Appropriate Use of Predictive Values in Clinical Decision Making and Evaluating Diagnostic Tests for TMD

Stephen R. Levitt, MD, PhD
Clinical Associate Professor of Psychiatry
University of North Carolina at Chapel Hill
Department of Psychiatry and Pain Program
Chapel Hill, North Carolina
and Director of Research
Pain Resource Center, Inc
Durham, North Carolina

Michael W. McKinney, PhD
Statistician
Pain Resource Center, Inc

Correspondence to:
Dr Steven R. Levitt
8770 Harmony Church Road
Efland, North Carolina 27243

Temporomandibular disorder literature contains serious misunderstandings and misapplications of statistical concepts, including predictive values, in evaluating diagnostic modalities and in clinical decision making. The use of general population prevalence data for temporomandibular disorders to evaluate positive predictive values of diagnostic modalities is shown to be invalid. The positive predictive value of a diagnostic tool should not be used to evaluate the efficacy of the tool or to confirm the presence of temporomandibular disorders when the pretest likelihood of temporomandibular disorder is low (eg, 10%). In such a situation, the TMJ Scale's negative predictive value of 98% supports the dentist's clinical impression of the absence of temporomandibular disorders. When the pretest likelihood of TMD is high (eg, 90%), the TMJ Scale's positive predictive value of 97% supports the dentist's clinical impression of the presence of temporomandibular disorders. The predictive values of the subscales of the TMJ Scale that measure joint dysfunction and stress may be used to further refine the diagnostic impression. When the dentist is unsure of the presence of TMD and makes a pretest estimate of 50%, the TMJ Scale's positive predictive value of 81% and negative predictive value of 83% substantially improve the accuracy of clinical decisions.
J OROFACIAL PAIN 1994;8:298–308.

A burgeoning interest is evident in the recent scientific and clinical literature on diagnostic tests for temporomandibular disorders (TMD).[1–7] At the same time, considerable controversy has evolved over which diagnostic modalities are valid and useful for screening and evaluating patients for TMD and for making clinical decisions.[8–16] The ADA Council on Dental Materials Instruments and Equipment (CDMIE) has published research guidelines that must be met for TMD diagnostic modalities to be considered for the ADA Acceptance Program.[17] The second edition of Temporomandibular Disorders: Guidelines for Classification, Assessment and Management,[18] published by the American Academy of Orofacial Pain (AAOP), emphasizes that the efficacy of all diagnostic tests should be established through studies that determine the sensitivity, specificity, and positive and negative predictive values of the test. It states that the sensitivity and specificity should generally be better than 70% and that these important characteristics are known for very few diagnostic tests for TMD.

A number of reviews of various diagnostic methods have appeared in the literature[19–27]; these are generally critiques of the research basis and performance of the various diagnostic techniques. Most of these reviews seem to conclude that no diagnostic modality has achieved sufficient sensitivity, specificity, and predic-

tive values to be considered useful for screening or clinical decision making.[19–25]

Critique of the Literature

Sensitivity, specificity, and predictive values are simply ratios expressing the probability of the occurrence of an event and are derived from statistical mathematics. However, many of the reviews of TMD diagnostic methods have used these mathematic concepts inappropriately by subjectively evaluating the results of these calculations and using them to infer "clinical truth."[28]

In a recent review of the theory and clinical application of these mathematic concepts to diagnosis of TMD, Christensen and Ash[28] discuss the shortcomings of many of these reviews of TMD diagnostic modalities. They point out that the establishment of the so-called "gold standard" for a clinical diagnosis, against which diagnostic techniques are tested, can be a highly spurious concept, and they discuss the fallacy of assuming that one can conclusively establish such a standard. They point out that TMJ imaging techniques and TMJ clinical examination techniques are effectively assigned a probability of 100% by certain researchers in arriving at a correct ("true") TMJ diagnosis. Their discussion emphasizes cause for concern, because the probability models of many clinical researchers, including Lund and Widmer,[23] Greene,[26] Goulet and Clark,[27] Widmer et al,[19] and Mohl et al,[20–22] "appear to be too optimistic and too oversimplified."[28]

Christensen and Ash[28] argue quite convincingly that many of these authors have used the concepts of probability to inappropriately infer the degree of belief in a particular diagnostic method. Their conclusion is that it is incorrect to subjectively evaluate, or pass judgment on, probability ratios such as sensitivity, specificity, and predictive values on the basis of an elusive gold standard, as is done in the TMD literature.[19–24,26,27] Christensen and Ash[28] conclude that the work of certain authors[24] uses the decision-making matrix as a framework for beliefs that are often far removed from "clinical truth."[28]

Dworkin et al[25] have recently discussed the validation of examination methods for TMD. Much attention was again paid to the importance of sensitivity, specificity, and predictive values. A prevalence of 10% for TMD in the general population was used as the basis for setting acceptable sensitivity and specificity as well as positive predictive values. Using this approach, they state that to achieve a positive predictive value of 75%, the specificity would have to be greater than 95% while the sensitivity could range between 70% and 100%. Various diagnostic methods are then evaluated using these concepts. However, the manner in which these concepts are employed once again represents a serious misunderstanding of the proper application of predictive values in evaluating diagnostic modalities and in clinical decision making. Specifically, the application of population prevalence data for evaluating diagnostic methods and for clinical decision making on individual patients is entirely inappropriate. Their misapplication of predictive values may lead to erroneous conclusions and calls into question much of their results.

A careful examination of the Dworkin et al[25] results shows that using the recommended TMD prevalence of 10%, no diagnostic tool with a sensitivity of even 100% and a specificity of 90%, nor any diagnostic tool with a specificity of 100% and a sensitivity of 90%, would have a positive predictive value even close to 75%. Clearly, the approach of using population prevalence data for TMD to evaluate positive predictive values of diagnostic tools eliminates the possibility that almost any tool can meet the acceptable criteria. Although many diagnostic tools do lack sufficient accuracy and science for general use, another problem is the repeated inappropriate application of statistical concepts in this field.

The incorrect application of predictive values to certain clinical situations is again exemplified by the suggestion that although the sensitivity (84%) and specificity (80%) of one particular diagnostic tool, the TMJ Scale (Pain Resource Center, Durham, NC), are acceptable, this tool is not suitable for screening for TMD because its positive predictive value is too low at a prevalence rate of 10% for TMD.[29] A prevalence of 10% was apparently again chosen because it represents the approximate prevalence of TMD in the general population. The error is again made of using general population prevalence rates for TMD to evaluate diagnostic tools in clinical settings. In addition, positive predictive values were used when, in fact, it is the negative predictive values that are required.[30]

It is not widely appreciated that the proper use of predictive values is based on using the clinician's best pretest estimate of the presence of disease or disorder in the individual patient being evaluated.[30] This estimate is always based on the history and examination of the particular patient and virtually never on population prevalence data. The ADA and AAOP have recommended that

every dental patient be screened for TMD.[18,31] This screening usually takes the form of a brief history and clinical examination covering certain key elements of TMDs.[18] The only time it would be appropriate to use population prevalence data as the best pretest estimate of TMD is during epidemiologic studies in the general population, where a diagnostic tool is used to screen for TMD and where no clinical information on the individuals being tested is available. In such a case, the positive predictive value of a diagnostic tool for TMD would necessarily be applied.

The use of general population prevalence data to estimate a patient's probability of having TMD is a last resort. Use of such an approach would be entirely invalid in the many dental practices that specialize in facial pain and TMD. In these practices the prevalence of TMD would be more accurately estimated at 50% to 100%. Even in such practice settings, the best estimate of the likelihood of a patient having TMD is a good history and clinical examination. The point is that population or practice prevalence data is never as good as an estimate based on the specific history and clinical examination results for the individual patient.

Further confusion has appeared in certain reviews regarding the proper selection of a diagnostic method and appropriate use of predictive values in clinical decision making.[29] As will be shown below, when the dentist's pretest estimate of TMD is low, eg, 10%, then a diagnostic tool should be chosen to help confirm the dentist's prediction of absence of TMD and not attempt to contradict the dentist's prediction by confirming presence of TMD.[30] In such a case, it is the negative predictive value that is used, not the positive predictive value.[30] Using one reviewer's example of the TMJ Scale,[29] the negative predictive value of the TMJ Scale at a pretest estimate of 10% for TMD is 98%, a very high and acceptable level of prediction.[32] Whenever the likelihood of the presence of a disease or disorder is low in a given patient, the positive predictive value of a diagnostic tool should never be used to contradict the clinician's impression by attempting to rule in the disorder.[30] The fallacy of this approach is that diagnostic tools in general have low positive predictive values when the pretest estimate of disorder is low. This leads to excessively high false-positive rates.

When the dentist believes that a TMD is probably present (pretest estimate of 51% to 99%), it may be decided to corroborate this prediction by appropriately employing a diagnostic tool that has a high positive predictive value for TMD. When the dentist believes that a TMD is probably absent (pretest estimate of less than 50%), then it may be decided to corroborate this prediction by appropriately employing a diagnostic tool that has a high negative predictive value for TMD.

In most instances, when the dentist feels that a TMD is not present, no diagnostic tool will be employed to further rule out TMD. The exception is the case of the orthodontist, surgeon, or other specialist who plans on performing an irreversible procedure. In such a case, the dentist may desire to more strongly support and corroborate the presumed absence of TMD through use of a diagnostic test with a high negative predictive value.

Certain reviews also contain some important omissions of published research findings and gross inaccuracies regarding the availability of sensitivity, specificity and predictive values for particular diagnostic tools. For example, in one such review in 1992, it was stated that "Assessing the diagnostic validity of the TMJ Scale was reported in one study" and "Unfortunately, no measures of sensitivity, specificity or positive predictive values were included in their study, and since there are no reported numbers of successful or unsuccessful diagnoses by their indices it is not possible to calculate sensitivity or specificity."[25] The sensitivity and specificity of all 10 subscales of the TMJ Scale were first reported in 1985 at the Annual Meetings of the International Association for Dental Research and American Pain Society by Lundeen et al.[34,35] They were again published in the TMJ Scale technical manual in 1986.[36] Sensitivity and specificity studies based on extensive cross-validation samples were published again in 1988.[37] This was followed by three separate publications on the predictive values of this diagnostic tool in 1990 and 1991.[32,38,39] A total of 16 publications on the TMJ Scale[32,34-48] were in print at least 1 year before the above incorrect statements,[25] and since then, additional published studies by numerous investigators continue to confirm the TMJ Scale's validity and clinical usefulness.[49-53]

The confusion generated by omission of important research findings and published misinformation (see erratum[33]) is compounded by an inconsistency and an apparent double standard in the literature. Certain authors,[19,20-22,25] demand adequate sensitivity, specificity, and predictive value studies for all diagnostic methods, yet they employ diagnostic methods of their own for which no sensitivity, specificity, or predictive value research has been published.[25] Specifically, drastically modified versions of two subscales of the SCL-90-R[54] are used extensively without having their internal structure

revalidated, with no internal consistency or test-retest reliability studies published, and with no published sensitivity, specificity, or predictive values.[25]

Focus of the Present Study

Few diagnostic tools actually make a direct prediction regarding presence of TMD. Therefore, predictive values for such tools are scarce. One diagnostic tool that does make a direct prediction of presence or absence of TMD is the TMJ Scale.[32,34–53] This diagnostic tool also makes similar predictions for the physical symptoms of TMD including pain, palpation pain, joint dysfunction, and range of motion limitation, as well as for psychological factors and stress, among other scales. This tool will be used to illustrate how the efficacy of a TMD diagnostic method should be evaluated and to demonstrate, using a number of clinical scenarios, the proper application of the theory of predictive values to clinical decision making.

Calculation of Sensitivity, Specificity, and Predictive Value

Sensitivity is the probability that a test result will be positive when the disease is present. Specificity is the probability that a test result will be negative when the disease is absent.[30] These mathematic concepts may be calculated using a binary decision matrix diagram (Fig 1) in which:

$$\text{Sensitivity} = TP/TP + FN$$

and

$$\text{Specificity} = TN/TN + FP$$

Sensitivity and specificity are expressions of probability as applied to groups of patients, as opposed to individual patients. Although a sensitivity and specificity of at least 70% are sometimes considered the minimum requirement for a useful TMD diagnostic method,[18] the sensitivity and specificity desired is actually situation specific. A high sensitivity is needed for screening or excluding a disease and high specificity is needed to confirm a disease.[30] Since the application of a diagnostic tool may vary depending on the specific clinical need, many psychometric diagnostic tools attempt to set the cut-off scores at a level that simultaneously maximizes both sensitivity and specificity.[36]

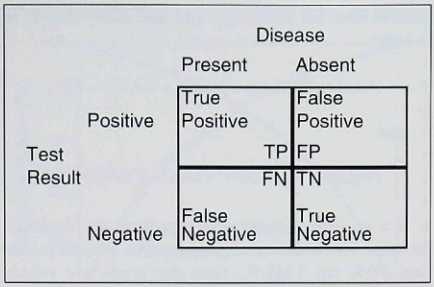

Fig 1 Binary decision matrix diagram.

Setting the cut-off score lower will increase sensitivity, ie, have a higher hit rate for true positives, while at the same time reduce specificity. Likewise, raising the cut-off score will increase specificity, ie, produce a higher hit rate for true negatives, while at the same time reduce sensitivity.

The consequences of low sensitivity or low specificity differ depending on the risks and costs of treating patients who do not have the disease compared to underdiagnosing and not treating patients who have the disease. Some believe that many TMDs are self-limiting and resolve without serious long-term effects.[55,56] It is felt that aggressive, irreversible, and costly treatments such as complex occlusal therapy and surgery should be avoided in most cases.[18] If this is true, then the costs and risks associated with low specificity (high false positives) would appear to be a greater problem than missing and not treating patients who have TMD. Also, since the prevalence of TMD is low (presumed about 10%) in the general population, there will be a small number of false negatives resulting from a lower sensitivity as opposed to a relatively large number of false positives resulting from a lower specificity.[25] For this reason, some researchers feel that specificity should be maximized at the expense of sensitivity.[25]

The predictive values of a test give the clinician a better estimate of its accuracy in a given clinical situation for the individual patient.[30] The positive predictive value of a diagnostic tool is the probability that a patient with a positive test result actually has the disease. The negative predictive value is the probability that a patient with a negative test result does not have the disease. These two measures of probability of occurrence of disease may be calculated using the same binary decision

matrix used for sensitivity and specificity (Fig 1) in which:

Positive Predictive Value = TP/TP + FP

and

Negative Predictive Value = TN/TN + FN

If a given sensitivity and specificity for diagnostic tests for TMD are determined to be acceptable (eg, 70% for TMD[18]), then the predictive values for those tests at all prevalence or pretest estimates of disease are also determined. It should be noted that the predictive values of a diagnostic test are affected by the prevalence or pretest likelihood of disease presence. In fact, calculation and use of predictive values depends on being able to make a pretest estimate of disease presence for the patient being tested.

Sensitivity, Specificity, and Predictive Values of the TMJ Scale

The TMJ Scale is a TMD symptom–based, psychometrically derived tool with an extensive research basis and numerous studies discussing its clinical application.[32,34–53] The theory used to construct and develop the TMJ Scale and the research basis for its validity, reliability, and clinical utility have been previously reported.[32,34–51,53] It contains subscales to measure both the physical and psychological components of TMDs. The sensitivities, specificities, and predictive values of the 10 subscales of the TMJ Scale have been thoroughly studied.[32,36–39]

The present study will focus on three subscales (Global, Joint Dysfunction, and Stress Scales) to explore the appropriate use of predictive values in clinical decision making and in evaluating diagnostic tests for TMD.

The Global Scale makes a direct prediction as to whether the patient has a TMD.[32,36,37] The patient's score on this scale can be used to help screen and rule in or exclude and rule out the presence of TMD. The Joint Dysfunction Scale is sensitive to TMJ noises (clicking, popping, grating, grinding, crepitus) and locking (open and closed).[36,37,39] The Stress Scale measures the overall level of reported stress.[36–38]

Table 1 shows the sensitivities and specificities of these subscales.[36,37] Figures 2 through 4 graphically show the results of predictive value calculations for these subscales.[32,38,39] Some general observations may be made at this point. The positive predictive value of any diagnostic method is relatively low at low pretest estimates of disease (disease in the present context means TMD or the particular symptom of TMD such as pain or joint dysfunction) and increases as the pretest estimate of disease increases from 0% to 100%. The negative predictive value behaves in an opposite manner, being relatively high at low pretest estimates of disease and decreasing as the pretest estimate increases. When the pretest estimate of disease is low, the negative predictive value is higher than the positive predictive value, and more confidence is placed in negative test results. When the pretest estimate of disease is high, the positive predictive value is greater than the negative predictive value and more confidence is placed in positive test results.[30]

Table 1 Sensitivity and Specificity of the Global Scale, Joint Dysfunction, and Stress Scales of the TMJ Scale[36,37]

Scale name	Study size	Sensitivity (%)	Specificity (%)
Global Scale	1215	84.2	80.3
Joint Dysfunction	808	76.3	73.1
Stress	197	73.7	71.1

From a clinical decision making point of view, these results for all diagnostic tests mean that a positive test result is not helpful in confirming the presence of a disease when the pretest likelihood of disease is low.[30] Thus, using the estimated population prevalence of TMD of 10% as the pretest estimate precludes the use of positive predictive values to evaluate the diagnostic method and prevents the use of a positive test result in screening for or confirming TMD. However, when the pretest likelihood of disease is low, a negative test result is helpful in excluding the disease.[30] This means that if, in the unlikely circumstance that the population prevalence of TMD of 10% is used as the pretest estimate, a diagnostic method with a high negative predictive value is desirable and a negative test helps the dentist rule out TMD.

The above observations represent the fundamental principles involved in using predictive values in clinical decision making. It is worth emphasizing again that predictive values should only be used in conjunction with the dentist's pretest estimate of the likelihood of presence of TMD or the particular symptom being evaluated. This estimate should be based on clinical information obtained from the history and examination of the patient, not on estimates of TMD in the general population.

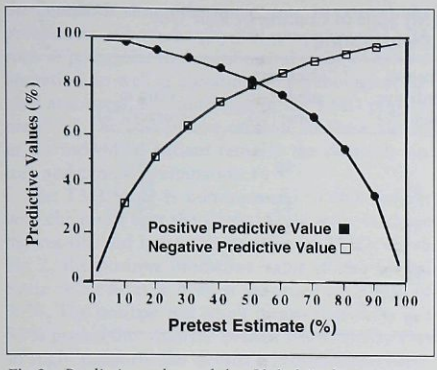

Fig 2 Predictive values of the Global Scale versus pretest estimates of the likelihood of TMD.

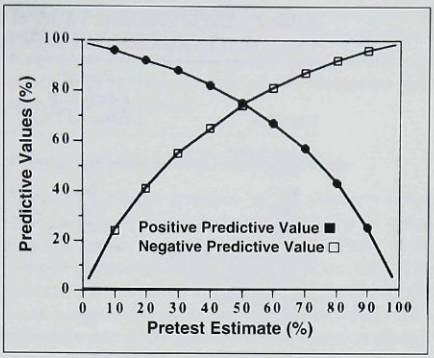

Fig 3 Predictive values of the Joint Dysfunction Scale versus pretest estimates of the likelihood of presence of joint dysfunction.

Discussion

Pretest Estimate of TMD is Low

Let it be supposed that a patient presents to the dentist's office with a chief complaint of facial pain. The history reveals that the patient has been having pain in the area of the left maxillary third molar for about 2 weeks. There has been a prior history of similar pain and dental work on that particular tooth in the past. Examination of the tooth reveals it to be sensitive to percussion, which reproduces the pain. The TMD screening history reveals mild pain on opening in the region of the left TMJ, no stiffness or difficulty on opening or masticating, and no history of locking. The patient does report an occasional clicking noise in the left TMJ. No history of injury is elicited. There is some generalized tenderness in the left masseter muscle. No preauricular or intrameatal TMJ tenderness is found. No click or crepitus is present on auscultation, and the range of motion of the mandible is 42 mm vertical and 7 mm bilateral laterotrusion. The occlusion is felt to be adequate. The remainder of the TMD screening examination is normal. A large carious lesion is noted on panoramic radiograph at the bottom of a prior restoration and in very close proximity to the dentin.

Treatment of the third molar carious lesion is indicated, and the possible need for a root canal is noted. Based on this history and examination, the dentist estimates that the likelihood of a TMD is no more than 10%. The dentist wishes to establish more firmly the absence of a TMD before an

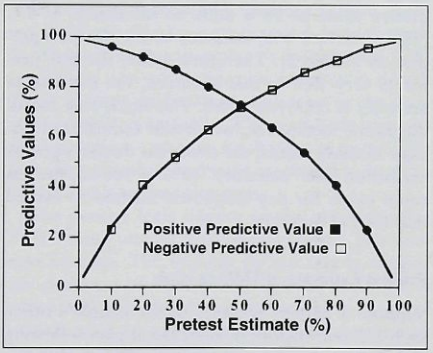

Fig 4 Predictive values of the Stress Scale versus pretest estimates of the likelihood of presence of stress.

attempt at restoration or referral to an endodontist for root canal. The dentist now needs to corroborate his impression that a TMD is absent. A diagnostic test is required that can rule out the presence of TMD with a high degree of certainty. The tool must have a high negative predictive value at a pretest estimate for presence of TMD of 10%.

The TMJ Scale is administered with the result that the Global Scale score is below the cut-off and indicates absence of TMD. Using Fig 2, the negative predictive value of the Global Scale is 98% at a pretest estimate of TMD of 10%. The negative test result means that there is a 98% probability

Table 2 Use of Predictive Values of the TMJ Scale to Confirm or Rule Out TMD, Joint Dysfunction, and Stress in the Dental Patient

Disorder	Pretest estimate of likelihood (%)	Positive predictive value (%)	Negative predictive value (%)	Gain in accuracy (%)
Global Scale[32]				
TMD	10	–	98	8
TMD	50	81	–	31
TMD	50	–	83	33
TMD	90	97	–	7
Joint Dysfunction Scale[39]				
Joint Dysfunction	80	92	–	12
Stress Scale[38]				
Stress	60	79	–	19

that the patient does not have TMD. This strongly supports the dentist's pretest impression.

Note that the dentist's pretest estimate of 10% means that he felt the likelihood of absence of TMD was 90%. The Global Scale's negative predictive value of 98% adds an increment of 8% (98% – 90% = 8%) accuracy to the dentist's prediction (Table 2). Therefore, while the test performs very well in this situation, the increase in accuracy is relatively small. This holds true for all diagnostic methods at low pretest estimates of disease. In other words, the closer the dentist's pretest prediction is to "certainty" (0% or 100%), the less room exists for any diagnostic method to exceed that level of accuracy.

Pretest Estimate of TMD is High

Suppose a patient presents to the dentist's office with a 2-year history of right facial pain following a fall during which the patient's face hit the ground with considerable force. The patient noticed immediate pain in the right jaw joint and surrounding areas. He was seen by his regular dentist who, after taking a radiograph, told the patient that there was no fracture. Over the next several weeks the patient noticed stiffness, pain, and difficulty in opening his jaw when yawning or masticating. A few months later the patient reports occasional episodes during which the jaw got stuck open about half way. There is a loud popping sound on opening wide and this is sometimes accompanied by pain in the jaw joint. All of these symptoms have persisted for 2 years. Over this time the patient has come to feel that his bite is off or changing. He also now experiences headaches, usually in the right temple region and across the right face. The patient reports feeling occasionally stressed to a mild degree by the pain and problems with jaw function, and that these problems have interfered with some of his usual leisure activities.

The dentist proceeds with a screening examination for TMD. The vertical range of motion is measured at 35 mm, with an opening click at 23 mm. Right and left laterotrusion are 6 mm and 7 mm, respectively. There is preauricular and intrameatal tenderness to palpation on the right side. Auscultation reveals crepitus in the right TMJ on opening and closing. The masseter and temporalis muscles are tender on the right side. There is a Class II, grade II malocclusion.

At this point in the evaluation the dentist believes that the data strongly support the presence of a TMD, probably primarily involving the right TMJ as well as the muscles of mastication on the right side. Before proceeding with more extensive examination, radiography, and other expensive imaging techniques to further help define the diagnosis, the dentist wishes to further confirm his impression of the presence of TMD and confirm TMJ involvement. He also is concerned about the patient's level of stress. The dentist makes a pretest estimate for presence of TMD of 90%. A diagnostic tool that can confirm and rule in the presence of TMJ with a high accuracy is needed. The tool must have a high positive predictive value at a pretest likelihood of TMD at 90%.

The dentist also feels, quite strongly, that the data suggest that TMJ dysfunction is involved and somewhat less strongly that the patient's stress level is abnormally high. The dentist makes a pretest estimate for the presence of TMJ dysfunction of 80% and for presence of elevated stress of 60%. Therefore, a diagnostic tool is needed to confirm and rule in the dentist's diagnostic impression of the presence of joint dysfunction and stress. It is desirable for such a test to have high positive predictive values for these symptoms at the pretest estimate levels of

the dentist. It should be noted that, in general, the prevalences of both the physical symptoms of TMD, such as pain, joint dysfunction, and range of motion limitation, as well as elevations in psychological distress and stress, are quite high in the TMD population.[36,37,40] The best pretest estimate of these factors in the individual patient remains the dentist's history and clinical examination.

The TMJ Scale is administered to this patient with the result that the Global Scale score is above the cut-off and indicates presence of TMD. Using Fig 2, the positive predictive value of the Global Scale is 97% at a pretest estimate of TMD of 90%. The positive test result means that there is a 97% probability that the patient has a TMD. This strongly supports the dentist's pretest impression. The increment in accuracy over the dentist's prediction is 7% (97% – 90% = 7%) (Table 2). Again, at 90% likelihood there is not much room for improvement.

The dentist also may use the TMJ Scale to help better define the TMD and the patient's response to it. The history and physical examination suggest TMJ involvement. The dentist's pretest prediction is 80% for TMJ dysfunction. The Joint Dysfunction Scale score is elevated above the cut-off and predicts the presence of joint dysfunction. The positive predictive value of the Joint Dysfunction Scale is 92% at a pretest estimate of 80%. This result means that the probability of the patient having joint dysfunction is 92% and this strongly confirms the dentist's prediction. The improvement in accuracy over the dentist's pretest prediction is 12% (92% – 80% = 12%) (Table 2). In a similar manner, the dentist may use the Pain, Palpation Pain, Range of Motion Limitation, and Perceived Malocclusion Scales of the TMD Scale to further refine diagnostic impression.[36,37,39]

The Stress Scale score is also above the cut-off and predicts the presence of stress as a significant factor in the patient's response to the TMD symptoms. The positive predictive value of the Stress Scale is 79% at a pretest estimate of 60%. The positive test result means that the probability of the patient experiencing significant stress is 79%, and this tends to confirm the dentist's prediction. It also improves the accuracy of the dentist's pretest prediction by 19% (79% – 60% = 19%) (Table 2). In a similar manner, the dentist may also use the Psychological Factors and Chronicity Scales of the TMJ Scale to further assess the patient's psychological state.[36–38]

With the resulting confirmations, the dentist may proceed with a higher degree of certainty in expending time and resources to further define the TMD. He may also discuss the test results with the patient and indicate that, based on the patient's responses, a referral for stress management appears appropriate as part of the treatment and management protocol.

Pretest Estimate of TMD is Intermediate

Suppose a patient presents to the dentist's office with a 3-month history of some difficulty in opening the mouth as wide as previously. There is no history of trauma to the head, neck, or face. The patient reports some occasional mild pain in the area of both jaw joints when chewing. The mandible has never gotten "stuck" open or closed. There is an infrequent clicking sound in the left TMJ, occurring only once or twice every 6 months or so. The jaws don't feel stiff and there is no history of headaches. The bite is not reported to feel off or changed. There is no prior history of treatment for a jaw-joint problem.

The clinical examination reveals a vertical opening of 40 mm and laterotrusion of 6 mm bilaterally. There is no preauricular or intrameatal TMJ tenderness and no clicking or crepitus is noted. There is some mild diffuse tenderness in the masseter muscles bilaterally. The occlusion appears normal and no excessive tooth wear is found. The face, dental arches, and jaws are found to be symmetric and in proper alignment.

The dentist feels unsure of the presence of a TMD based on this history and clinical examination findings. The findings do not clearly suggest the presence of TMD, although the findings of occasional pain, borderline limitation in range of motion, and an occasional click seem suspicious. Suppose in this situation the dentist's pretest estimate for presence of TMD is 50%. This is precisely the clinical situation in which the dentist may be helped most by diagnostic screening tools for TMD.

Whenever the dentist is very sure of the presence of TMD (pretest likelihood of TMD estimated at 90% or greater) or very sure of the absence of TMD (pretest likelihood of TMD estimated at 10% or less), no diagnostic test can improve much on the dentist's accuracy. The maximum improvement in accuracy of prediction is less than 10% even if the diagnostic tool has a positive and negative predictive value of 100% (not possible).

In examining the positive and negative predictive value curves (Figs 2 through 4), it is easily seen that a combination of high predictive values combined with substantial improvements in accuracy over the dentist's pretest estimates occur over the midrange of pretest estimates. This means a diag-

nostic tool can be of most use when the dentist is unsure of the presence of the disorder or entity for which a test is being made.

For example, the positive predictive value of the Global Scale (Fig 2) varies from 74% at a pretest estimate of presence of TMD of 40% (producing a 34% increase in accuracy) to 86% at a pretest estimate of presence of TMJ of 60% (producing a 26% increase in accuracy). The negative predictive value of the Global Scale varies from 77% at a pretest estimate of 60% (producing a 37% increase in accuracy), to 88% at a pretest estimate of 40% (producing a 28% increase in accuracy).

Using the dentist's pretest estimate for presence of TMD of 50%, the Global Scale has a positive predictive value of 81% and a negative predictive value of 83%. If the Global Scale score is above the cut-off and predicts the presence of TMD, then there is an 81% probability that the patient has TMD and use of the TMJ Scale has produced a gain in accuracy of 31% (Table 2). If, on the other hand, the Global Scale score is below the cut-off and predicts the absence of TMD, then there is an 83% probability that the patient does not have TMD and use of the TMJ Scale has produced a gain in accuracy of 33% (Table 2). In either case, use of the TMJ Scale helps the dentist make a clinical decision regarding presence or absence of TMD by providing estimates of probability of occurrence of TMD that are high and substantially above the probability estimates made by the dentist.

Conclusion

The results of probability studies as related to sensitivity, specificity, and predictive value ratios frequently have been judged subjectively and used inappropriately to infer the degree of belief in a particular diagnostic method for TMD. The TMD literature contains serious misunderstandings of the proper application of statistical concepts, including predictive values, in evaluating diagnostic modalities and in clinical decision making The frequent approach of using general population prevalence data for TMD to evaluate positive predictive values of diagnostic modalities, and thereby the efficacy of these modalities, is shown to be invalid and highly misleading. The use of predictive values depends on having available the dentist's best pretest estimate of the presence of TMD. This estimate is always based on a proper history and clinical examination, not prevalence data. When the dentist's pretest estimate for the presence of TMD is low, ie, below 50%, a diagnostic method is needed to confirm the absence of TMD and the modality chosen should have a high negative predictive value. When the dentist's pretest estimate for presence of TMD is high, ie, greater than 50%, a diagnostic method is needed to confirm the presence of TMD and the modality chosen should have a high positive predictive value. The positive predictive value of any diagnostic tool is not helpful and should not be used to confirm the presence of TMD when the pretest likelihood of TMD is low. Similarly, in situations where the pretest likelihood of TMD is low, the positive predictive value of a diagnostic tool should not be used to evaluate the efficacy of that tool.

The use of general population estimates for prevalence of TMD of about 10% should not be used to pass judgement on diagnostic modalities for TMD. In clinical settings, the best pretest estimate of TMD is always based on the history and clinical examination of the individual patient. When the pretest likelihood of TMD is low (eg, 10%), a diagnostic modality such as the TMJ Scale, which has a negative predictive value of 98%, can be used to support the dentist's clinical impression of absence of TMD. When the pretest likelihood of TMD is high (eg, 90%), a diagnostic modality such as the TMJ Scale, which has a positive predictive value of 97%, can be used to support the dentist's clinical impression of presence of TMD.

The subscales of the TMJ Scale that measure physical symptoms, such as joint dysfunction, or psychological factors, such as stress, have predictive values of such a magnitude that they too may be used in a similar manner to further refine the diagnostic impression by confirming or ruling out the presence of TMJ involvement and stress in the TMD patient. The use of the TMJ Scale in this manner adds a substantial amount of accuracy to the dentist's clinical decision making. In the situation where the dentist is unsure of the presence of TMD and makes a pretest estimate of 50%, the TMJ Scale's positive predictive value of 81% and negative predictive value of 83% allow this diagnostic and screening tool to help the dentist make a clinical decision regarding presence or absence of TMD with a substantially higher degree of accuracy than without the use of this tool.

References

1. Hatcher DC. Craniofacial imaging. J Calif Dent Assoc 1991;19:27–34.
2. Kozeniauskas JJ, Ralph WJ. Bilateral arthrographic evaluation of unilateral TMJ pain and dysfunction. J Prosthet Dent 1988;60:98–105.

3. Helms CA, Doyle GW, Orwig D, McNeill C, Kaban L. Staging of internal derangement of the TMJ with magnetic resonance imaging: Preliminary observations. J Craniomandib Disord Facial Oral Pain 1989;3:93–99.
4. Toolsen GA, Sadowsky C. An evaluation of the relationship between temporomandibular joint sounds and mandibular movement. J Craniomandib Disord Facial Oral Pain 1991;5:187–196.
5. Schroeder H, Siegmund H, Santibanez H-G, Kluge A. Causes and signs of temporomandibular joint pain and dysfunction: An electromyographical investigation. J Oral Rehabil 1991;8:301–310.
6. Berry DC, Yemm R. Variations in skin temperature on the face in normal subjects and in patients with mandibular dysfunction. Br J Oral Surg 1971;8:242–247.
7. Davidson SL. Doppler auscultation. J Craniomandib Disord Facial Oral Pain 1988;2:128–132.
8. Knoernschild KL, Aquilino SA, Ruprecht A. Transcranial radiography and linear tomography: A comparative study. J Prosthet Dent 1991;66:239–250.
9. Dixon DC, Graham GS, Mayhew RB, Oesterle LJ, Simms D, Pierson WP. The validity of transcranial radiography in diagnosing TMJ anterior disk displacement. JADA 1984;108:615–618.
10. Fava C, Gatti G, Cardes E, Parchetti R, Rocca G, Preti G. Possibilities and limits in identifying the TMJ articular meniscus with the CT scanner: A comparative anatomoradiological study. J Craniomandib Disord Facial Oral Pain 1988;2:141–147.
11. Helms CA, Morrish RB, Kircos LT, Katzberg RW, Dolwick WF. Computed tomography of the meniscus of the temporomandibular joint: Preliminary observations. Radiology 1982;145:719–722.
12. Kircos LT, Ortendahl DA, Mark AS, Arakawa M. Magnetic resonance imaging of the TMJ disc in asymptomatic volunteers. J Oral Maxillofac Surg 1987;45:852–854.
13. Velasco J, Tasaki T, Gale E. Study of pantographic tracings of TMD patients and asymptomatic subjects [abstract]. J Dent Res 1991;70(special issue).
14. Rugh JD, Davis SE. Accuracy of diagnosing MPD using electromyography [abstract]. J Dent Res 1990;69(special issue):273.
15. Gratt BM, Sickles EA, Graff-Radford SB, Solberg W. Electronic thermography in the diagnosis of atypical odontalgia: A pilot study. Oral Surg Oral Med Oral Pathol 1989;68:472–481.
16. Widmer CG. Temporomandibular joint sounds: A critique of techniques for recording and ana1ysis. J Craniomandib Disord Facial Oral Pain 1989;3:213–217.
17. Council on Dental Materials, Instruments and Equipment. American Dental Association Acceptance Program guidelines for instruments as aids in the diagnosis of temporomandibular disorders. American Dental Association, Chicago, IL, August 1991.
18. American Academy of Orofacial Pain [McNeill C, ed]. Temporomandibular Disorders—Guidelines for Classification, Assessment, and Management, ed 2. Chicago: Quintessence, 1993.
19. Widmer CG, Lund JP, Feine JS: Evaluation of diagnostic tests for TMD. J Calif Dent Assoc 1990;18:53–60.
20. Mohl ND, McCall WS, Lund JP, Plesh O. Devices for the diagnosis and treatment of temporomandibular disorders: Part I. Introduction, scientific evidence, and jaw tracking. J Prosthet Dent 1990;63:198–201.
21. Mohl ND, Lund JP, Widmer CG, McCall WD. Devices for the diagnosis and treatment of temporomandibular disorders. Part II. Electromyography and sonography. J Prosthet Dent 1990 63:332–336.
22. Mohl ND, Ohrbach RK, Crow HC, Gross AJ. Devices for the diagnosis and treatment of temporomandibular disorders. Part III. Thermography, ultrasound, electrical stimulation and EMG biofeedback. J Prosthet Dent 1990;63:472–477.
23. Lund JP, Widmer CG. Evaluation of the use of surface electromyography in the diagnosis, documentation and treatment of dental patients. J Craniomandib Disord Facial Oral Pain 1989:3:125–137.
24. Stockstill JW, Mohl ND. Evaluation of temporomandibular joint sounds. Diagnostic analysis and clinical applications. Dent Clin N Am 1991;35:75–80.
25. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examination and specifications, critique. J Craniomandib Disord Facial Oral Pain 1992;6:301–355.
26. Greene CS. Can technology enhance TM disorder diagnosis? J Calif Dent Assoc 1990;18:21–24.
27. Goulet J-P, Clark GT. Clinical TMJ examination methods. J Calif Dent Assoc 1990;18:25–33.
28. Christensen LV, Ash MM. Remarks on probability theory and TMJ diagnosis. J Oral Rehabil 1992;19:561–567.
29. Mohl ND. Current status of diagnostic modalities. Scientific Frontiers in Clinical Dentistry—An update at the National Institute of Dental Research, Bethesda, April 15–16, 1993.
30. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures: Principles and applications. Ann Int Med 1981; 94:553–600.
31. Griffith RH. Report of the president's conference on the examination, diagnosis, and management of temporomandibular disorders. JADA 1983;106:75–78.
32. Levitt SR. Predictive value: A model for dentists to evaluate the accuracy of diagnostic tests for temporomandibular disorders as applied to the TMJ Scale. J Prosthet Dent 1991;66:385–390.
33. Dworkin SF, et. al. Errata. J Orofacial Pain 1993;7:234.
34. Lundeen TF, Levitt SR, McKinney MW. TMJ Scale: Discriminative ability of a new diagnostic tool [abstract]. J Dent Res 1985;64(special issue):357.
35. Lundeen TF, Levitt SR, McKinney MW. The discriminative ability of the TMJ Scale: Age and gender differences. Am Pain Soc 1985; 5th Ann Meeting, Dallas, Oct 17–20.
36. Levitt SR, Lundeen TF, McKinney MW. The TMJ Scale Manual. Durham, NC: Pain Resource Center, 1987.
37. Levitt SR, McKinney MW, Lundeen TF. The TMJ Scale: Cross-validation and reliability studies. J Craniomand Pract 1988;6:17–25.
38. Levitt SR. The predictive value of the TMJ Scale in detecting psychological factors and non-TM disorders in the patient with a temporomandibular disorder. J Craniomand Pract 1990;8:225–233.
39. Levitt SR: Predictive value of the TMJ Scale in detecting clinically significant symptoms of temporomandibular disorders. J Craniomandib Disord Facial Oral Pain 1990; 4:177–185.
40. Lundeen TF, Levitt, SR, McKinney MW. Evaluation of temporomandibular joint disorders by clinician ratings. J Prosthet Dent 1988;59:202–211.
41. Levitt SR, Lundeen TF, McKinney MW. Initial studies of a new assessment method for temporomandibular joint disorders. J Prosthet Dent 1986;59:490–495.
42. Lundeen TF, Levitt SR, McKinney MW. TMJ Scale: Discriminative ability of the TMJ Scale: Age and gender

differences. J Prosthet Dent 1986;56:84–92.
43. Lundeen TF, Levitt SR, McKinney MW. Clinical applications of the TMJ Scale. J Craniomand Pract 1988;6: 339–345.
44. Lundeen T, Scruggs R, McKinney M, Daniel S, Levitt S. TMD symptomatology among denture patients. J Craniomandib Disord Facial Oral Pain. 1990;4:40–45.
45. Spiegel EP, Levitt SR. Measuring symptom severity and treatment outcome of temporomandibular disorders with the TMJ Scale: Case report. J Craniomand Pract 1990; 8:353–358.
46. Spiegel EP, Levitt SR. Measuring symptom severity with the TMJ Scale. J Clin Orthodont 1991;25:21–26.
47. Levitt SR, Spiegel EP, Claypoole WH. The TMJ Scale and undetected brain tumors in patients with temporomandibular disorders. J Craniomand Pract 1991;9: 152–158.
48. Levitt SR, Lundeen TF. Quantitative measurements of symptoms and treatment results. TMJ Update 1987; 5:77–80.
49. Pocock P, Mamandras A, Bellamy N. Evaluation of an anamnestic questionnaire as an instrument for investigating potential relationships between orthodontic therapy and temporomandibular disorders. Am J Orthod Dentofac Orthop 1992;102:239–243.
50. Levitt SR, McKinney MW, Willis WA. Measuring the impact of a dental practice on TM disorder symptoms. J Craniomand Pract 1993;11:211–216.
51. Levitt SR, McKinney MW. Validating the TMJ Scale in a national sample of 10,000 patients: Demographic and epidemiologic characteristics. J Craniomandib Disord Facial Oral Pain 1994;8:25–35.
52. Willis WA. The effectiveness of an extreme canine-protected splint with limited lateral movement in treatment of temporomandibular dysfunction. Am J Ortho Dentofac Orthop 1993 (In press).
53. Chibnall JT, Duckro PN, Greenberg, MS. The TMJ Scale: Evidence for construct validity. Submitted for publication, 1993.
54. Derogatis L. SCL-90-R: Administration, Scoring and Procedures Manual—II. Towson, MD: Clinical Psychometric Research, 1983.
55. Greene CS, Laskin DM. Long-term evaluation of treatment for myofascial pain-dysfunction syndrome. A comparative analysis. JADA 1983;107:235–238.
56. Mejersjo C, Carlsson GE. Long term results of treatment for temporomandibular pain-dysfunction. J Prosthet Dent 1983;49:809–815.

Resumen

La utilización apropriada de valores estimativos para la toma de decisiones clínicas y la evaluación de exámenes diagnósticos para los desórdenes temporomandibulares

La literature relacionada a los desórdenes temporomandibulares (DTM) contiene una serie de malentendidos y aplicaciones erradas de conceptos estadísticos, incluyendo los valores estimativos, en la evaluación de las modalidades diagnósicas y en la tome de decisiones clínicas. El uso de la información sobre la frecuencia de DTM en la población general para evaluar los valores estimativos positivos de las modalidades diagnósticas ha side inválida. El valor estimativo positivo de un medio de diagnóstico no debería ser utilizado para evaluar la eficacia del medio o para confirmar la presencia de DTM cuando la probabilidad de la prueba preliminar del DTM es baja, por ejemplo, un 10%. En tal situación, el valor estimativo negativo de la Escala de la ATM del 98%, soporta la impresión clínica del dentista sobre la ausencia de DTM. Cuando la probabilidad de la prueba preliminar de los DTM es alta, o sea como un 90%, el valor estimative positivo de la Escala de la ATM del 97% soporta la impresíon clínica del dentista de la presencia de DTM. Los valores estimativos de las subescalas de la Escala de la ATM que miden la disfunción articular y el stress pueden ser utilizadas para refinar mas la impresión diagnóstica. Cuando el dentista no está seguro de la presencia de DTM y hace una estimación preliminar del 50%, el valor estimative positivo de la Escala de la ATM del 81% y el valor estimativo negativo del 83% substancialmente mejoran la exactitud de las decisiones clínicas.

Zusammenfassung

Der richtige Gebrauch von Voraussagewerten im klinischen Entscheidungsprozess und in der Evaluation von diagnostischen Tests für Myoarthropathien des Kausystems (MAP)

Die Literatur zu den MAP enthält grobe Missverstandnisse und Fehlanwendungen von statistischen Konzepten — insbesondere der Voraussagewerte — zum Auswerten von diagnostischen Modalitäten und zum klinischen Entscheidungsprozess. Der Gebrauch von Prävalenzdaten von MAP für ganze Populationen ist unbrauchbar zum Evaluieren des positiven Voraussagewertes für diagnostische Modalitäten. Der positive Voraussagewert eines diagnostischen Instrumentes sollte nicht gebraucht werden zur Prüfung der Tauglichkeit des Instrumentes oder zur Bestätigung der Präsenz von MAP, wenn die Wahrscheinlichkeit für MAP im Vortest tief, also z. B. 10% ist. In einer solchen Situation unterstützt der negative Voraussagewert der TMJ Scale von 98% den klinischen Eindruck des Zahnarztes, dass keine MAP vorhanden sei. Wenn die Wahrscheinlichkeit im Vortest für MAP hoch ist, etwa 90%, so unterstützt der positive Voraussagewert von 97% den klinischen Eindruck des Zahnarztes, dass eine MAP vorliegt. Die Voraussagewerte der Subskalen der TMJ Scale, die die Gelenkdysfunktion und den Stress beschreiben, können den klinischen Eindruck weiter verfeinern. Falls der Zahnarzt unsicher ist, ob MAP vohanden sind und eine Vortestschätzung von 50% macht, verbessert der positive Voraussagewert der TMJ Scale von 81% und der negative Voraussagewert von 83% die Genauigkeit der klinischen Entscheidung.