017;20:9–19.
16. Gibbs CH, Hirschfeld JW, Lee JG, et al. Description and clinical evaluation of a new computerized periodontal probe – the Florida probe. J Clin Periodontol 1988;15:137–144.
17. Connert T, Judenhofer MS, Hülber JM, et al. Evaluation of the accuracy of nine electronic apex locators by using Micro-CT. Int Endod J 2018;51:223–232.
18. Abdelaziz M, Krejci I. DIAGNOcam – a Near Infrared Digital Imaging Transillumination (NIDIT) technology. Int J Esthet Dent 2015;10:158–165.
19. Schierz O, Klinger N, Schön G, Reissmann DR. The reliability of computerized condylar path angle assessment. Int J Comput Dent 2014;17:35–51.
20. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. S2k-Leitlinie (Langversion): Instrumentelle zahnärztliche Funktionsanalyse, 2015. [cited 2017]. Available from: http://www.awmf.org/uploads/tx_szleitlinien/083-017l_S2k_Funktionsanalyse_2015-12.pdf. Accessed 28 December 2017.
21. Ahlers MO, Jakstat HA. CMDcheck – CMD-Kurzbefund nach Ahlers/Jakstat, ed 4. Hamburg: dentaConcept, 2015.
22. Ahlers MO, Jakstat HA. Klinische Funktionsanalyse. In: Ahlers MO, Jakstat HA (eds). Klinische Funktionsanalyse • Manuelle Strukturanalyse • Interdisziplinäre Diagnostik. dentaConcept Arbeitsbücher, ed 4. Hamburg: dentaConcept; 2011:171–284.
23. Ahlers MO, Jakstat HA. Manuelle Strukturanalyse. In: Ahlers MO, Jakstat HA (eds). Klinische Funktionsanalyse • Manuelle Strukturanalyse • Interdisziplinäre Diagnostik. dentaConcept Arbeitsbücher, ed 4. Hamburg: dentaConcept; 2011:285–316.
24. Deutsche Gesellschaft für Zahn- Mund- und Kieferheilkunde. Wissenschaftliche Stellungnahme: Klinische Funktionanalyse 2009. Available from: https://www.dgfdt.de/documents/266840/266917/05_Stellungnahme_Klinische_Funktionsanalyse/082c3e7d-de0b-4310-90ab-098ec0ea71c3. Accessed 28 December 2017.
25. DIMDI. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10-GM) Köln: Deutsches Institut für Medizinische Dokumentation und Information; 2017 [10. Revision Version 2017]. Available from: http://www.dimdi.de/static/de/klassi/icd-10-gm/kodesuche/onlinefassungen/htmlgm2017/. Accessed 28 December 2017.
26. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. J Craniomandib Disord1992;6:301–355.
27. 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.
28. Okeson J. Management of temporomandibular disorders and occlusion, ed 7. St. Louis, Missouri: Mosby, 2013.
29. Jakstat HA, Ahlers MO. Initaldiagnosen bei CMD. In: Ahlers MO, Jakstat HA (eds). Klinische Funktionsanalyse • Manuelle Strukturanalyse • Interdisziplinäre Diagnostik. dentaConcept Arbeitsbücher, ed 4. Hamburg: dentaConcept, 2011:329–364.
30. Ahlers MO, Jakstat HA, Freesmeyer WB, Simonis A, Hugger A, Meyer G. Vorschlag eines modernen Diagnoseschemas zur therapiespezifischen Erfassung von Anamnesen und Befunden bei CMD. 34 Jahrestagung der Deutsche Gesellschaft für Funktionsdiagnostik und -therapie (DGFDT) in der Deutschen Gesellschaft für Zahn-, Mund- und Kieferheilkunde (DGZMK). Bad Homburg. Düsseldorf: Deutsche Gesellschaft für Funktionsdiagnostik und -therapie (DGFDT) in der Deutschen Gesellschaft für Zahn-, Mund- und Kieferheilkunde (DGZMK), 2001.
31. Ahlers MO, Jakstat HA. Computer assistance in clinical functional analysis. Int J Comput Dent 2002;5:271–284.
32. Ahlers MO, Jakstat HA. CMDfact – Klinische Funktionsanalyse für Windows und Mac 4.0. Hamburg: dentaConcept, 2017.
33. Köneke C, Landeweer G. Das ist CMDexcellence, 2015. Available from: https://www.cmdexcellence.de. Accessed 20 December 2017.
34. Bumann A, Frank E. MSA Software. Besigheim: pst – praxis systemtechnik GbR, 2009.
35. Ahlers MO, Jakstat HA. Computer-aided examination reports in clinical functional analysis of TMD. Int J Comput Dent 2010;13:121–139.
36. Türp JC, Nilges P. Muskuloskeletale Gesichtsschmerzen. In: Kröner-Herwig B, Frettlöh J, Klinger R, Nilges P (eds). Schmerzpsychotherapie: Grundlagen – Diagnostik – Krankheitsbilder – Behandlung. Berlin, Heidelberg: Springer, 2011:419–430.
37. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133–149.
38. Türp JC, Nilges P, Jakstat HA, Ahlers MO. Fragebogen Chronische Schmerzen (1C), ed 1. Hamburg: dentaConcept, 2012.

39. Ahlers MO, Jakstat HA. Befundbogen Klinische Funktionsanalyse (4), ed 4. Hamburg: dentaConcept, 2012.
40. Ahlers MO, Jakstat HA, Wiesner J, Freesmeyer WB. Manuelle Strukturanalyse/Isometrie, ed 2. Hamburg: dentaConcept, 2010.
41. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152:726–732.
42. Egger M, Jüni P, Bartlett C, CONSORT Group. Value of flow diagrams in reports of randomized controlled trials. JAMA 2001;285:1996–1999.
43. Gwet KL. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. A handbook for researchers, practitioners, teachers & students, ed 2. Gaithersburg, Md.: Advances Analytics, 2010.
44. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960;20:37–46.
45. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971;76:378–382.
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
47. Ahlers MO, Bernhardt O, Jakstat HA, et al. Motion analysis of the mandible: guidelines for standardized analysis of computer-assisted recording of condylar movements. Int J Comput Dent 2015;18:201–223.
48. Jokstad A. Computer-assisted technologies used in oral rehabilitation and the clinical documentation of alleged advantages – a systematic review. J Oral Rehabil 2017;44:261–290.

Address

Kai Becker
Doctoral Candidate
Department of Prosthetic Dentistry
School of Dental Medicine
University Medical Centre Hamburg-Eppendorf
Martinistrasse 52
20251 Hamburg, Germany
Tel. +49 40 7410 5 3267
E-Mail: kaibeckerhh@me.com

Qualitätsverbesserung zahnärztlich-funktionsanalytischer Diagnostik durch computerassistierte Diagnosestellung: Eine randomisierte kontrollierte Studie

Schlüsselwörter: *computerassistierte Diagnostik, kraniomandibuläre Dysfunktion (CMD), klinische Funktionsanalyse, manuelle Strukturanalyse, randomisierte kontrollierte Studie, Diagnoseschema, CMDfact,*

Zusammenfassung

Die funktionsanalytischen Untersuchungen „Klinische Funktionsanalyse“ und „Manuelle Strukturanalyse“ münden in der Stellung qualifizierter Diagnosen. Bislang erfolgt dies durch gedankliche Zuordnung von Befunden zu Diagnosen. In dieser randomisierten kontrollierten experimentellen Diagnostikstudie wurde geprüft, ob die computer-assistierte Diagnostik bei kraniomandibulärer Dysfunktion (CMD) Qualitätsvorteile gegenüber der traditionellen Diagnosestellung bringt.

Probanden und Methode: 39 Hamburger Zahnmedizinstudenten im fünften Studienjahr (Probanden) erhielten eine gemeinsame Instruktion in die Diagnosestellung nach klinischer Funktionsanalyse und manueller Strukturanalyse. Studiengrundlage waren die anonymisierten Befunde von zehn aus einem Schwerpunktzentrum konsekutiv rekrutierten Patienten. Die Probanden wurden in zwei Gruppen randomisiert aufgeteilt. Jeder Proband wertete anschließend im AB/BA-Design jeweils fünf Patientenfälle mit traditionellem Vorgehen und fünf Fälle computerassistiert mit der Software CMDfact 4 aus. Die Diagnosen wurden anschließend mit Fall-Muster-Lösungen (Goldstandard) und untereinander verglichen.

Ergebnisse: Bei der Auswertung nach Cohens Kappa waren die Übereinstimmungen mit den Musterlösungen bei computerassistierter Diagnostik (Median 0,692) signifikant höher ($p < 0,001$) als ohne (0,553). Die Auswertung nach Fleiss' Kappa zeigte, dass auch die Konsistenz der Diagnosestellung zwischen den Probanden computerassistiert (Mittelwert 0,497) signifikant ($p < 0,001$) über den Werten für das traditionelle Verfahren lag (0,271). Auch die Anzahl falsch-positiver und falsch-negativer Diagnosen war bei der computerassistierten Diagnostik signifikant geringer ($p < 0,001$).

Schlussfolgerung: Die Studie zeigt, dass Zahnärzte mit weniger Erfahrung und/oder ohne Spezialisierung in zahnärztlicher Funktionsdiagnostik bei computerassistierter Diagnosestellung eine deutlich bessere und konsistentere Diagnosequalität erreichen. Es erscheint daher sinnvoll, Befunde weiterer Untersuchungen (Kondylenpositionsanalyse, Bewegungsanalyse) in die computerassistierte Auswertung einzubeziehen.