Diagnostic Tool Using the Diagnostic Criteria for Temporomandibular Disorders: A Randomized Crossover-Controlled, Double-Blinded, Two-Center Study

Andrew Young, DDS, MSD
Department of Diagnostic Sciences
Arthur A. Dugoni School of Dentistry
University of the Pacific
San Francisco, California, USA

Samantha Gallia, BS
Arthur A. Dugoni School of Dentistry
University of the Pacific
San Francisco, California, USA

John F. Ryan, BS
University of California
San Diego School of Medicine
La Jolla, California, USA

Atsushi Kamimoto, DDS, PhD
Department of Comprehensive Dentistry and Clinical Education, Division of Dental Education
Dental Research Center, Nihon University School of Dentistry
Tokyo, Japan

Olga A. Korczeniewska, PhD
Center for Orofacial Pain and Temporomandibular Disorders
Department of Diagnostic Sciences, Rutgers School of Dental Medicine
Rutgers, The State University of New Jersey
Newark, New Jersey, USA

Mythili Kalladka, BDS, MSD
Junad Khan, BDS, MSD, MPH, PhD
Orofacial Pain and TMJD
Eastman Institute for Oral Health
University of Rochester Medical Center, School of Medicine and Dentistry
Rochester, New York, USA

Noboru Noma, DDS, PhD
Department of Oral Diagnostic Sciences
Nihon University School of Dentistry and Clinical Research Division, Dental Research Institute
Nihon University, Tokyo, Japan

Correspondence to:
Dr Andrew Young
Department of Diagnostic Sciences
Arthur A. Dugoni School of Dentistry
155 Fifth Street
San Francisco, CA 94103, USA
Email: ayoung@pacific.edu
Fax: 415-929-6624

Submitted May 9, 2021; accepted May 25, 2021.

Aims: To assess the speed and accuracy of a checklist user interface for the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD). **Methods:** A diagnostic tool formatted as a checklist was developed and compared to an existing diagnostic tool, the DC/TMD diagnsostic decision trees. Both types of tools use the DC/TMD and were tested by dental students, interns, and residents in the USA and Japan for diagnosis of hypothetical patients. The comparisons were done in a randomized, crossover, controlled, double-blinded trial. **Results:** Overall, subjects using the experimental tool answered 25% more correct diagnoses ($P < .001$) and missed 27% fewer diagnoses ($P < .01$). They were also able to finalize their diagnoses faster than those using the control tool, in 16% less time ($P < .05$). The difference in accuracy was more pronounced in complex cases, while the difference in speed was more pronounced in simple cases. **Conclusion:** This checklist is an alternative user interface for the DC/TMD. *J Oral Facial Pain Headache 2021;35:241–252. doi: 10.11607/ofph.3008*

Keywords: *DC/TMD, diagnosis, Diagnostic Criteria for Temporomandibular Disorders, temporomandibular disorders, TMD*

According to the American Academy of Orofacial Pain, temporomandibular disorders (TMDs) encompass a group of musculoskeletal and neuromuscular conditions that involve the TMJs, the masticatory muscles, and all associated tissues.[1] The National Institute of Craniofacial Research estimates the prevalence of TMD to range from 5% to 12%.[2] The National Health Interview Survey reported that 5% of adults self-reported jaw or face pain within a 3-month period.[3] A systematic review specified the prevalence to be 10% for muscular disorder TMD, 11% for disc displacements, and 3% for temporomandibular joint (TMJ) disorders.[4]

The responsibility for diagnosing and managing TMDs lies primarily with the dentist. However, most providers, including dentists, have received minimal training in TMDs, and those with more intensive training are few. Thus, there is great need for more clinicians to develop the capability to diagnose and manage TMDs.[5]

The Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD)[6] was published in 1992, with the expectation that it would be tested and improved. Numerous validation studies, symposia, workshops, and field tests were conducted, leading to the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) in its current form in 2014.[7] While the authors of the DC/TMD stress that the criteria are still a work in progress, the criteria bring a significant degree of standardization to the diagnosis of TMD for research and clinical purposes.

In conjunction with the DC/TMD, three diagnostic decision trees[8] were made available on the DC/TMD host website as diagnostic tools to aid in the application of the DC/TMD. Once a clinician has collected all of the necessary clinical information from a patient, the decision trees map the clinician to a diagnosis or diagnoses. These trees serve as a user interface for practitioners who are not well versed in the DC/TMD. Their mention in the original DC/TMD publication[7] did not include validation testing. A literature search also did not find testing by other researchers. Therefore, to the present authors' knowledge, the deci-

sion trees have not undergone validation testing. Yet, while they are not a gold standard, they are the only diagnostic tool that uses the DC/TMD.

The diagnostic tool, or user interface, is a crucial link between the clinician and the criteria. In the clinical setting, clinicians who are unfamiliar with the criteria are not likely to use it. If the format of the user interface is similar to other tools with which an individual is adept, comfortable, and amenable, that individual is more likely to use the criteria.[9] A smooth experience will increase the likelihood that the individual will use the criteria again. Conversely, an unfamiliar format or a poor experience will decrease the likelihood of using the DC/TMD again. Yet individuals differ on what formats they are adept and comfortable with and amenable toward. Therefore, more offerings of tool formats will increase the number of clinicians who use the DC/TMD. This would then increase the number of clinicians who manage TMDs and who do so accurately.

The present authors therefore developed another diagnostic tool for the DC/TMD, formatted as a checklist rather than as a decision tree (Fig 1), while still using the same general wording as the DC/TMD. The objective of this study was to test the speed and accuracy of this tool compared to the DC/TMD decision trees.

Materials and Methods

This randomized crossover-controlled double-blinded study was conducted among dental students, interns, and residents in two centers: *(1)* Arthur A. Dugoni School of Dentistry, University of the Pacific (UOP), San Francisco, California, USA; and *(2)* Nihon University (NU) School of Dentistry, Tokyo, Japan. All consent, recruitment, and testing were done between December 2019 and April 2020. This study was approved by the Institution Review Board of the UOP (protocol 19-117) and the Ethical Committee of the NU School of Dentistry (protocol EP17 D001) and adhered to the Helsinki Guidelines.

Control/Standard Diagnostic Tool

This tool[8] was designed as three diagnostic decision trees; one for pain conditions (myalgias, arthralgia, and headache attributed to TMD), one for intra-articular disc disorders (disc displacements), and one for degenerative joint disorder. Subjective details are arranged at the top, objective details in the middle, and diagnoses on the bottom. To use this tool, the clinician first asks the patient questions regarding the details of the symptoms and then performs a physical exam. The clinician then selects the appropriate tree and uses that information to progress through the tree, arriving at the appropriate diagnosis.

Experimental Diagnostic Tool

This tool (Fig 1) was designed with the purpose of easing the diagnostic process when multiple conditions are present in a single patient. It was tested for accuracy and ease of use in hypothetical patients in seminars (3 to 20 individuals) of senior dental students, orthodontic residents, and nonfaculty practicing dentists at UOP from October 2017 to March 2019. Feedback from the roughly 300 participants in these seminars and clinical consultations influenced numerous modifications to the checklist that brought it to its current form. Starting in January 2018, it was also made available to all students and residents in the clinic and has since then been used in live patients during chairside TMD consultations.

This tool was designed as a checklist. The different types of TMD are arranged in rows, with subjective details on the left side, objective details in the middle, and diagnoses on the right side. These details are the diagnostic criteria for each diagnosis. Technically, for a diagnosis to be made, all items/criteria in the condition's row must be checked.

To use this tool, as with the diagnostic decision trees, a clinician first asks the patient questions regarding the details of the symptoms and then performs a physical exam. All positive answers and findings are checked. The clinician can then identify which conditions/rows have met all of the diagnostic criteria. However, there will be times when not all the criteria are checked. For example, for patients presenting with myofascial pain of the lateral pterygoid and medial pterygoid muscles, such muscles cannot be palpated. Thus, the criterion "Familiar pain with palpation of masticatory muscles" would be left unchecked for the condition "Local Myalgia." However, when noting that three of the four criteria for "Local Myalgia" are checked, the clinician may reason that local myalgia is a possible diagnosis.

Subjects

The inclusion criteria were dental students, interns, and residents. Exclusion criteria were previous exposure to either the experimental or control tool. Sample size estimates determined that 100 subjects were needed, based on a half-confidence interval of 0.1 and a specificity and sensitivity of 0.5.

A total of 155 individuals participated in this study (Fig 2). As this study was crossover in design, there were 155 subjects in the control group and 154 subjects in the experimental group, effectively totaling 309 subjects.

At UOP, the 98 dental student subjects were in their second year. At the time of their participation in the study, the students had only had 2 hours of lecture describing the different types of TMDs and how an examination is performed. The 7 orthodontic resident

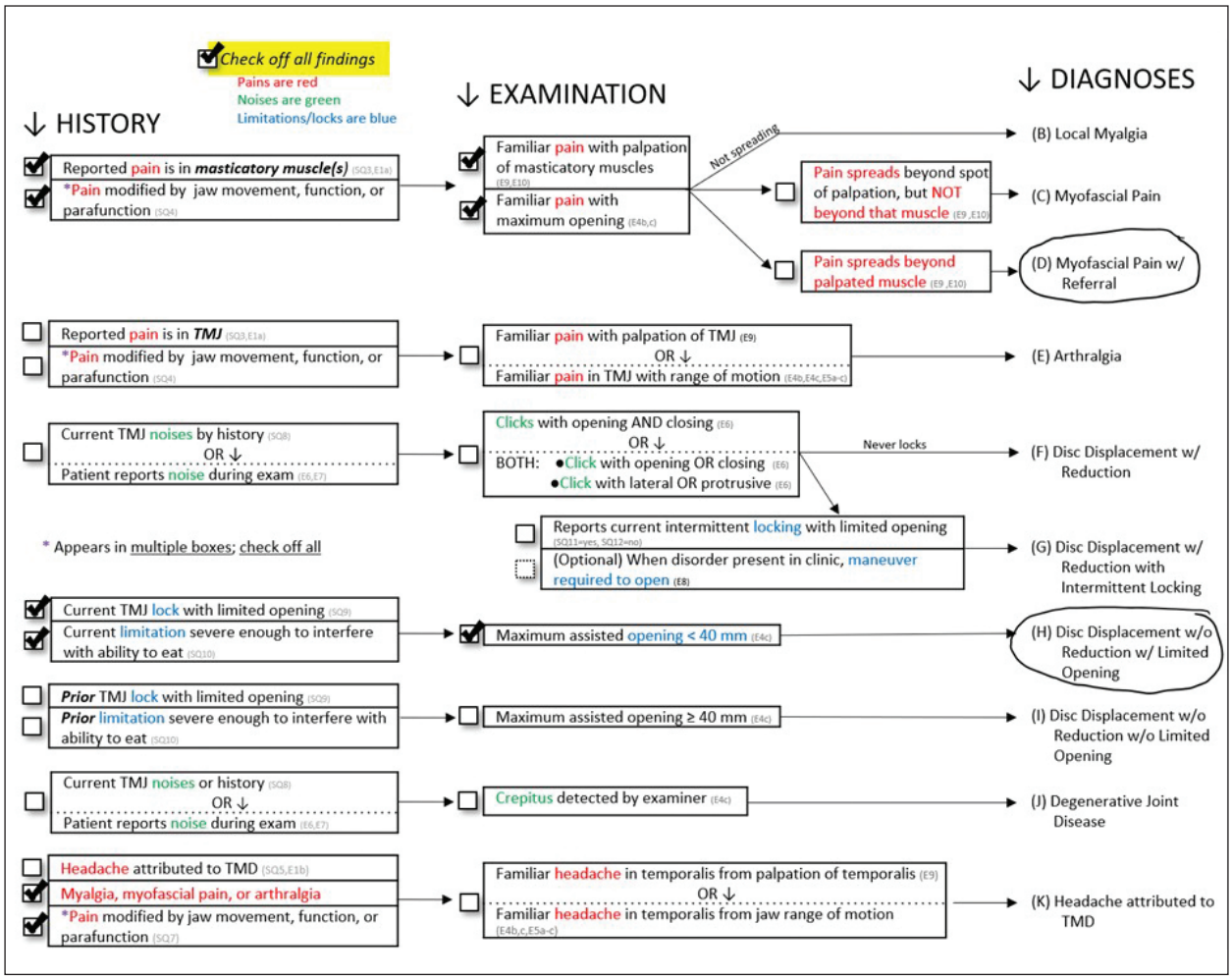

Fig 1 Experimental checklist used in the study.

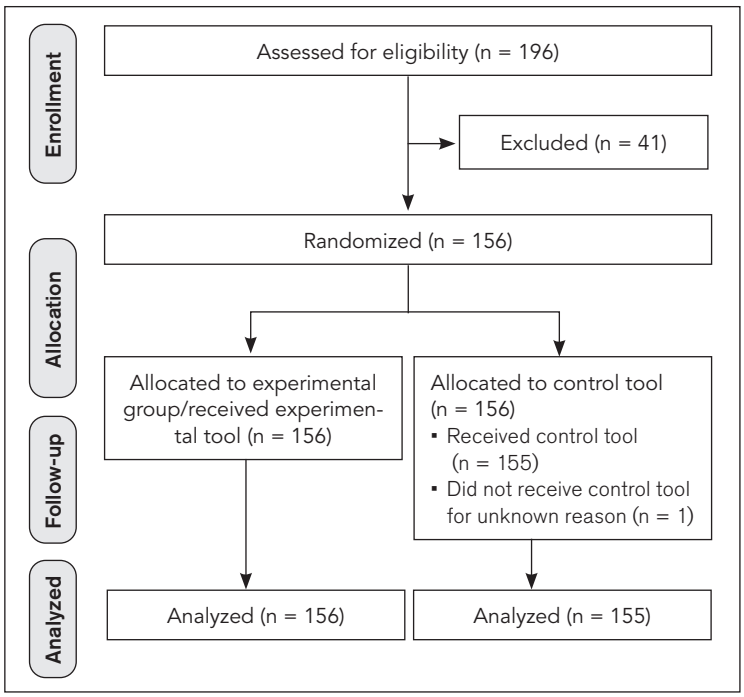

Fig 2 Flowchart of progression of participants through the study.

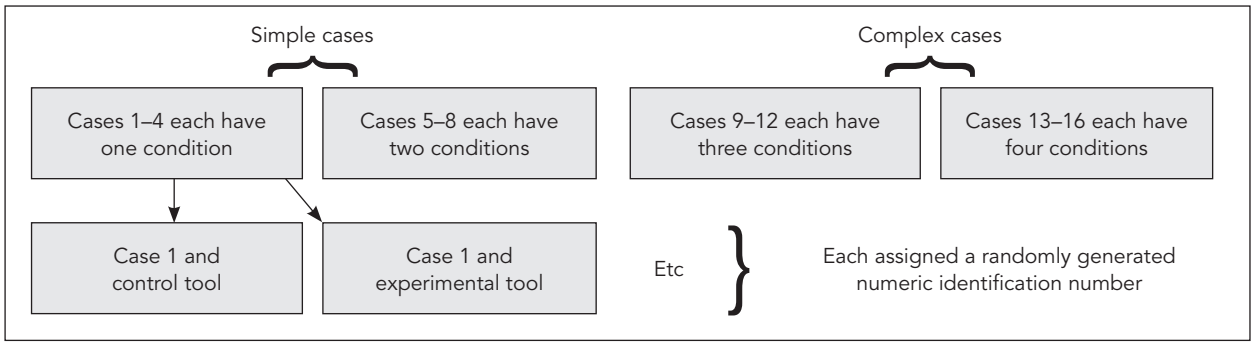

Fig 3 Types of case complexity.

subjects were in their first year at the same school and received the same lecture on TMD, though it is unknown what amount of TMD education each received in their prior dental school education. Neither the students nor the residents had yet been exposed to either tool in lectures, seminars, or clinics.

At NU, the 50 intern and resident subjects were in the Department of Oral Diagnosis. It is unknown what amount of TMD education each received in their prior dental school education.

This study was performed during normal class times. All who presented for class were introduced to the study, invited to participate, and given a consent form. Those who signed the form were thereby enrolled as subjects in the study. One author (A.Y.) handled enrollment at UOP, and another author (A.K.) handled enrollment at NU. All residents and interns who were recruited agreed to participate and completed the whole study. Of the 140 dental students recruited, 98 agreed to participate. One NU participant completed the experimental tool, but not the control tool; the reason for completing only one tool by that student is unknown.

Cases

A total of 16 cases were created by one author (A.Y.; Fig 3). Four cases had a single TMD condition (for example, one case had myofascial pain with referral, another case had disc displacement with intermittent locking, etc). Four cases had two simultaneous TMD conditions (for example, one case had both local myalgia and arthralgia, another case had both right-sided disc displacement with reduction and left-sided disc displacement without reduction with limited opening, etc). Four cases had three simultaneous TMD conditions, and four cases had four simultaneous TMD conditions. Each TMD condition's description contained all the mandatory signs and symptoms of the DC/TMD for that condition, and no additional signs or symptoms were described. Cases were identified with three-digit numeric codes generated by a Microsoft Excel random number generator.

Protocol

Both UOP and NU followed the following protocol (Fig 4).

Every subject received a first packet that contained a simple case description (one or two simultaneous TMD conditions), either the experimental or control tool to use for diagnosing that case, and an answer sheet. Packets were handed out to seated subjects during a lecture or seminar. Simple randomized allocation was achieved with the use of an Excel random number generator to dictate the order in which cases were stacked and distributed. Concealment of allocation sequence was achieved by labeling each case with only the randomly generated numeric code. One author (A.Y.) generated both the numeric codes and the randomized allocation sequence.

The start time was noted on the answer sheet. Then the subject read the case description, used the given diagnostic tool, marked the diagnosis (or diagnoses) on the answer sheet, and turned in the entire packet. The end time was noted on the answer sheet.

At the end of the first packet, the subject read instructions on which second packet to obtain. The second packet contained a complex case (three or four simultaneous TMD conditions), either the experimental tool or control tool for diagnosing that case, and an answer sheet. Packets were paired so that if the subject received the experimental tool for the first case, they received the control tool for the second case, and vice versa (Fig 5). In this way, every subject served in both the experimental group and the control group.

Subjects had never seen either of the diagnostic tools prior to the study, so while they were not blind to the structure of the diagnostic tool they were using, they were unaware of which was the more established tool (control) and which was the experimental tool. Thus, the subjects were blinded.

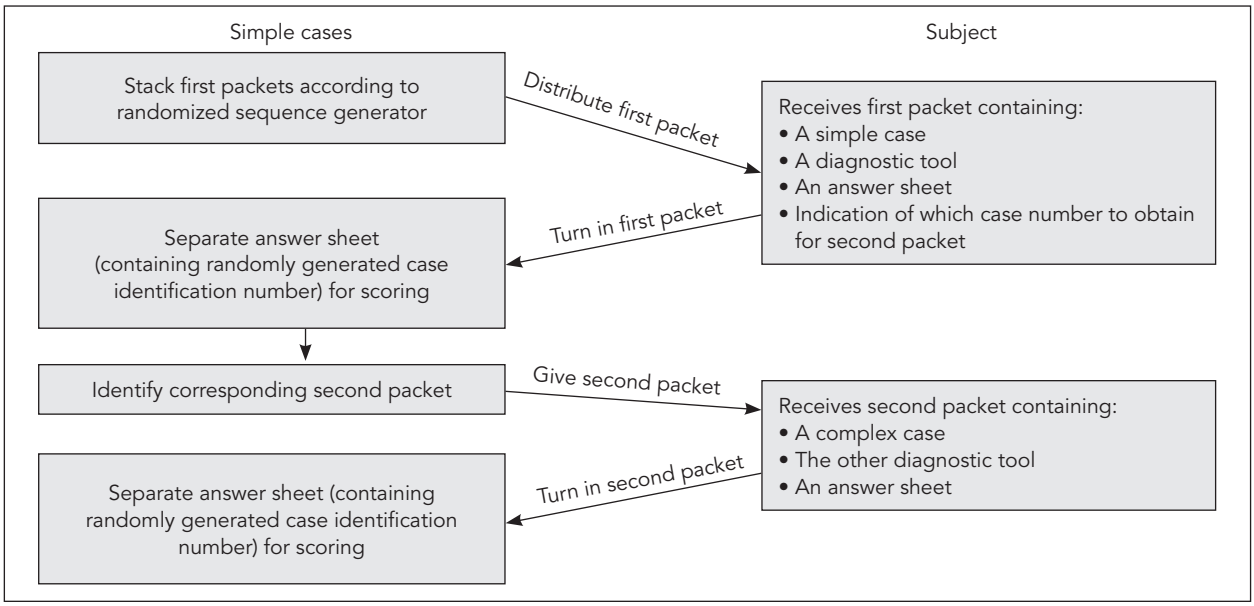

Fig 4 Study protocol for study personnel and participants.

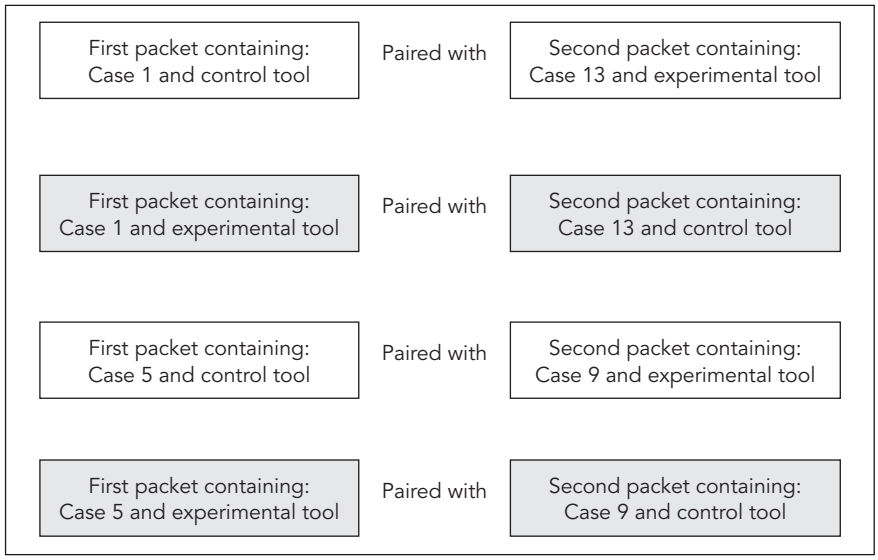

Fig 5 Example contents of participant packets.

For each case, subjects were not told how many simultaneous conditions the case had; they were only told that the number of conditions could range from 1 to 4. To ensure no advantages or disadvantages were imparted by a tool being used for the first case vs the second case, the experimental tool was used for the first case roughly the same number of times that the control tool was used for the first case. The simple randomized allocation sequence determined who received the control tool first and who received the experimental tool first. Also, each individual case was tested on both the control tool and the experimental tool, though by different subjects.

To ensure blinding of the examiners, each case was identified only with a randomly generated identification number. Thus, when the examiners were handling packets and grading answer sheets, they did not know which cases they were handling and which diagnostic tool had been used for that case. Additionally, the answer sheets were separated from the rest of the packet for grading, further ensuring the grading examiner would not know whether that answer sheet was from the experimental or control group.

Scoring

Subject responses for each case were evaluated by the following parameters:

- Time needed to diagnose
- Number of correct diagnoses
- Number of missed diagnoses
- Number of incorrect added diagnoses (for example, if the correct diagnosis was solely myalgia, and a subject diagnosed myalgia and arthralgia, arthralgia would be an added incorrect diagnosis)

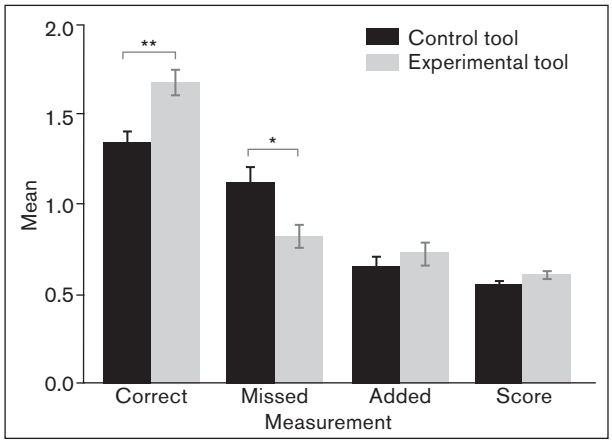

Fig 6 Comparison of the new (experimental) and existing (control) diagnostic tools for the DC/TMD. Participants using the experimental tool identified significantly more correct diagnoses and missed significantly fewer diagnoses compared to the control tool. The mean number of incorrectly added diagnoses and the overall score did not differ between the two tools. *P < .05. **P < .01.

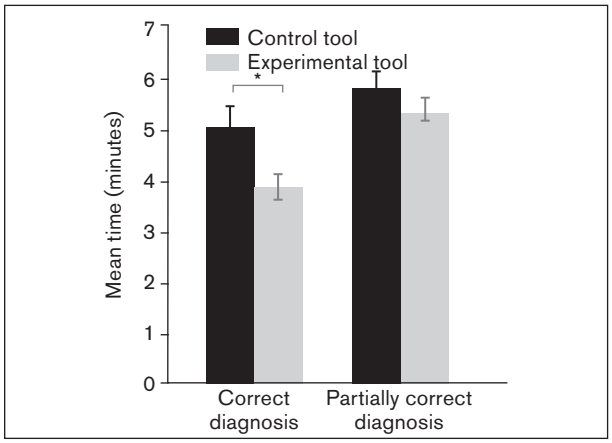

Fig 7 Mean time to identify correct or partially correct diagnoses using experimental and control diagnostic tools for the DC/TMD. Participants using the experimental tool identified correct diagnosis significantly faster compared to those who used the control tool. There was no significant difference in the time it took to identify a partially correct diagnosis (ie, if one of the multiple conditions present was correctly identified) between the two tools. *P < .05.

- Score = (number of correct diagnoses) / [(number of correct diagnoses) + (number of missed diagnoses) + (number of incorrect added diagnoses)]

Statistical Analyses

Data calculations and analyses were performed using Microsoft Excel (2016) and JMP Pro version 15.0.0 (SAS Institute). The distribution of outcome measures (number of correct diagnoses, number of missed diagnoses, number of added diagnoses, score, and time) was tested for normality using the goodness of fit test. All of the outcome measures of interest were significantly skewed. Therefore, nonparametric Wilcoxon rank sum test was used to compare the outcomes between control and experimental tools. The Steel-Dwass all pairs test was used to control for type I error rate in multiple comparisons. Contingency analysis of individual conditions by tool type was performed using Fisher exact test (small sample sizes) or Pearson chi-square test (large samples). Two-tailed significance level alpha was set at .05 for all analyses. Results are presented as mean and standard error of the mean (SEM) unless otherwise specified.

Results

A total of 155 participants were recruited from two locations, UOP (n = 105) and NU (n = 50). UOP participants included second-year dental students (n = 98) and orthodontic residents (n = 7). NU participants included interns (n = 28), residents (n = 12), and unknown (n = 10). NU subjects who were identified as unknown did not specify whether they were interns or residents. All of the participants, except for one from NU who completed only the control tool but not the experimental tool, completed two parts of the study (ie, used both tools, control and experimental, to diagnose assigned cases).

Overall Comparison of Diagnostic Tools

Participants identified significantly more correct diagnoses when using the experimental tool (1.7 ± 0.93) compared to the control tool (1.3 ± 0.81, P = .002). Additionally, participants missed significantly fewer diagnoses when using the experimental tool (0.8 ± 0.87) compared to the control tool (1.1 ± 1.1, P = .024). There were no significant differences in the number of incorrectly added diagnoses or overall score between the two tools (Fig 6). Participants identified a correct diagnosis significantly faster when using the experimental tool (3.9 ± 2.2 minutes) compared to the control tool (5.1 ± 3.7 minutes, P = .016; Fig 7). The two tools did not significantly differ in the time it took to identify a partially correct diagnosis.

Performance of the Two Diagnostic Tools with Simple and Complex Cases

Simple cases had either one TMD condition or two simultaneous TMD conditions. For such cases, subjects using both the control and experimental diagnostic tools identified correct diagnoses (control tool: 1.1 ± 0.6; experimental tool: 1.1 ± 0.5, P = .68) and missed diagnoses (control tool: 0.4 ± 0.6; experimental tool: 0.4 ± 0.7, P = .54) at similar rates (Fig 8). Participants who used the experimental tool correctly diagnosed simple cases significantly faster

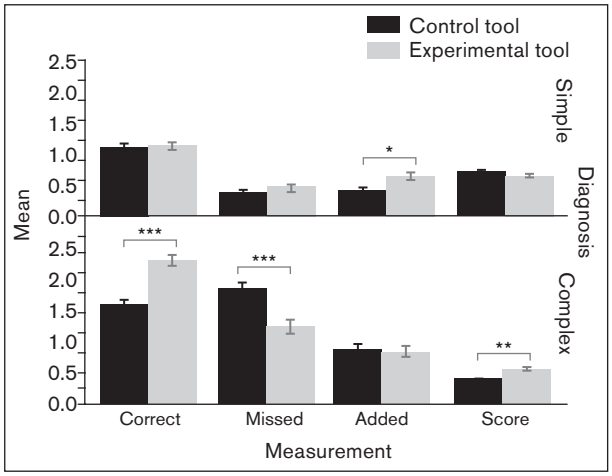

Fig 8 Comparison of the experimental and control diagnostic tools for the DC/TMD by case complexity. The experimental and the control tools performed comparably for simple cases (1 or 2 TMD conditions present) in terms of the number of correctly identified diagnoses, number of missed diagnoses, and overall score. Participants using the experimental tool added significantly more incorrect diagnoses compared to those who used the control tool. For complex cases, the experimental tool performed significantly better for all of the parameters tested (number of correctly identified diagnoses, number of missed diagnoses, and overall score) except for the number of incorrectly added diagnoses, where the two tools performed comparably. *P < .05. **P < .001. ***P < .0001.

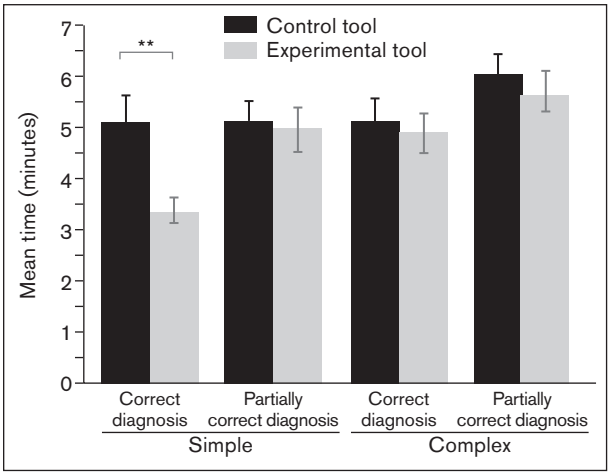

Fig 9 Mean time to identify a correct or partially correct diagnosis using the experimental and control diagnostic tools for DC/TMD by case complexity. Participants using the experimental tool correctly identified simple diagnoses significantly faster compared to those using the control tool. The time to identify partially correct simple or complex diagnoses did not differ between the two tools. **P < .001.

(3.28 ± 2.0) compared to those who used the control tool (5.05 ± 4.3, P < .001; Fig 9). Partially correct diagnosis of simple cases took a comparable amount of time when using either one of the diagnostic tools.

Complex cases had either three or four simultaneous TMD conditions. For such cases, participants using the experimental tool made significantly more correct diagnoses (2.3 ± 0.9) compared to participants using the control tool (1.6 ± 0.9, P < .001). Additionally, participants who used the experimental tool missed significantly fewer diagnoses (1.2 ± 0.9) compared to participants who used the control tool (1.8 ± 1.0, P < .001). The two tools did not significantly differ in the number of incorrectly added diagnoses. The experimental tool resulted in a significantly higher overall score (0.6 ± 0.3) compared to the control tool (0.4 ± 0.3, P < .001; Fig 8). Complex case diagnoses took a similar amount of time irrespective of the tool used (Fig 9).

Simple Cases

Forty participants used the control tool and 37 participants used the experimental tool to diagnose one-condition cases. For the one-condition cases, there were no significant differences between the two diagnostic tools for any of the parameters tested (Fig 10). Participants who used the experimental tool diagnosed one-condition cases significantly faster (3.24 ± 1.96) compared to those who used the control tool (5.23 ± 4.15, P = .001).

Thirty-five participants used the control tool and 44 participants used the experimental tool to diagnose two-condition cases. For the two-condition cases, the two instruments did not significantly differ in the mean number of correctly identified diagnoses (P = .38) or missed diagnoses (P = .16). The experimental tool resulted in significantly more incorrectly added diagnoses (0.8 ± 0.7) compared to the control tool (0.3 ± 0.5, P = .001). Additionally, with the two-condition cases, the experimental tool had a significantly lower overall score (0.5 ± 0.3) compared to the control tool (0.7 ± 0.3, P = .011; Fig 10). For two-condition cases, the experimental and diagnostic tools did not differ in terms of the time it took to diagnose such cases (P = .20).

The experimental tool resulted in faster diagnoses by 1.31 seconds (or 25% less time, P < .05; Fig 9).

Complex Cases

Forty-four participants used the control tool and 35 participants used the experimental tool to diagnose three-condition cases. For such cases, participants

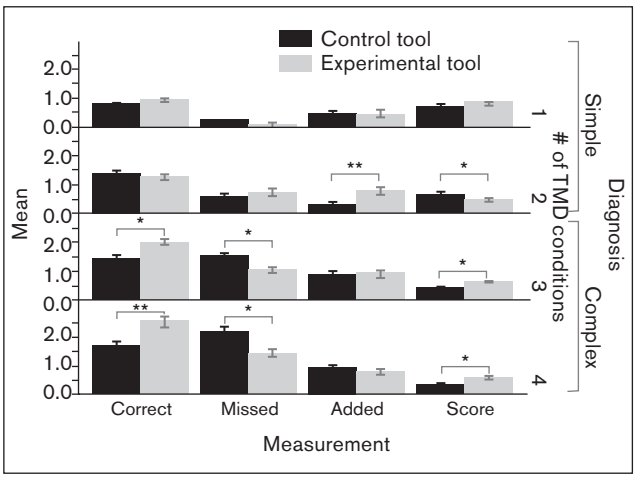

Fig 10 Comparison of the performance of the experimental and control diagnostic tools for the DC/TMD by number of TMD-associated conditions present. For cases with 1 TMD-related condition, the experimental and control diagnostic tools performed comparably across all the measurements tested. For cases with 2 TMD-related conditions, the two tools performed comparably in terms of the average number of correctly identified diagnoses and the number of missed diagnoses. For 2-condition cases, participants using the experimental tool added significantly more incorrect diagnoses compared to those who used the control tool, and the overall score was significantly higher for the control compared to the experimental tool. For complex cases with 3 or 4 TMD-related conditions, the experimental tool performed significantly better compared to the control tool in terms of the number of correctly identified diagnoses, number of missed diagnoses, and overall score. *P < .05. **P < .001.

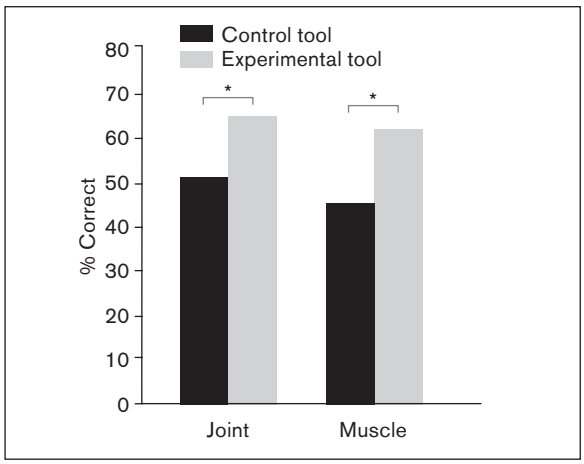

Fig 11 Comparison of correct diagnoses (%) between the control and experimental diagnostic tools for DC/TMD by TMD condition type. The experimental tool performed significantly better for correctly diagnosing both joint- and muscle-related TMD conditions compared to the control tool. *P < .05.

using the experimental tool made significantly more correct diagnoses (2.0 ± 0.6) and missed significantly fewer diagnoses (1.0 ± 0.6) compared to those using the control tool (correct: 1.5 ± 0.9, P = .002; missed: 1.5 ± 0.9, P = .004). The two tools did not significantly differ in the number of incorrectly added diagnoses (P = .82). The experimental tool resulted in a significantly higher overall score (0.6 ± 0.2) compared to the control tool (0.4 ± 0.3, P = .029; Fig 10). For three-condition cases, the experimental and diagnostic tools did not differ in terms of the time it took to diagnose such cases (P = .73).

Thirty-six participants used the control tool and 38 participants used the experimental tool to diagnose four-condition cases. For such cases, participants using the experimental tool made significantly more correct diagnoses (2.6 ± 1.0) and missed significantly fewer diagnoses (1.5 ± 1.0) compared to those using the control tool (correct: 1.7 ± 0.9, P < .001; missed: 2.2 ± 1.0, P = .001. The two tools did not significantly differ in the number of incorrectly added diagnoses (P = .57). The experimental tool resulted in a significantly higher overall score (0.6 ± 0.3) compared to the control tool (0.4 ± 0.3, P = .004; Fig 10). For four-condition cases, the experimental and diagnostic tools did not differ in terms of the time it took to diagnose such cases (P = .53).

Performance by Condition

The control and experimental diagnostic tools were compared for a frequency of correctly identified diagnoses by TMD condition type (joint- and muscle-type conditions). Participants using the experimental tool correctly identified joint- (67.7%) and muscle-type (64.9%) conditions more frequently compared to the control tool (joint: 53.5%, P = .002; muscle: 48%, P = .018; Fig 11). The tools were then compared for how frequently a given condition was correctly diagnosed. Except for local myalgia, the experimental tool correctly diagnosed all of the tested conditions more frequently, and the difference was statistically significant for myofascial pain with referral (experimental tool: 67.7%, control tool: 44.4%, P = .007), disc displacement with reduction with intermittent locking (experimental tool: 79.6%, control tool: 60.4%, P = .048), disc displacement without reduction with limited opening (experimental tool: 76.6%, control tool: 50.84%, P = .004), and disc displacement without reduction without limited opening (experimental tool: 100%, control tool: 60%, P = .029; Fig 12).

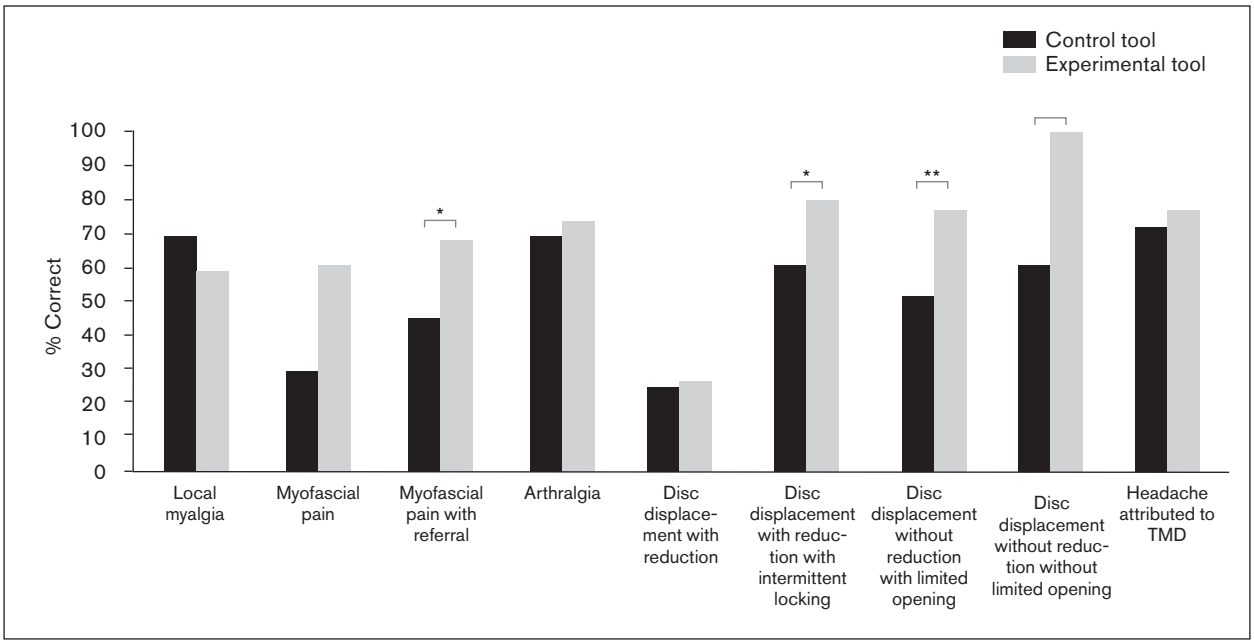

Fig 12 Comparison of the performance of the control and experimental tools by TMD condition. Participants using the experimental diagnostic tool more frequently identified correct diagnoses for most of the TMD-related conditions, except for local myalgia. The differences were statistically significant for myofascial pain with referral, disc displacement with reduction with intermittent locking, disc displacement without reduction with limited opening, and disc displacement without reduction without limited opening. *P < .05. **P < .01.

Consistency Between Sites

There were no significant differences between the two locations (NU and UOP) for either the experimental or the control tool for the outcomes tested (number of correct diagnoses, number of missed diagnoses, number of incorrectly added diagnoses, score; Fig 13a). On average, NU participants took significantly more time to diagnose cases using either the control (NU mean ± SD = 7.68 ± 4.52 minutes, UOP mean ± SD = 4.36 ± 1.87; P < .001) or experimental (NU mean ± SD = 5.76 ± 3.80 minutes, UOP mean ± SD = 4.06 ± 1.81; P = .009) tool compared to UOP participants (Fig 13b). NU participants took significantly less time to diagnose cases using the experimental tool (5.76 ± 3.80 minutes) compared to the control tool (7.68 ± 4.52 minutes, P = .006). There was no significant difference for the time it took to diagnose cases using either the experimental (4.06 ± 1.81) or control (4.36 ± 1.87, P = .35) tool for UOP participants (Fig 13c).

Analysis of Study Participants

There were no significant differences between the two locations for the outcomes tested with either the experimental or the control tool; therefore, two locations were combined in the analysis investigating differences between the study participants. Five "types" of study participants were analyzed, and they included second-year dental students (n = 98), orthodontic residents (n = 7), oral diagnosis interns (n = 28), oral diagnosis residents (n = 12), and oral diagnosis unknown (intern/resident, n = 10). There were no significant differences between study participants for the outcomes tested for either the experimental tool (number of correct diagnoses, P = .93; number of missed diagnoses, P = .71; number of incorrectly added diagnoses, P = 1.00; and score, P = .94) or the control tool (number of correct diagnoses, P = .96; number of missed diagnoses, P = .94; number of incorrectly added diagnoses, P = .44; and score, P = .96).

When using the control tool, second-year dental students took significantly less time to diagnose cases (4.29 ± 1.69) compared to oral diagnosis interns (7.71 ± 4.28, P < .0001), oral diagnosis residents (8.92 ± 6.29, P = .0003), and oral diagnosis unknown (6.10 ± 1.85, P = .005). There were no significant differences with respect to the time it took to diagnose cases between other groups.

When using the experimental tool, second-year dental students took significantly less time to diagnose cases (4.01 ± 1.72) compared to oral diagnosis residents (6.58 ± 3.55, P = .007) and interns (5.81 ± 4.06, P = .031). There were no significant differences with respect to the time it took to diagnose cases between other groups.

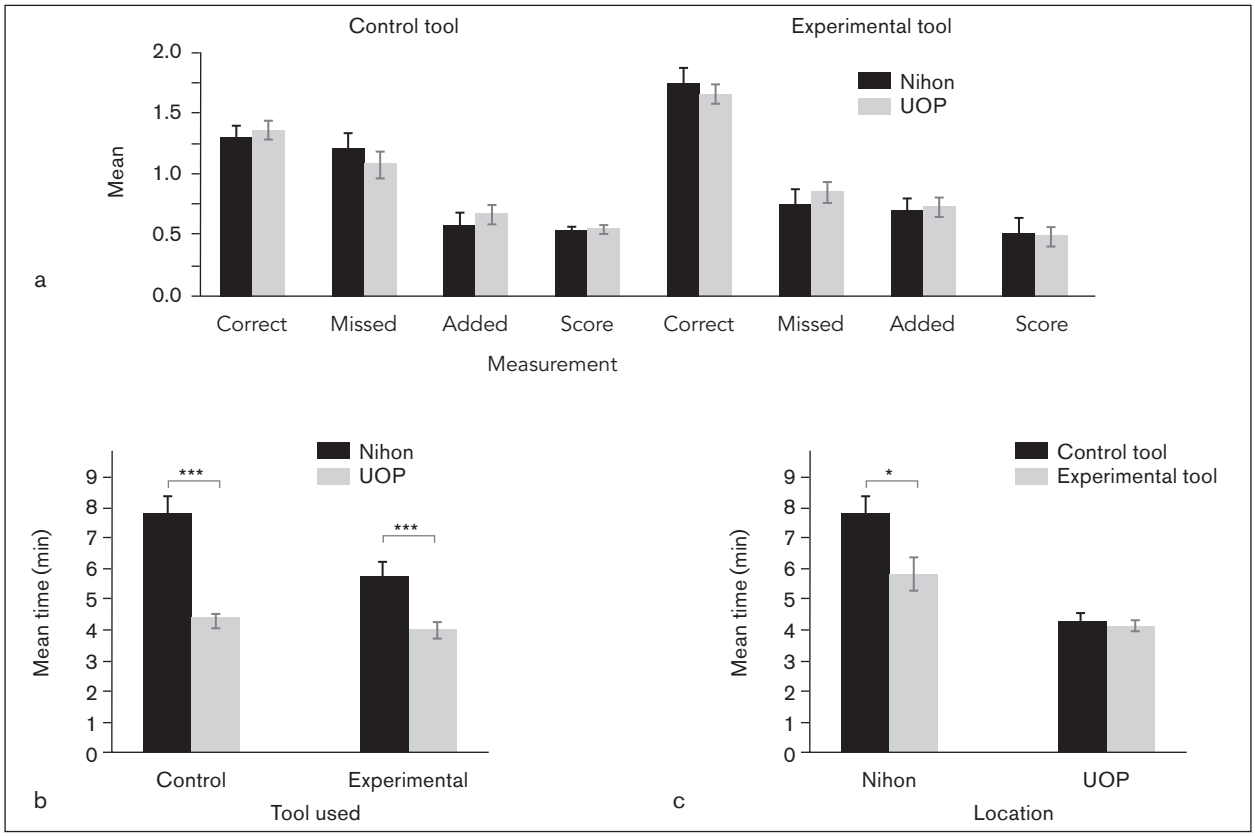

Fig 13 Comparison of performance of the experimental and control diagnostic tools for the DC/TMD between Nihon University (NU) and University of Pacific (UOP) participants. **(a)** Performance of the two tools with respect to the parameters measured (number of correct diagnoses, number of missed diagnoses, number of incorrectly added diagnoses, and score) was consistent between the NU and UOP. **(b)** On average, NU participants took significantly more time to diagnose cases using either the control or experimental tool compared to UOP participants. **(c)** NU participants took significantly less time to diagnose cases using the experimental tool compared to the control tool. UOP participants took a comparable amount of time to diagnose cases using either one of the tools. *$P < .05$. ***$P < .0001$.

Discussion

Overall, the experimental tool resulted in 30% more correct diagnoses and 27% fewer missed diagnoses. The experimental tool also arrived at the correct diagnosis in three-quarters the time that the control tool took. A possible trend was also seen in which the experimental tool performed better in cases that had more simultaneous conditions. When used for cases that had three or four simultaneous conditions, the experimental tool performed better, with 44% more correct diagnoses and 33% fewer missed diagnoses.

While all of these differences were statistically significant, the clinician must decide if these differences are clinically significant. The choice of which tool to use will depend partly on personal preference. Some will prefer one layout over another for reasons pertaining to familiarity, problem-solving style, etc. Introducing an additional user interface for the DC/TMD increases the number of clinicians who will use the DC/TMD. This will in turn increase the number of clinicians managing TMDs and improve the completeness and accuracy of their diagnoses.

The experimental tool was designed to aid in a number of specific situations. First, the nonspecialist, at times, will not initially know whether a patient's TMD is muscular or joint-related in origin. The single-checklist format of the experimental tool does not require the clinician to choose between a muscular and joint TMD tool because it works for both types. Second, for patients with more than one condition, the clinician may not know which signs and symptoms apply to which condition, but such knowledge is necessary for a decision-tree format of the control

tool, which diagnoses one condition at a time. The checklist format of the experimental tool allows all signs and symptoms to be entered at the same time and then diagnoses any number of conditions simultaneously. This may explain why test subjects using the experimental tool arrived at significantly more correct diagnoses than those using the control tool for cases that had three or four simultaneous conditions, but not for the cases that had one or two conditions. Third, patients will often exhibit only some of the signs and symptoms of their condition(s). For example, a myalgia may be in the medial pterygoid muscle and thus not fulfill the myalgia criterion "familiar pain with palpation of masticatory muscles." In the decision-tree format of the control tool, this missing piece of data would result in the conclusion "Investigate other pain diagnoses." In the checklist format of the experimental tool, the clinician would see that three of the four criteria for myalgia have been fulfilled and can conclude that the patient has probable myalgia.

It is unclear why the experimental tool added more incorrect diagnoses for the cases that had two diagnoses. The conditions represented in those cases were fairly evenly present in the three- and four-condition cases, where the experimental tool performed well. The number of subjects who used the experimental and control tools in the two-condition cases (n = 35 and 44, respectively) were comparable to the one-, three-, and four-condition cases. But to some degree, it is somewhat consistent with the overall trend of better performance in cases with more simultaneous conditions.

This comparison used hypothetical patients rather than live patients for several reasons. First, a larger number of subjects can test both tools for each condition. Second, all conditions in the DC/TMD can be more feasibly tested in sufficient numbers. Third, live patients would require the subjects to collect the data by patient interview and physical examination. The completeness and accuracy of data that they collect would then depend on their skill level in those two tasks, and that would affect the completeness and accuracy of their diagnoses. With these additional variables, it would be difficult to know how much of the differences in the performance of the two diagnostic tools was due to the tools, and how much was due to their skill in information-gathering.

This study has limitations. It was performed in students, interns, and residents, but not in practicing dentists. The purpose of the participant selection was to minimize the variability in TMD knowledge, but it is less representative of practicing dentists than if the participants came from practices. Future studies involving practicing dentists would have considerable value.

Some conditions were not tested in as many subjects as others (myofascial pain and disc displacement without reduction without limited opening). However, when designing the study, the focus was on having adequate sample sizes by number of conditions (cases with one condition, cases with two conditions, etc) and by general groupings (myalgias, arthralgia, headache, and disc disorders), rather than by individual conditions. Also, while hypothetical patients allow for larger sample sizes and more uniform testing conditions, they lack the real-world validation that live patients would provide. Neither the control tool nor the experimental tool have been tested in live patients.

Because there were no significant differences in how accurately the tools performed in the two sites, their results can be considered in whole. The most likely explanation for the longer time needed to complete the cases at NU is that while the diagnostic tools were translated into Japanese, the instructions, case descriptions, and answer sheets were not. If this explanation is correct, then that time difference is an artifact of the study design, not an indication of the tool's performance in Japan.

Finally, the DC/TMD has been expanded[10] beyond the conditions tested here. However, the objective of this study was to compare the experimental format to the validated control format, which does not include the expanded taxonomy.

Conclusions

In this study, the proposed checklist performed with a shorter time for identification of the correct diagnosis, comparable performance for simple TMD cases, and improved performance for complex TMD cases, with significantly more correct and fewer missed diagnoses compared to the existing diagnostic decision tree tool. However, further validation is needed before it can be considered an alternative diagnostic tool.

Highlights

- The experimental diagnostic tool arrived at correct TMD diagnoses in less time.
- The experimental diagnostic tool had more correct and fewer missed TMD diagnoses for complex cases.

Acknowledgments

Author contributions: A.Y.: design and execution of the study, handling of the data, and writing of the manuscript; S.G. and J.F.R.: handling of the data and the analysis; A.K.: execution of the study; O.A.K.: data analysis and writing of the manuscript; M.K. and J.K.: writing of the manuscript; N.N.: design and execution of the study and writing of the manuscript. The authors report no conflicts of interest and no external funding.

References

1. de Leeuw R, Klasser G (eds). Orofacial Pain: Guidelines for Assessment, Diagnosis, and Management, ed 6. Batavia, IL: Quintessence, 2019.
2. National Institute of Dental and Craniofacial Research. Prevalence of TMJD and its Signs and Symptoms. Updated July 2018. nidcr.nih.gov/research/data-statistics/facial-pain/prevalence. Accessed 12 July 2021.
3. Maixner W, Fillingim RB, Williams DA, Smith SB, Slade GD. Overlapping chronic pain conditions: Implications for diagnosis and classification. J Pain 2016;17(suppl 9):T93–T107.
4. Manfredini D, Guarda-Nardini L, Winocur E, Piccotti F, Ahlberg J, Lobbezoo F. Research diagnostic criteria for temporomandibular disorders: A systematic review of axis I epidemiologic findings. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2011;112:453–462.
5. Yost O, Liverman CT, English R, Mackey S, Bond EC (eds). Temporomandibular Disorders: Priorities for Research and Care. Washington, DC: National Academies Press, 2020.
6. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord 1992;6:301–355.
7. 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.
8. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD): Diagnostic Decision Trees. Pain-Related TMD and Headache. Intra-articular Joint Disorders. https://www.iadr.org/Portals/69/docs/Groups/INfORM/DC-TMD-Decision-trees_2013_06_08.pdf. Accessed 12 July 2021.
9. Raju PS. Product familiarity, brand name, and price influences on product evaluation. In: Perreault WD JR (ed). Advances in Consumer Research, volume 4. Atlanta: Association for Consumer Research, 1977:64–77.
10. Peck CC, Goulet JP, Lobbezoo F, et al. Expanding the taxonomy of the diagnostic criteria for temporomandibular disorders. J Oral Rehabil 2014;41:2–23.