Evaluation of Diagnostic Tests Using Information Theory for Multi-Class Diagnostic Problems and its Application for the Detection of Occlusal Caries Lesions

Umut Arslan[1], Ergun Karaağaoğlu[1], Gökhan Özkan[2], Aydan Kanlı[3]

[1]Department of Biostatistics, Hacettepe University Faculty of Medicine, Ankara, Turkey
[2]Department of Oral and Maxillofacial Radiology, Adnan Menderes University Faculty of Dentistry, Aydın, Turkey
[3]Department of Oral and Maxillofacial Radiology, Hacettepe University Faculty of Dentistry, Ankara, Turkey

Background: Several methods are available to evaluate the performance of the tests when the purpose of the diagnostic test is to discriminate between two possible disease states. However multi-class diagnostic problems frequently appear in many areas of medical science. Hence, there is a need for methods which will enable us to characterize the accuracy of diagnostic tests when there are more than two possible disease states.

Aims: To show that two information theory measures, information content (IC) and proportional reduction in diagnostic uncertainty (PRDU), can be used for the evaluation of the performance of diagnostic tests for multi-class diagnostic problems that may appear in different areas of medical science.

Study Design: Diagnostic accuracy study.

Methods: Sixty freshly extracted permanent human molar and premolar teeth suspected to have occlusal caries lesions were selected for the study and were assessed by two experienced examiners. Each examiner performed two evaluations. Histological examination was used as the gold standard. The scores of the histological examination were defined as sound (n=11), enamel caries (n=22) and dentin caries (n=27). Diagnostic performance of i) visual inspection, ii) radiography, iii) laser fluorescence (LF) and iv) micro-computed tomography (M-CT) caries detection methods was evaluated by calculating IC and PRDU.

Results: Micro-computed tomography examination was the best method among the diagnostic techniques for the diagnosis of occlusal caries in terms of both IC and PRDU. M-CT examination supplied the maximum diagnostic information about the diagnosis of occlusal caries in the first (IC: 1.056; p<0.05), (PRDU: 70.5%) and second evaluation (IC: 1.105; p<0.05), (PRDU: 73.8%) for the first examiner. M-CT examination was the best method among the diagnostic techniques for the second examiner in both the first (IC:1.105; p<0.05), (PRDU:73.8%) and second evaluation (IC:1.061; p<0.05), (PRDU:70.8%). IC and PRDU were lowest for visual inspection.

Conclusion: The present study demonstrates that IC and PRDU can be used to evaluate diagnostic test performance when multiple disease states are being evaluated.

(*Balkan Med J* 2014;31:214-8).

Key Words: Diagnostic test, information content, occlusal caries, proportional reduction in diagnostic uncertainty

Diagnostic tests are widely used in medical science for diagnosing the presence or absence of a disease. These tests may not discriminate between subjects with and without the disease or disorder with perfect accuracy. Several methods are available to evaluate the performance of diagnostic tests when the purpose of the test is to discriminate between two possible disease states. Different performance indices have been proposed to determine the accuracy of diagnostic tests when both the disease state and the test result are binary. However, in some cases, the disease state may have more than two categories. For this reason, there is a need for other methods to characterize the accuracy of diagnostic tests. In this study, we propose a method which is useful in terms of selecting the most appropriate diagnostic test when the disease state and test results have three or more classes. In this study, information content (IC) and proportional reduction in diagnostic uncertainty (PRDU) were introduced as measures of diagnostic test accuracy. To demonstrate the applicability of these statistics, we evaluated the diagnostic performance of four different diagnostic methods which are used for dental caries detection: visual inspection, radiography, laser fluorescence (LF) and micro-computed tomography (M-CT) examination.

MATERIAL AND METHODS

Entropy is the main concept in information theory. It is an appropriate way to quantify the amount of uncertainty of a random variable (1, 2). The amount of uncertainty or entropy of a random variable, D, with probability distribution, $p(d_i)$, can be calculated as shown Equation (1). This function is called the Shannon entropy.

$$H(D) = -\sum_{i=1}^{n} p(d_i) \log_2 p(d_i) \qquad (1)$$

When the base of the logarithm in Equation (1) is two, the unit of measurement is bits (binary digits) (3, 4). Shannon entropy can be used to quantify diagnostic uncertainty. The uncertainty related to the disease before the diagnostic test is applied is referred to as "a priori uncertainty". The uncertainty related to the disease after the test result is obtained is referred to as "a posteriori uncertainty". The difference between the a priori uncertainty and the expected value of the a posteriori uncertainty gives the gain in information provided by the diagnostic test. It is referred to as the diagnostic information or IC of the diagnostic test (4-6). This difference is also called mutual information. It is also commonly expressed in bits (4).

Metz et al. (5) introduced the equation for IC for the binary case in terms of the sensitivity and specificity of the test. Benish (4) provided the expression for IC for the more general case as follows:

$$I(D; T) = \sum_{j=1}^{m} \sum_{i=1}^{n} p(t_j|d_i) p(d_i) \log_2 \frac{p(t_j|d_i)}{\sum_{i=1}^{n} p(t_j|d_i) p(d_i)} \qquad (2)$$

where $p(d_i)$ is the prior probability of the disease state i and $p(t_j|d_i)$ is the probability of test result j given disease state i.

We also consider the percent reduction in diagnostic uncertainty. The reduction in uncertainty or gain in information is described as the explained percentage of information or PRDU supplied by the diagnostic test. Entropy known as uncertainty can also be defined as the qualitative variance. For this reason, the PRDU can be interpreted as the total variance and residual variance in the diagnostic state. These indices measure the degree of relationship between the disease state and diagnostic test and can be interpreted in the same way as the coefficient of determination obtained in Pearson's correlation coefficient (7, 8). PRDU is a measure of the bivariate association between disease state and diagnostic test (7). PRDU can be defined as shown in Equation (3).

$$PRUD = \frac{[H(D) - H(D|T)]}{H(D)} \quad \text{or} \quad PRUD = \frac{I(D;T)}{H(D)} \qquad (3)$$

where H (D) is a priori uncertainty of the disease state and H(D|T) is the expected value of the a posteriori uncertainty of the disease state. Interpretation of diagnostic test performance in bits can be difficult in clinical practice. PRDU, on the other hand, can be simply interpreted as the percentage of the a priori uncertainty that is eliminated, on average, by diagnostic testing. The independence of the disease state (D) and the test result (T) can be statistically tested. 2n (IC) has an asymptotically chi square distribution with degrees of freedom equal to (d-1) (t-1), where n is the sample size, d is the number of possible disease states, and t is the number of possible test results (8). When the null hypothesis is rejected at a specified level, there is a significant association between the disease state and diagnostic test. The test provides a significant amount of information about the disease state. Afterwards, the performance of the diagnostic test can be interpreted using PRDU.

Study procedures

Sixty freshly extracted permanent human molar and premolar teeth suspected to have occlusal caries lesions were selected for the study and were assessed by two experienced examiners from the Department of Dento-Maxillofacial Radiology. The teeth were cleaned using a rotary brush and hand instruments and then stored in formalin solution. Occlusal surfaces were photographed and one spot, the most suspected area, was selected as the test site for every single tooth. Occlusal caries were detected by two observers (AK, GÖ) using four diagnostic methods. All measurements were performed by the two examiners and then repeated a week later. The Kappa coefficient was used to assess the intra-rater agreement. The four methods used to diagnose dental carries were visual inspection, radiography, LF and M-CT examination. In this study, we evaluated the diagnostic performance of four different diagnostic methods for occlusal caries using IC and PRDU. Histological examination was used as the gold standard. The scores of histological examination were sound, enamel caries and dentin caries. The final decision about the histological examination score was a consensus between the two observers.

Visual inspection

The sites were examined with a light reflector and without magnification. The teeth were first examined wet and then after 5 seconds of drying with compressed air. The examiners did not use a mouth mirror or sound instruments.

Radiographic method

The x-ray device (Belmont Phot-X II, Osaka, Japan) was set to 70 kVp, 7 mA and 0.13 s exposure time. E speed radiographic films (Eastman Kodak, Rochester, NY) and a bitewing film holder were used under standard conditions. The focus-to-film distance was 21 cm. After the exposure process,

the films were developed automatically and viewed independently using a light box without magnification.

Laser fluorescence

The selected occlusal sites were assessed using an LF device (DIAGNOdent Pen) (Kavo, Biberach, Germany) with the appropriate cylindrical tip according to the manufacturers' instructions. Two examiners carried out the assessments independently after 5 seconds of air drying. The measurements were repeated three times by each examiner to prevent individual mistakes. The highest value was recorded for each selected site.

Micro-computed tomography examination

First, tooth crowns were separated from their roots to reduce ring artifacts. Images of the tooth crowns were obtained using a SkyScan 1774 unit (SkyScan, Kontich, Belgium). The device was set to 50 kV, 800 µA, a pixel size of 15 µm at 1024x1024 resolution and 1° rotation step. The average exposure time was 73 minutes. NRecon software (SkyScan, Kontich, Belgium) was used to obtain coronal and sagittal images.

Histological examination

Histological examination was used as the gold standard for the validation of lesion depth. Tooth crowns were embedded in methyl methacrylate and two to four sections were obtained for each tooth from the suspected area with a diamond saw placed on the microtome device (Isomet 1000, Buehler Ltd, Lake Bluff, IL, USA) in the buccolingual direction. After the grinding and polishing process, crown slices with a thickness of 80-100 µm were examined under a light microscope (Leica Dental Corp, Wetzlar, Germany). When a disagreement occurred, consensus was reached by repeating the examination.

The three possible scores for visual inspection, radiography, LF, M-CT examination and histological examination were i) sound, ii) enamel caries and iii) dentin caries (9-13).

RESULTS

The dataset was derived from 60 freshly extracted permanent human molar and premolar teeth. The findings from the histologic examination were 11 sound teeth, 22 teeth with enamel caries, and 27 teeth with dentin caries. Prior diagnostic uncertainty of occlusal caries was obtained as 1.498 bits. In our study, all diagnostic methods were applied to extracted permanent human molar and premolar teeth and there were no missing data.

The IC and PRDU for each diagnostic method for the first examiner are presented in Table 1. IC gives a reduction in diagnostic uncertainty in units of bits. A larger value of IC or PRDU indicates a greater reduction in diagnostic uncertainty resulting from testing. The M-CT examination was the best

TABLE 1. The information content and PRDU (%) of each diagnostic method for the first examiner

Diagnostic methods	First examiner			
	First evaluation		Second evaluation	
	Information content (in bits)	PRDU (%)	Information content (in bits)	PRDU (%)
Visual inspection	0.582*	38.9	0.539*	36.0
Radiographic method	0.856*	57.2	0.839*	56.0
LF device	0.755*	50.4	0.817*	54.6
M-CT examination	1.056*	70.5	1.105*	73.8

*p<0.005

PRDU: proportional reduction in diagnostic uncertainty; LF: laser fluorescence; M-CT: micro-computed tomography

TABLE 2. The information content and PRDU (%) of each diagnostic method for the second examiner

Diagnostic methods	Second examiner			
	First evaluation		Second evaluation	
	Information content (in bits)	PRDU (%)	Information content (in bits)	PRDU (%)
Visual Inspection	0.589*	39.3	0.544*	36.3
Radiographic method	0.730*	48.8	0.788*	52.6
LF device	0.844*	56.3	0.793*	53.0
M-CT examination	1.105*	73.8	1.061*	70.8

*p<0.005

PRDU: proportional reduction in diagnostic uncertainty; LF: laser fluorescence; M-CT: micro-computed tomography

method among the diagnostic techniques for the diagnosis of occlusal caries in terms of IC or PRDU. The M-CT examination supplied the most diagnostic information and yielded the greatest reduction in diagnostic uncertainty about occlusal caries for the first examiner for both the first evaluation (IC:1.056, p<0.05 and PRDU:70.5) and the second evaluation (IC:1.105, p<0.05 and PRDU 73.8). Table 2 shows that M-CT examination was the best method among the diagnostic techniques for the second examiner in both the first (IC:1.105; p<0.05 and PRDU:73.8%) and second evaluation (IC:1.061; p<0.05 and PRDU:70.8%). As shown in Table 3, M-CT was also associated with greater measures of intra-rater agreement (Kappa coefficients) than measures for the other tests for the first examiner.

Visual inspection had the minimum discriminatory power regarding the diagnosis of occlusal caries in the first (IC:0.582; p<0.05, PRDU:38.9%) and second evaluation (IC:0.539; p<0.05, PRDU:36.0%) for the first examiner, as shown in Table 1. Table 2 shows that visual inspection had similar results for the diagnosis of occlusal carries for the second examiner in the first (IC:0.589; p<0.05, PRDU:39.3%) and second evaluation (IC:0.544 p<0.05, PRDU:36.3%). As

Arslan et al. Multi-Class Diagnostic Problems and Diagnosis of Disease

217

TABLE 3. Kappa coefficient values of the first and second examiner for intra-rater agreement

Diagnostic methods	Kappa coefficient	
	First examiner	Second examiner
Visual inspection	0.726*	0.800*
Radiographic method	0.922*	0.949*
LF device	0.814*	0.809*
M-CT examination	0.975*	0.898*

*$p < 0.005$

LF: laser fluorescence; M-CT: micro-computed tomography

shown in Table 3, the visual inspection method was also associated with the lowest measures of intra-rater agreement. The radiographic method was second in terms of diagnostic information in the first (IC:0.856; $p < 0.05$, PRDU:57.2%) and second evaluation (IC:0.839; $p < 0.05$, PRDU:56.0%) for the first examiner. On the other hand, the LF device took second place in terms of diagnostic information in the first (IC:0.844; $p < 0.05$, PRDU:56.3%) and second evaluation (IC:0.793; $p < 0.05$, PRDU:53.0%) for the second examiner.

DISCUSSION

Sensitivity and specificity are commonly used to characterize diagnostic test performance. These statistics apply to the case where the disease state and test results are binary. However, most of the diagnostic tests used in medical science are not binary, and few of them have positive and negative results (14, 15). Hence, there is a need for new methods to evaluate diagnostic test performance. Studies of diagnostic test performance using information theory began with the work of Metz et al. (5) in 1973. They provided a formula for calculating IC in terms of disease prevalence, sensitivity and specificity. Disease state is not always binary, and most of those who have and do not have a certain disease are divided into sub groups within themselves. Mossman (16) discussed how the performance of tests would be evaluated when the disease condition has more than two classes. Benish (4) introduced a formula for the general case (multiple possible disease states and multiple possible test results).

In addition to providing a demonstration of the clinical applicability of an information theory approach to quantifying diagnostic test performance, one of the contributions of our study is the demonstration that IC can be evaluated for its statistical significance. Previous studies have not considered whether or not the associations identified by IC might have occurred by chance. Another contribution of this study is demonstrating how to evaluate the performance of a test by using PRDU. Researchers have to evaluate clinical significance together with statistical significance. It is at the researcher's discretion as to how much information gain or percentage of information constitutes clinical significance.

In this study, we have focused on the role of PRDU in the interpretation of the performance of the diagnostic test. Our analysis interpreted the performance of visual inspection, radiographic examination, LF and M-CT examination methods for the diagnosis of occlusal caries using IC and PRDU. PRDU in percent can be used for the performance of diagnostic tests instead of IC in bits due to the easy interpretation of PRDU.

Visual and radiographic examination methods are the most commonly used methods for caries detection in dentistry. These diagnostic methods have some disadvantages for the detection of caries lesions. Visual examination is a subjective method that changes from physician to physician, particularly in caries lesions (17, 18). In the present study, the PRDU of visual examination for both examiners was lower than for the other methods. This method had minimum discriminatory power regarding the diagnosis of occlusal caries in the first and second evaluation for both examiners. It also had the minimum Kappa coefficient for intra-rater agreement between the two evaluations for both examiners. The ICDAS II visual examination method is an improved method for carries detection (9, 10, 19, 20). However, the ICDAS II visual examination method showed lower performance than other studies, according to our results. According to some of the results, the radiographic method was found to have worse performance than the visual examination and LF device methods when compared using sensitivity, specificity, accuracy, and AUC values (20-23). This could be due to the depth of the caries lesion on the radiograph, difficulty in noticing hidden caries because of the fissure system and not removing all of the colorization on the occlusal region (21, 24, 25). Therefore, new methods are being researched for early caries lesions. The LF device is one of the latest methods (12).

The radiographic method had a greater intra-rater agreement than the LF device. Although the Kappa value of the radiographic method was high and there was high concordance between the first and second readings, the PDRU values for this method were not the highest. This may have been caused by the fact that detecting a caries lesion on a radiograph requires at least 30% demineralization. Another cause may have been the experience level of the physician. In this study, the radiographic method and the LF device showed similar performance for the diagnosis of caries in terms of the reduction in diagnostic uncertainty. Kavvadia et al. (26) reported that the LF device did not show better performance than the other methods after the evaluation of sensitivity, specificity, accuracy, and AUC values. The LF device had provided results than the radiographic method according to the assessment of sensitivity, specificity, accuracy, AUC, and likelihood ratio values. (12, 21, 22, 27). This could be caused by different cut-off values of the device, different environmental conditions,

218

Arslan et al. Multi-Class Diagnostic Problems and Diagnosis of Disease

or various other reasons. The radiographic method and the LF device supplied more diagnostic information than the visual method. This indicates that visual examination should be combined with radiographic or laser fluorescence methods. In a study by Kavvadia et al. (26), when an LF device was combined with visual examination, this combination improved the performance of caries detection.

Micro-computed tomography showed the best performance for both examiners for both evaluations, according to PRDU. Detecting caries lesions with an M-CT device was the most valuable system among all the methods. Although the M-CT device showed the best performance, its clinical use is impossible because of the high radiation dose while scanning the object. It might have a role as an alternative or adjunct to the histological method. There is disagreement about using M-CT as the gold standard; different claims are available in the literature, some using sensitivity, specificity, accuracy, and AUC, as well as negative and positive predictive values (28-30). This method had high intra-rater agreement at the same time. The high intra-rater agreement for this method suggests that the high resolution of the M-CT allowed to the physician to obtain a more accurate reading of the reconstructed images.

REFERENCES

1. Thomas JA. Elements of information theory. 2nd ed. New York: John Wiley & Sons; 2006.
2. Asch DA, Patton PJ, Hershey JC. Knowing for the sake of knowing: The value of prognostic information. *Med Decis Making* 1990;10:47-57. **[CrossRef]**
3. Klir GJ. Uncertainty and information. New York: John Wiley & Sons; 2006.
4. Benish WA. Mutual information as an index of diagnostic test performance. *Methods Inf Med* 2003;42:260-4.
5. Metz CE, Goodenough DJ, Rossmann K. Evaluation of receiver operating characteristic curve data in terms of information theory with applications in radiography. *Radiology* 1973;109:297-303. **[CrossRef]**
6. Mossman D, Somoza E. Diagnostic tests and information theory. *J Neuropsychiatry Clin Neurosci* 1992;4:95-8.
7. Kim JO. PRU Measures of association for contingency table analysis. *Social Methods Res* 1984;13:3-44. **[CrossRef]**
8. Golden LL, Brockett PL, Zimmer MR. An information theoretic approach for identifying shared information and asymmetric relationships among variables. *Multivariate Behav Res* 1990;25:479-502. **[CrossRef]**
9. Ismail AL, Sohn W, Tellez M, Amaya A, Sen A, Hasson H, et al. The International Caries Detection and Assessment System (ICDAS): An integrated system for measuring dental caries. *Dent Oral Epidemiol* 2007;35:170-8. **[CrossRef]**
10. Extrand KR, Martignon S, Ricketts DJ, Qvist V. Detection and activity assessment of primary coronal caries lesions: a methodologic study. *Oper Dent* 2007;32:225-35. **[CrossRef]**
11. Extrand KR, Ricketts DJ, Kidd EA. Reproducibility and accuracy of three methods for assessment of demineralization depth of the occlusal surface: an in vitro examination. *Caries Res*1997;32:225-235.
12. Lussi A, Hellwig E. Performance of a new laser fluorescence device for the detection of occlusal caries in vitro. *J Dent* 2006;34:467-71. **[CrossRef]**
13. Downer MC. Concurrent validity of an epidemiological diagnostic system for caries with the histological appearance of extracted teeth as validating criterion. *Caries Res* 1975;9:231-46. **[CrossRef]**
14. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continious measurements. *Stat Med* 2004;23:3437-49. **[CrossRef]**
15. Somoza E, Mossman D. Neuropsychiatric decision making: Designing non-binary diagnostic tests. J Neuropsychiatry Clin Neurosci 1991;3;197-200.
16. Mossman D. Three-way ROCs. *Med Decis Making* 1999;19:78-89. **[CrossRef]**
17. Ekstrand K, Qvist V, Thylstrup A. Light microscope study of the effect of probing in occlusal surfaces. *Caries Res* 1987;21:368-374. **[CrossRef]**
18. Sawle RF, Andlaw RJ. Has occlusal caries become more difficult to diagnose? *Br Dent J* 1988;164:209-11. **[CrossRef]**
19. Rodrigues JA, Hug I, Diniz MB, Lussi A. Performance of fluorescence methods, radiographic examination and ICDAS II on occlusal surfaces in vitro. *Caries Res* 2008;42:297-304. **[CrossRef]**
20. Ekstrand KR, Luna LE, Promisiero L, Cortes A, Cuevas S, Reyes JF, et al. The reliability and accuracy of two methods for proximal caries detection and depth on directly visible proximal surfaces: an in vitro study. *Caries Res* 2011;45:93-9. **[CrossRef]**
21. Neuhaus KW, Rodrigues JA, Hug I, Stich H, Lussi A. Performance of laser fluorescence devices, visual and radiographic examination for the detection of occlusal caries in primary molars. *Clin Oral Investig* 2011;15:635-41. **[CrossRef]**
22. Sridhar N, Tandon, Rao NA. Comparative evaluation of DIAGNOdent with visual and radiography for detection of occlusal caries: an in vitro study. *Indian J Dent Res* 2009;20:326-31. **[CrossRef]**
23. Rodrigues JA, Hug I, Diniz MB, Lussi A. Performance of fluorescence methods, radiographic examination and ICDAS II on occlusal surfaces in vitro. *Caries Res* 2008;42:297-304. **[CrossRef]**
24. White SC, Pharoah MJ. Oral Radiology Principles and interpretation. 6th ed. New Delhi: St. Louis Mosby; 2009.
25. Bader JD, Shugars DA, Bonito AJ. Systematic reviews of selected dental caries diagnosis and management methods. *J Dent Educ* 2001;65:960-8.
26. Kavvadia K, Lagouvardos P, Apostolopoulo D. Combined validity of DIAGNOdent and visual examination for in vitro detection of occlusal caries in primary molars. *Lasers Med Sci* 2012;27:313-9. **[CrossRef]**
27. Chu CH, Lo EC, You DS. Clinical diagnosis of fissure caries with conventional and laser-induced fluorescence techniques. *Lasers Med Sci* 2010;25:355-62. **[CrossRef]**
28. Kamburoğlu K, Kurt H, Kolsuz E, Öztaş B, Tatar İ, Çelik HH. Occlusal caries depth measurements obtained by five different imaging modalities. *J Digit Imaging* 2011;24:804-13. **[CrossRef]**
29. Mitropoulos P, Rahiotis C, Stamatakis H, Kakaboura A. Diagnostic performance of the visual caries classification system ICDAS II versus radiography and micro-computed tomography for proximal caries detection: an in vitro study. *J Dent* 2010;38:859-67. **[CrossRef]**
30. Soviero VM, Leal SC, Silva RC, Azevedo RB. Validity of MicroCT for in vitro detection of proximal carious lesions in primary molars. *J Dent* 2011;40:35-40. **[CrossRef]**