Community Dental Health (2011) **28,** 238–242
Received 2 November 2009; Accepted 1 April 2010

doi:10.1922/CDH_ 2612Nelson05

Dental examiners consistency in applying the ICDAS criteria for a caries prevention community trial

S. Nelson[1], H. Eggertsson[2], B. Powell[3], J. Mandelaris[3], M. Ntragatakis[3], T. Richardson[1] and G. Ferretti[3]

[1]Case Western Reserve University, Department of Community Dentistry, Cleveland, Ohio, USA. [2]Indiana University School of Dentistry, Department of Preventive and Community Dentistry, Indiana, USA. [3]Case Western Reserve University, Department of Pediatric Dentistry, Cleveland, Ohio, USA.

Aim: To examine dental examiners' one-year consistency in utilizing the International Caries Detection and Assessment System (ICDAS) criteria after baseline training and calibration. ***Methods:*** A total of three examiners received baseline training/calibration by a "gold standard" examiner, and one year later re-calibration was conducted. For the baseline training/calibration, subjects aged 8-16 years, and for the re-calibration subjects aged five to six years were recruited for the study. The ICDAS criteria were used to classify visual caries lesion severity (0-6 scale), lesion activity (active/ inactive), and presence of filling material (0-9 scale) of all available tooth surfaces of permanent and primary teeth. The examination used a clinical light, mirror and air syringe. Kappa (weighted: Wkappa, unweighted: Kappa) statistics were used to determine inter-and intra-examiner reliability at baseline and re-calibration. ***Results:*** For lesion severity and filling criteria, the baseline calibration on 35 subjects indicated an inter-rater Wkappa ranging from 0.69-0.92 and intra-rater Wkappa ranging from 0.81-0.92. Re-calibration on 22 subjects indicated an inter-rater Wkappa of 0.77-0.98 and intra-rater Wkappa ranged from 0.93-1.00. The Wkappa for filling was consistently in the excellent range, while lesion severity was in the good to excellent range. Activity kappa was in the poor to good range. All examiners improved with time. ***Conclusions:*** The baseline training/calibration in ICDAS was crucial to maintain the stability of the examiners reliability over a one year period. The ICDAS can be an effective assessment tool for community-based clinical trials.

Key words*: Caries severity, ICDAS, dental examiners, baseline training, dental caries, community based study*

Introduction

Currently there exists no single standard caries detection and/or assessment system universally agreed upon among scientists (Ismail, 2004). Several caries detection systems are in place which attempt to describe and diagnose the caries process (Bader, 2001; Ismail, 2004). But, the evidence points to poor performance of these methods due to high false negative diagnosis in the presence of carious lesions and low to moderate false positive diagnosis in the absence of carious lesions (Bader, 2001). In 2004, a total of 29 visual and visual-tactile caries detection methods across the globe dating from the 1950s to 2000 were investigated and no single method met content validity standards (Ismail, 2004). This lack of consistency in detection, reporting, description, and interpretation of dental caries in these assessment systems limits the comparability of outcomes in clinical studies (Ismail *et al*., 2007).

ICDAS is a newly developed system that addresses the incompatibility of prior systems by utilizing their best elements and showing demonstrated capability in standardizing caries detection (Topping and Pitts, 2009). This system consolidates features of several caries classification systems into one universal system using a six-point ordinal scale ranging from non-cavitated to extensive cavitated lesions to describe the carious process and a two point scale to describe lesion activity (ICDASCC, 2007; Pitts, 2004). This quantitative description of the disease makes ICDAS ideal and unique for communicating the diagnosis of dental caries thus providing concise and descriptive flexibility to researchers and clinicians (ICDASCC, 2007).

Although ICDAS is not the only caries detection and assessment system available, it has been tested against other caries detection systems such as the World Health Organization and laser fluorescence and has shown to be effective in the detection and assessment of dental caries (Braga *et al.,* 2009; Kuhnisch *et al.,* 2008). During in-vitro evaluation of extracted teeth, the ICDAS lesion codes were highly correlated with the histological examination (Ekstrand *et al.,* 2007; Ismail *et al.,* 2007; Jablonski-Momeni *et al.,* 2008; Ricketts *et al,* 2002). Furthermore, the ICDAS system has demonstrated good inter- and intra- examiner reliability and validity (content and discriminant) in clinical studies (Ismail *et al.,* 2007).

ICDAS criteria for lesion severity and filling have been used in community-based clinical studies (Cook *et al.,* 2008; Ismail *et al.,* 2008; Varma *et al.,* 2008) to report the decayed, missing, filled surfaces/teeth (DMFS/T), but none have reported on reliability of the lesion activity criteria of ICDAS which is important for assessing treatment effectiveness in clinical trials. Further, to our knowledge, only one study has reported on the

Correspondence to: Professor Suchitra Nelson,School of Dental Medicine, Case Western Reserve University, Cleveland, Ohio 44106-4905, USA. E-mail: suchitra.nelson@case.edu

examiners training/calibration on long-term consistency in using ICDAS (Ismail *et al.,* 2007) which is important for longitudinal investigations. Thus, this study examined dental examiners inter- and intra-rater reliability after one year in diagnosing lesion severity, activity, and filling criteria of the ICDAS following baseline training and calibration.

Method

A "gold standard" examiner (A) trained and calibrated three dental examiners (B, C, D). The three trainee examiners had all graduated from dental school within the last three years. Baseline training consisted of a four hour didactic session, followed by examination of 80 extracted teeth, and an additional two hour session of diagnosis from photographs of extracted teeth. Clinical training involved thorough review of clinical procedures and communication with recorders, with subsequent in-vivo examination of subjects. During the clinical training, the trainee examiners were given time to identify differences in scoring, and to converse with the trainer. Overall, the training took two days. This was followed by two days of calibration exams with in-vivo subjects who were independently examined by the "gold standard" and the trainee examiners. At the one year examination, all three trainee examiners completed re-calibration with examiner A, preceded by a one hour review of the criteria.

During baseline training/calibration subjects aged 8–16 years were recruited from the Pediatric Dental and Endodontic clinics of Case Western Reserve University (CWRU) School of Dental Medicine. Institutional Review Board (IRB) approval was obtained from University Hospitals CWRU Medical Center. Informed consent and assent were obtained from the parent and child respectively. The re-calibration with more randomly selected subjects were part of a larger school-based clinical trial. All re-calibration subjects also had IRB approval as part of the parent study.

The time was also recorded during re-calibration to determine approximate length of examination, assessed from when the first tooth was called to the recorder until the recording of the last tooth.

All in-vivo examinations during training and baseline calibration were completed at the CWRU School of Dental Medicine in regular dental operatories. Examinations performed during re-calibration were accomplished with portable dental units with subjects seated in the dental chair in a reclined position.. Exams used a clinical light, mirror, and air syringe, but no explorer. All teeth, regardless of their condition, were included in the baseline training/calibration and re-calibration. Each surface of the erupted tooth was examined, scored, and recorded: occlusal, buccal, lingual, mesial, and distal. Lingual grooves of upper molars and buccal pits of lower molars were scored as separate entities. A comprehensive visual intra- oral exam was completed of all hard and soft tissues. Any abnormal findings were recorded on the data sheet and disclosed to the subject's carer. Referrals for further care were given as necessary.

The examiner used toothbrush and water to remove plaque from occlusal and smooth surfaces, and floss to clean proximal surfaces of all teeth. Teeth were then wiped clean with a 2x2 gauze as needed. Cotton rolls were used for isolation. A standardized visual examination was completed of all study teeth by numbering the teeth in a clockwise fashion beginning with upper right most posterior tooth across the anterior teeth and over to the upper left most posterior tooth. The examination was continued through the lower teeth beginning on the lower left most posterior tooth across the lower anterior teeth and completing the examination on the lower right most posterior tooth. Visual examination was commenced on a moist tooth and then continued as the tooth was dried using air-spray over approximately five seconds.

The ICDAS criteria were used in this study. A separate score was used for lesion severity, lesion activity, and for the presence and type of filling materials, as presented in Table 1. Each surface of the study tooth received a score which was called out to a trained study recorder and recorded on the data sheet.

Inter-examiner repeatability with the "gold standard" (A) and intra-examiner repeatability were evaluated using Wkappa statistics for lesion severity and fillings at baseline and re-calibration. Wkappa accounts for the ordered nature of the data by allowing 'closer' disagreements to show better agreement than disagreements which were farther apart. The Wkappa was calculated using

Table 1. Lesion and filling codes for ICDAS II

Lesion Codes	
0	Sound; no caries change after air drying (five seconds); or non-carious change such as stain, hypoplasia, wear, erosion and other non-caries phenomena.
1	First visual change in enamel, seen after air drying, or coloured change limited to the confines of the pit and fissure area.
2	Distinct visual change in enamel seen when wet, white or coloured, wider than the fissure/fossa.
3	Localized enamel breakdown, with no visible dentin, widening of fissure.
4	Underlying dark shadow from dentin with or without localized enamel breakdown.
5	Distinct cavity with dentin exposed at the base of the cavity.
6	Extensive cavity with dentin visible at base and walls of the cavity (or ½ surface).
Filling Codes	
1	Sealant, partial
2	Sealant, complete
3	Tooth coloured
4	Amalgam
5	Stainless steel crown
6	Crown, gold or porcelain
7	Filling gone
8	Temporary
9	Other
Lesion Activity Codes	
1	Not active
2	Active

Table 2. Inter-rater reliability Kappa

	Baseline Calibration Examiners			*Re-Calibration Examiners*		
	B	*C*	*D*	*B*	*C*	*D*
Lesion*	0.69 (5040)	0.81 (5040)	0.75 (5040)	0.77 (1560)	0.88 (2252)	0.83 (2270)
Activity†	0.41 (150)	0.45 (147)	0.20 (162)	0.48 (93)	0.08 (124)	0.36 (113)
Filling*	0.87 (5040)	0.92 (5040)	0.82 (5040)	0.77 (1560)	0.98 (2252)	0.96 (2270)

() Number of surfaces used for kappa calculation
* squared distance (Fleiss-Cohen weights) weighted kappas
** As compared to "Gold Standard" A
† For activity unweighted kappas were calculated with surfaces which had a lesion severity code.

Table 3. Intra-rater reliability Kappa

	Baseline Calibration Examiners				*Re-Calibration Examiners*			
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*
Lesion*	0.88 (2100)	0.91 (2100)	0.81 (2100)	0.83 (2100)	0.99 (998)	0.93 (474)	0.97 (990)	0.95 (994)
Activity†	0.79 (181)	0.56 (151)	0.11 (130)	0.44 (135)	0.96 (90)	** (66)	0.94 (85)	0.58 (78)
Filling*	0.90 (2100)	0.92 (2100)	0.81 (2100)	0.82 2100)	1.00 (998)	1.00 (474)	1.00 (990)	1.00 (994)

() Number of surfaces used for kappa calculation
* squared distance (Fleiss-Cohen weights) weighted Kappas
† For activity unweighted kappas were calculated with surfaces which had a lesion severity code.
** Kappa undefined as all 66 judgements were the same code, 2.

the Fleiss-Cohen (1973) squared distance method. The Fleiss-Cohen weights were calculated as

$$1 - \left[\frac{(\text{row score} - \text{column score})^2}{(\text{highest score} - \text{lowest score})^2}\right].$$

For lesion severity and fillings, all tooth surfaces including sound were used for the WKappa calculation. The rationale for using all filling codes was that much of the disagreements between the gold standard and examiners were in codes 1-3 (partial, full sealants, tooth coloured, while there was 100% agreement in codes 7 and 8 (lost and temporary restoration). For lesion activity, unweighted kappas were calculated using only tooth surfaces with a lesion (severity code 1-6).

Results

The baseline training was completed on 10 subjects, and baseline calibration was completed on 35 subjects aged 8–16 years on whom 14 examinations were repeated. Re-calibration was accomplished by evaluating 22 subjects (aged 5-6 years) of which 10 received repeat examinations. Repeat examinations were done at least two hours after the initial examination. Both the baseline calibration and re-calibration groups of children were predominantly African-American and of low socio-economic status. The prevalence of decay (ICDAS lesion code ≥ 3) or fillings (ICDAS filling code ≥3) in one or more surfaces was 72% in the baseline calibration group, and 70% in the re-calibration group as determined by the "gold standard" examiner. Thus, the extent of disease burden was similar in the two groups of calibration subjects.

Each examiner scored 7,140 surfaces for the baseline calibration, and 2,034 to 3,264 surfaces for the re-calibration. Tables 2 and 3 indicate that for baseline calibration, inter-rater Wkappa for lesion severity ranged from 0.69-0.81 and intra-rater Wkappa ranged from 0.81-0.91, indicating good to excellent reliability. After one year, all examiners Wkappa scores improved with the inter-rater Wkappa for lesion severity ranging from 0.77-0.88 and intra-rater Wkappa ranging from 0.93-0.97, indicating excellent reliability. For baseline calibration, the reliability for activity was in the poor to good range (inter-rater: 0.20-0.45; intra-rater: 0.11-0.56) and most examiners showed improvement in kappa scores after one year. For baseline calibration, inter- and intra-rater reliability for fillings were in the excellent range (>0.81), and consistently maintained after one year. Over time examiner (B) had a slight decrease in inter-rater Wkappa score for filling but still in the excellent range. Examiner C had a decrease in inter-rater kappa score for activity after one year that resulted in the poor range.

The mean time for an ICDAS examination in a school-based setting based on time recorded for 10 subjects during re-calibration, was 1.14 (sd 0.10) minutes.

Discussion

In clinical trials assessing treatment effectiveness, there exists a need to train and calibrate dental examiners in a standardized assessment protocol which would yield

consistent and reliable caries outcome data over time. Currently, the ICDAS criteria is being used for an ongoing parent multi-site clinical trial assessing the effectiveness of xylitol gummy bears versus placebo gummy bears where kindergarten children are being followed until second grade in five urban elementary schools. As part of this community-based trial multiple examiners collect longitudinal caries outcome data after a baseline training/calibration. Thus, examiners long-term stability in maintaining consistency of caries measurements is crucial.

The present findings indicate dental examiners had good to excellent inter-rater reliability (0.69-0.81) for the baseline calibration, and after one year weighted kappa scores improved and were consistently in the excellent range (0.77-0.88) for diagnosing lesion severity. These results are very similar to a prior community based observational study (Ismail *et al.,* 2007) which reported good to excellent inter-rater reliability for their main examiners at baseline (0.63-0.75) and one to two years later (0.68-0.76). The intra-rater reliability was higher than inter-rater and consistently excellent for baseline (0.81-0.91), and at one year (0.93-0.97), a finding similar to prior studies (Cook *et al.,* 2008; Ismail *et al.,* 2007; Varma *et al.,* 2008).

To our knowledge, our study would be the first to report on the reliability of lesion activity for ICDAS in a community based setting. The dental examiners were in the poor to good range for consistency in recognizing lesion activity with the "gold standard," which was slightly better for two of the examiners at the one year re-calibration. However, the intra-rater reliability was mostly in the good to excellent range. The present study also found the dental examiners had excellent inter- and intra-rater reliability with fillings that was maintained after one year which was also similar to a previous study (Ismail *et al.,* 2007). This indicates dental examiners with no prior experience in epidemiological dental exams/research can be reliably trained and calibrated in using the ICDAS criteria and maintain their reliability over a one year period, a very important goal for any observational longitudinal or clinical trial study.

During re-calibration, a subset of ICDAS exams was timed to test the feasibility of dental examiners using this protocol in a community setting to detect and assess dental caries. The complete examination of the dentition took a little over one minute for children aged five to six years old. Logistically, this finding is a good indicator for the use of ICDAS in community based studies where the exam can be performed quickly, easily, and with clearly defined criteria for visual caries detection. A prior study (Braga *et al.,* 2009) reported the ICDAS criteria took twice as long as the World Health Organization criteria without any actual time reported, and thus it is difficult to compare our findings with this study.

The experience gained from using the ICDAS criteria indicated the following:

1) the logistical details of coordinating such calibration activities are detailed and time-consuming, but extremely essential for in-depth and detailed calibration of examiners for large scale epidemiological research. The baseline training/calibration took four complete days with didactic, in-vitro, and in-vivo exams on a separate group of subjects who were recruited solely for the calibration study. The calibration exercise was a study in itself. So, researchers wanting to use the ICDAS should be aware that training of dental examiners has to be done prior to the start of data collection and appropriate time and funds should be allocated for such activities;
2) to maintain the consistency of longitudinal caries measurements, re-calibration of examiners is essential. In our study, pediatric dental residents and pediatric dental faculty are being used as dental examiners. One method of maintaining reliability is to consistently use ICDAS criteria in the clinic setting or when seeing patients so using this protocol becomes a habit. Although our dental examiners did not examine research subjects during the interim one year period, they were increasingly using the ICDAS criteria in the clinic while seeing patients thus enhancing their overall consistency in using ICDAS over time;
3) it is also essential for dental examiners in clinical studies to understand ICDAS criteria can be effectively used in the clinic setting. This standardized clinical visual scoring system of recording enamel and dentinal caries on tooth/surface level in a reliable manner has demonstrated usefulness to form integrated and personalized treatment plans involving both early preventive and treatment options that can be easily monitored over time (Pitts and Richards, 2009; Topping and Pitts, 2009).

There are some limitations to our study. First, we used different age groups for the baseline and re-calibration exercise: baseline calibration with 8-16 year olds who had predominantly permanent dentition; re-calibration with 5-6 year olds with predominantly primary dentition and same age as the larger parent clinical trial. However, the disease burden was similar in the two groups of subjects. We deliberately excluded children younger than 8 years from the baseline training/ calibration due to the inability of young children to sit in a dental chair for multiple exams by four dental examiners lasting nearly 45-60 minutes. But, as a result of the baseline calibration our examiners were well trained in diagnosing lesion severity in both primary and permanent dentition. This may have actually enhanced their ability to have excellent reliability during re-calibration perhaps due to clear-cut signs of lesions in the primary dentition. Additionally, it is possible the better results during re-calibration may be due to fewer diagnostic challenges in the primary dentition of 5-6 year olds compared to the more challenging differential diagnosis that is often needed in 8-16 year olds. Second, the repeat exams were conducted in a two hour time frame that could have enhanced intra-examiner reliability. It is possible that examiners were able to recall their previous judgement on the subject better than if there was a longer time lapse between exams. Since, the subjects were recruited either solely for the baseline calibration or seen in a school setting, logistical reasons such as funds to pay for later visits, minimizing time out of the classroom precluded a longer time span which is essential for future researchers to consider. Third, the weighted kappa scores can be calculated by different methods, with the commonly used ones being Fleiss-Cohen (squared distance) and Cicchetti-

Allison (linear). In our study we used the Fleiss-Cohen weighting method whose calculated value is consistently higher than the Cicchetti Allison weighting method (Ismail *et al.,* 2007). But this should not have affected the conclusion of our findings since a prior study (Ismail *et al.,* 2007) found regardless of the method the conclusions remained similar. Fourth, the present study found the dental examiners were less consistent in recognizing lesion activity compared to lesion severity and fillings. Lesion activity is not well-conceptualized for clinicians as part of their dental training and should be emphasized in the future. Most clinical studies use the DMFS/ T without much regard to the activity of the lesions, but this is especially important to consider in clinical trials if an active lesion becomes inactive as a result of treatment (Ekstrand, 2004). One would expect dental examiners would have 100% repeatability in recognizing the filling codes, but this was not the case as suggested by our study. While our examiners had excellent repeatability, the difficulty was in identifying full, partial sealant, and tooth coloured fillings especially in the occlusal surface. The examiners were very consistent in recognizing filling codes >3 (amalgam, crowns, lost and temporary restorations). Widespread use and high quality of any kind of resin restorations has made their identification difficult under clinical conditions. Also, with the use of flowable composites as the main occlusal restorative material and for preventive resin restoration, has made it difficult to assess whether a surface is covered with a sealant or a restoration.

Conclusions

The present study demonstrates the feasibility of using ICDAS in a community based clinical trial which is both easy and quick to use and the good to excellent reliability of the examiners in lesion and filling ICDAS criteria was sustained over a one year period. The extensive training and calibration in the ICDAS protocol made the study dental clinicians aware of its usefulness in clinical research and/or clinical practice.

Acknowledgement

We acknowledge Mr George J. Eckert, Indiana University School of Medicine and Dr Charles Spiekerman from the University of Washington for calculating the kappa statistics. The following individuals are acknowledged for data entry**:** Ms Anjum Kherani, Ms Sara Segel, and Ms Sarah Park. This research was supported in part by HRSA R40MC07838. The authors have no financial relationship relevant to this article to disclose.

References

Bader, J. (2001): *Diagnosis and management of dental caries. Agency for Healthcare Research and Quality Number 36.* AHRQ Publication No.01-E055, http://www.ahrq.gov/clinic/dentsumm.htm.

Braga, M.M., Oliveira, L.B., Bonini, G.A., Bonecker, M. and Mendes, F.M. (2009): Feasibility of the International Caries Detection and Assessment System II in epidemiological surveys and comparability with standard World Health Organization Criteria. *Caries Research* **43**, 245-249.

Cook, S., Martinez-Mier, A., Dean, J., Weddell, J., Sanders, B., Eggertsson, H., Ofner, S. and Yoder, K. (2008): Dental caries experience and association to risk indicators of remote rural populations. *International Journal of Pediatric Dentistry* **18**, 275-283.

Ekstrand, K.R. (2004): Improving clinical visual detection—potential for caries clinical trials. *Journal of Dental Research* **83 (Spec No C)**, C67-C71.

Ekstrand, K.R., Martignon, S., Ricketts, D.J.N. and Qvist, V. (2007): Detection and Activity Assessment of Primary Coronal Caries Lesions: a Methodologic Study. *Operative Dentistry* **32**, 225-235.

Fleiss, J.L. and Cohen, J. (1973): The equivalence of weighted kappa and the intra-class correlation coefficient as measures of reliability. *Educational and Psychological Measurement* **33**, 613-619.

International Caries Detection and Assessment System (ICDAS) Coordinating Committee (2007): *Rationale and Evidence for the International Caries Detection and Assessment System* II Early detection of dental caries: proceedings of the 2005 Indiana conference, Indianapolis: Indiana University.www.dundee.ac.uk/dhsru/docs/Rationale%20and%20Evidence%20ICDAS%20II%20September%2011.doc

Ismail, A.I. (2004): Visual and visuo-tactile detection of dental caries. *Journal of Dental Research* **82**, C56-C66.

Ismail, A.I., Sohn, W., Tellez, M., Amaya, A., Sen, A., Hasson, H. and Pitts, N.B. (2007): The International Caries Detection and Assessment System: an integrated system for measuring dental caries. *Community Dentistry and Oral Epidemiology* **35**, 170-178.

Ismail, A.I., Sohn, W., Tellez, M., Willem, J.M., Betz, J. and Lepkowski, J. (2008): Risk indicators for dental caries using the International Caries Detection and Assessment System. *Community Dentistry and Oral Epidemiology* **36**, 55-68.

Jablonski-Momeni, A., Stachniss, V., Ricketts, D.N., Heinzel-Gutenbrunner, M. and Pieper, K. (2008): Reproducibility and accuracy of the International Caries Detection and Assessment System II for detection of occlusal caries in vitro. *Caries Research* **42**, 79-87.

Kuhnisch, J., Berger, S., Goddon, I., Senkel, H., Pitts, N. and Heinrich-Weltzien, R. (2008): Occlusal caries detection in permanent molars according to the World Health Organization basic methods International Caries Detection and Assessment System II and laser fluorescence measurements. *Community Dentistry and Oral Epidemiology* **36**, 475-484.

Pitts, N.B. (2004): An international system for caries detection and assessment being developed to facilitate caries epidemiology, research and appropriate clinical management. *Community Dental Health* **21**, 193-198.

Pitts, N.B. and Richards, D. (2009): Personalized Treatment Planning. *Monographs in Oral Science* **21**, 128-143.

Ricketts, D.N.J., Ekstrand, K.R., Kidd, E.A.M. and Larsen, T. (2002): Relating visual and radiographic ranked scoring systems for occlusal caries detection to histological and microbiological evidence. *Operative Dentistry* **27**, 231-237.

Topping, G.V. and Pitts, N.B. (2009): Clinical Visual Caries Detection. *Monographs in Oral Science* **21**, 15-41. Epub Jun3.

United States Department of Health and Human Services (2000): *Oral health in America: a report of the Surgeon General. United States Department of Health and Human Services,* National Institute of Dental and Craniofacial Research National Institutes of Health.

Varma, S., Banerjee, A. and Bartlett, D. (2008): An in-vivo investigation of associations between saliva properties, caries prevalence and potential lesion activity in an adult United Kingdom population. *Journal of Dentistry* **36**, 294-299.