Community Dental Health (2017) **34,** 248–253
Received 2 December 2016; Accepted 25 April 2017
© BASCD 2017
doi:10.1922/CDH_4102Christian06

Examiner calibration in caries detection for populations and settings where *in vivo* calibration is not practical

Bradley Christian[1,3], Emily Amezdroz[1], Hanny Calache[2], Mark Gussy[3], Rachel Sore[4] and Elizabeth Waters[1]

[1]*Evidence and Child Health, The University of Melbourne;* [2]*Centre for Population Health Research, Deakin University;* [3]*Dentistry and Oral Health, La Trobe Rural Health School, La Trobe University,* [4]*Statistical Consulting Centre, The University of Melbourne*

Aim: To study examiner calibration methods for populations and settings where clinical (*in vivo*) calibration is not practical. ***Methods:*** Study design was cross-sectional and fully-crossed. The units of analysis were 880 tooth surfaces, from ten children ages 3 to 4 years. The study had three data components: (1) Examiner training and calibration using the *International Caries Detection and Assessment System* (ICDAS) e-Learning programme (2) *In vivo* community-based visual examination and (3) Intra-oral digital photographs of the same tooth surfaces from the *in vivo* visual examination. Kappa and weighted kappa scores were used to study reliability estimates. Systematic differences in caries assessments were determined using the Stuart Maxwell test. Data were analysed using STATA 13.1 and SAS 9.2. ***Results:*** Weighted kappa scores for the *in vivo* component ranged from 0.50 to 0.66 and from 0.64-0.74, for inter- and intra-examiner reliability, respectively. Caries lesions detected *in vivo* were also detected on photographs, albeit with increased severity when using photographs. For example, of 46 tooth surfaces assessed as being sound in the *in vivo* examination, 22 (48%) of these were assessed as having caries when photographs were used as the diagnostic method. ***Conclusions:*** From this research it appears that good quality photographs alone may be used for training and calibration among challenging populations or settings without adversely affecting data quality.

Key words: *dental caries, calibration, photographs, epidemiology, caries detection, oral diagnosis, caries assessment*

Introduction

Data quality in oral health research ensures valid estimates of disease are obtained (Nelson *et al.*, 2011). Validity is the ability of a test to distinguish between those with and without disease, and is commonly measured using sensitivity, specificity and positive predictive and negative predictive values. Reliability is the ability of the test to produce the same results when repeated (Chattopadhyay, 2010). Standardization in caries assessments can be achieved by training and calibrating examiners. Training involves examiners learning examination protocols and calibration is a process that allows measurements to be compared to reference standards (validity), and repeated scoring by the same and different examiners (intra-and inter-examiner reliability).

It is generally recommended that *in vivo* training and calibration exercises be conducted following photograph slide-based training, to give trainee examiners the opportunity to apply their training to the *in vivo* setting, where the clarity of a visual examination may not be as good as a photograph of a caries lesion. This step also allows examiners to study, discuss and evaluate their scoring of the *in vivo* presentation of caries lesions before data collection. *In vivo* calibration entails several examinations on the same person by different examiners, making it a rigorous process requiring maximum co-operation from study participants. Hence, while *in vivo* training and calibration is good practice, it may not always be possible, for reasons such as timing of data collection, lack of resources, specific population limitations (very young children, people with a disability and the elderly). For example, calibration exercises among pre-school children can be demanding on the child, parents/care-givers and examiners. Children, particularly those under five years, are more temperamental, unpredictable and reactive to their environment, thus making examinations more difficult (Aminabdi *et al.*, 2011). *In-vitro* methods for training and calibration could therefore be useful, but require assessment before they are used.

Several studies have investigated the use of intra-oral photographs as a caries diagnostic tool, but fewer have assessed their validity (Boye *et al.*, 2012, Elfrink *et al.*, 2009). A series of studies comparing intra-oral photographs with visual *in vivo* clinical assessments for caries detection demonstrated good comparability between the two methods (Boye *et al.*, 2013a, Boye *et al.*, 2013b, Boye *et al.*, 2013c). However, the unit of analysis in these studies was overall caries experience score (dmft) per study participant, which describes gross error rather than caries diagnosis at the tooth or surface level. In addition, most research has focused on caries diagnosis at the level of cavitation (dentine lesions) which does not recognize early non-cavitated stages of the disease process. More precise and sensitive diagnosis is particularly important for prevention and health promotion (Bader *et al.*, 2002). There has also been limited research in pre-school children, and to our knowledge no comparative studies in a community-based setting (Bader *et al.*, 2002, Gomez *et al.*, 2013).

Correspondence to: Bradley Christian, Evidence and Child Health, Melbourne School of Population and Global Health, The University of Melbourne, Level 5, 207 Bouverie Street, Victoria 3010, AustraliaTel: +61 3 90359926, Email: B.Christian@latrobe.edu.au

The purpose of this research was to study examiner calibration methods for populations and settings where clinical (*in vivo*) calibration is not practical. The objectives were to:

1. Determine whether photographs alone can be used in the training and calibration of examiners without adversely affecting data quality? This was achieved by assessing inter-and intra-examiner reliability in diagnosing the severity of caries lesions at the surface level, using: (a) Photographs of caries lesions sourced from the ICDAS e-Learning training programme and (b) *In vivo* community-based visual examination of children's teeth.
2. Assess systematic bias in *in vivo* caries diagnosis.
3. Compare caries diagnosis using intra-oral photographs of selected tooth surfaces with their corresponding *in vivo* community examination scores; to assess systematic differences in caries diagnosis.

Materials and Methods

Ethics approval for this study was provided by the Human Research Ethics Committee at the University of Melbourne (Ethics ID 1442648).

Study design and setting

This cross-sectional study used a fully-crossed design with the same examiners examining all study participants. This approach allowed systematic bias to be assessed and controlled for in reliability estimates (Hallgren, 2012).

Study participants were a sub-sample of 3- to 4-year-olds from the VicGeneration (VicGen) cohort of children residing in Melbourne, Australia. This age range was chosen for the full complement of primary teeth. The VicGen oral health birth cohort study began in 2008, to gain an understanding of the multifactorial nature of early childhood caries development (de Silva-Sanigorski *et al.*, 2010).

Sample size and participants

Based on estimates from the VicGen cohort, and with the intention to set the minimum acceptable weighted kappa level as 0.61, an estimated sample size of 50 teeth (250 surfaces) was required to detect a weighted kappa score of 0.87 (95% CI: 0.64, 1.00). The sample size calculation was performed by manipulating the estimated distribution of ratings until a sufficiently narrow 95% confidence interval was obtained. For this study, 880 tooth surfaces from 10 children were analysed. Ten children were purposively selected (based on caries activity) and invited to participate. This selection process ensured presence of caries lesions to reflect a range of ICDAS caries codes.

Data collection

Data collection involved three components: (1) Examiner training and calibration using the ICDAS e-Learning programme (2) *In vivo* community-based visual examination and (3) Intra-oral digital photographs of the same tooth surfaces from the *in vivo* visual examination.

Components (1) and (2) relate to objective one, which was to determine whether sufficient reliability was achieved in the *in vivo* assessments, when training and calibration was conducted using photographs from the ICDAS e-Learning programme (Topping *et al.*, 2008). Components (2) and (3), relate to objective three (Tables 5 & 6), where the *in vivo* assessments for selected surfaces were compared with their corresponding intra-oral photograph scores to determine whether caries lesions detected *in vivo* would also be detected on photographs to determine their validity.

Three dental examiners were used. Caries was diagnosed and scored using the International Caries Detection and Assessment System (ICDAS) (Ismail, 2007). ICDAS codes (Table 1) were used due to distinguish differing severity of lesions as compared with the WHO and BASCD methods, which only record dentinal lesions (cavities) (Altarakemah *et al.*, 2013, Braga *et al.*, 2009, Pine *et al.*, 1997). Teeth were not air-dried as this was considered impractical in the community setting. This lack of air drying prevented the use of ICDAS caries code 1 on smooth surfaces. The *in vivo* dental examinations were exclusively visual without the use of a dental probe.

Table 1. ICDAS Codes

Treatment Codes	Caries Codes
0 Sound	0 Sound surface
1 Sealant, partial	1 First visual change in enamel
2 Sealant, full	2 Distinct visual change in enamel
3 Tooth coloured restoration	3 Enamel breakdown: no dentine visible
4 Amalgam restoration	4 Dentinal shadow (not cavitated into dentine)
5 Stainless steel crown	5 Distinct cavity with visible dentine
6 Porcelain or gold or PFM crown or veneer	6 Extensive distinct cavity with visible dentine
7 Lost or broken restoration	
8 Temporary restoration	

ICDAS e-Learning programme training and calibration

The 90 Minute e-Learning programme by the ICDAS Foundation was used to train examiners in the application of ICDAS (Topping *et al.*, 2008). An experienced clinician and researcher (HC), facilitated the session, followed by a discussion of the scoring criteria, until agreement on the reference standard ICDAS code was achieved. For the calibration, examiners scored four sets of eight photographs each (32 photographs in total) of the various ICDAS codes, sourced from the e-Learning programme. Examiners were blinded to each other's scoring. Scoring was repeated to measure intra-examiner reliability.

In vivo community-based visual dental assessments

Each participant was visited at home twice, one week apart. At each visit the participant was examined by each of the three examiners who were blinded to each other's assessment. Five minute breaks were provided between examinations for each child. In total each child had six assessments.

Children were examined lying down on a flat surface with the examiner situated behind them. The examination involved standardised light sources (Henry Schein MirrorLites) and infection control protocol. A mouth mirror was used to view the teeth and gauze used to remove debris where required. Each quadrant was examined in a standardised manner. A trained research assistant recorded the caries scores onto paper forms which were then double-entered into an electronic database and checked for data entry consistency.

Intra-oral photographs

Intra-oral digital photographs of all tooth surfaces of participating children were taken to compare with the corresponding *in vivo* examinations. These photographs were taken at the first or second visit depending on the participant's convenience. All the photographs were taken by the same examiner, experienced in the use of the intra-oral camera and trained in the examination protocol. The Sopro Acteon 717 intra-oral camera had its own light source and the following settings: intra-oral mode, colour images and output format as JPG. The protocol for taking intra-oral photographs was similar to the *in vivo* examination. The images were captured digitally and stored directly to an electronic folder. The data were backed up onto an external hard drive at the end of each day.

For the analysis, 71 photographs were chosen based on clarity and to reflect a range of ICDAS codes. The photographs were scored by the same three examiners who conducted the *in vivo* examinations. A standardised environment was used to score the photographs, with the photographs projected onto a screen approximately 3 meters from the examiners. The room was dimmed during the session and each photograph was viewed for approximately 20 seconds. One examiner selected and collated the photographs. To avoid bias this examiner's scores were not used.

Data analysis

The unit of analysis was the tooth surface. A child with a complete primary dentition of 20 teeth will have 88 tooth surfaces in total. Five tooth surfaces for each molar and four for incisors and canines. Though both ICDAS codes were recorded, the focus for this analysis were the ICDAS caries codes.

Kappa and weighted kappa scores were used to determine inter- and intra-examiner reliability for tooth surface level scoring of caries lesions. Weighted kappa is recommended for determining examiner reliability for ordinal data such as ICDAS caries codes (0 (sound surface) – 6 (extensive decay)). This statistic incorporates the factor of agreement by chance alone and also weights proportional agreement (Fleiss, 1973, Maclure and Willett, 1987)1987. The 'kapci' command in STATA was used to generate kappa scores and 95% confidence intervals (95% CI), with the 'rep' option set at 999. This option uses a bootstrap method to generate estimates of the 95% CI.

Clustering effects (surfaces within teeth and teeth within individuals) were explored and found to be minimal and hence, estimates did not require adjustment for clustering (Litaker *et al.*, 2013, Masood *et al.*, 2015).

To assess systematic bias in scoring caries severity, the Stuart-Maxwell test for marginal homogeneity was conducted. This tests for the marginal homogeneity across all ICDAS categories simultaneously and determined if examiners were systematically scoring differently from other examiners or within themselves.

Data were analysed using STATA 13.1 and SAS 9.2.

Results

Reliability estimates for the two caries diagnostic methods are provided in Table 2. Using the e-Learning photographs, Kappa scores for inter-examiner reliability ranged from 0.56 to 0.70 (moderate to good). In the *in vivo* examination, using ICDAS caries codes only, weighted Kappa scores for inter-examiner reliability ranged from 0.50 to 0.64 (moderate to good).

Systematic differences in scoring caries lesions *in vivo* are summarised in Tables 3 and 4. Testing for marginal homogeneity between the three examiners showed that in some instances significant (p<0.05) differences were observed, both between and within examiners. From Table 3, it appears that examiner 2 was scoring differently to examiners 1 and 3, as determined by the non-significant p-value of 0.330 for examiner 1 against 3. This difference is shown in the marginal distribution of ICDAS scores in Table 4.

Tables 5 and 6 compare caries diagnoses of the same tooth surfaces using the *in vivo* examination and intra-oral photographs of the same teeth. To assess systematic differences, the scores for the two methods from one examiner were cross tabulated (Table 5). More caries lesions were diagnosed from photographs (Stuart-Maxwell p=0.003). For example, of the 46 tooth surfaces diagnosed as being sound (ICDAS code 00) *in vivo*, only 24 (52%) were diagnosed as sound using photographs. The remaining 22 experienced caries, off which 16 were non-cavitated caries lesions (ICDAS code 02). Seventeen of these 22 diagnostic differences occurred on posterior teeth.

Table 2. Estimates (95% CI) for inter- (at time 1) and intra-examiner reliability

Examiners	**1**	**2**	**3**
1	0.96 (0.88-1.00)[1] 0.60 (0.54-0.66)[2] 0.74 (0.65-0.81)[3]	0.60 (0.41-0.80) [1] 0.55 (0.48-0.61) [2] 0.52 0.40-0.63) [3]	0.70 (0.54-0.89) [1] 0.54 (0.47-0.60) [2] 0.64 (0.51-0.74) [3]
2		1.00[1] 0.65 (0.58-0.71) [2] 0.64 (0.51-0.74) [3]	0.56 (0.35-0.76) [1] 0.54 (0.48-0.61) [2] 0.50 (0.37-0.63) [3]
3			0.96 (0.88-1.00) [1] 0.56 (0.49-0.63) [2] 0.66 0.52-0.77) [3]

[1]Kappa scores from the ICDAS e-Learning training photographs (n=32 photographs)
[2]Kappa scores (all ICDAS codes) *in vivo* community dental examinations (n=880 tooth surfaces)
[3]Weighted Kappa scores (ICDAS caries codes only) *in vivo* community dental examinations (n=880 tooth surfaces)

Table 3. Stuart-Maxwell test for marginal homogeneity to assess systematic differences in the *in vivo* assessment of caries lesions, both between and within examiners

Examiners	**1**	**2**	**3**
1	P<0.001	*P*=0.030	*P*=0.330
2		*P*=0.093	*P*=0.020
3			*P*=0.444

Discussion

This methodological paper explored issues with examiner calibration in settings where *in vivo* calibration is not practical. The findings present an interesting scenario. On the one hand, moderate to good reliability, in *in vivo* caries diagnosis, was achieved when training and calibration was conducted used photographs. On the other, more caries lesions were detected from intra-oral photographs than *in vivo* on the same tooth surfaces. Two important lines of discussion arise from these observations: (1) the training & calibration of dental examiners and (2) caries detection, both of which are key to collecting quality epidemiological data.

Table 4. Marginal distribution of ICDAS caries codes by examiner from the *in vivo* community-based dental examination

ICDAS Caries Codes (n=880)	Examiners		
	1 (%)	2 (%)	3 (%)
00	89.5	89.2	91.0
01	0.5	0.1	0.1
02	6.4	8.0	6.2
03	1.1	0.9	0.5
04	0.9	1.3	0.6
05	0.8	0.2	0.8
06	0.5	0.1	0.5
Column totals	100	100	100

As stated earlier, calibration using participants in epidemiological studies is a rigorous process and may not always be possible because of the timing of data collection, lack of resources and/or specific population limitations. In such cases, alternative methods such as photographs and extracted teeth have been used, without an *in vivo* component, which may reduce data quality. *In vivo* data quality, which is affected by factors such as: magnification, saliva and debris on the tooth surfaces, lack of appropriate equipment and participants' behavior is rarely assessed following the use of in-vitro calibration. Similarly, the quality of intra-oral photographs is affected by factors such as the inability to move the camera head to a different angle to confirm diagnosis, image size, screen resolution, distance of camera head from tooth, number of teeth per photograph and difficulty in accessing particular tooth surfaces.

This study found that moderate to good reliability can be achieved in *in vivo* caries detection when training and calibration used photographs only. However, when the *in vivo* scores were assessed for systematic bias, one examiner scored differently to the others. This may have been the result of the presentation of caries lesions in vivo. Though the difference was statistically significant, whether it was clinically significant is questionable. This is because the difference was mainly observed in the scoring of non-cavitated lesions and sound surfaces, ICDAS caries codes 0-2.

Table 5. Cross tabulation (marginal frequencies) of caries diagnosis (ICDAS codes) using *in vivo* visual examinations and intra-oral photographs of the same tooth surfaces for one examiner

In vivo ICDAS Scores	Intra-oral Photograph ICDAS Scores												
	00	01	02	03	04	05	06	30	32	34	36	84	Total
00	24	0	16	0	3	2	0	1	0	0	0	0	46
01	0	0	2	0	0	0	0	0	0	0	0	0	2
02	1	0	5	1	0	0	0	0	1	0	0	1	9
03	0	0	3	1	0	0	1	0	1	0	0	0	6
04	0	0	0	0	0	0	0	0	0	0	0	0	0
05	0	0	0	0	0	2	0	0	0	0	0	0	2
06	1	0	0	0	0	1	0	0	0	0	0	0	2
30	0	0	1	0	0	0	0	0	0	0	0	0	1
32	0	0	0	0	0	0	0	0	0	2	0	0	2
34	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	1	0	1
84	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	26	0	27	2	3	5	1	1	2	2	1	1	71

Stuart Maxwell test for marginal homogeneity: p=0.003

Table 6. Distribution of 22 (of the 46) tooth surfaces that were diagnosed as sound (ICDAS code 00) in the *in vivo* examination but carious when using photographs in Table 5, by tooth location and surface.

Tooth location	Tooth surfaces					Total
	Distal	Occlusal1	Mesial	Buccal	Lingual	
Anterior	0	n/a	1	2	2	5
Posterior	2	6	0	5	4	17
Total	2	6	1	7	6	22

[1] Anterior teeth did not have an option to score occlusal surfaces and are hence not applicable (n/a)

Though there is no standard classification for the Kappa score, it is generally accepted that scores of over 0.5 indicate moderate to good reliability (Landis and Koch, 1977). The kappa estimates in this study were similar to those reported in the paper on the development of ICDAS (Ismail, 2007), in spite of being conducted in community conditions and in a challenging population.

Reliability estimates are influenced by the diagnostic method used, as well as by disease prevalence. Greater precision risks greater disagreement. A system like ICDAS with 7 levels of disease severity is more prone to disagreement than the WHO dmf/DMF index, which is dichotomous at the tooth level. Hence, the reason for WHO dmf/DMF achieving Kappa scores of 0.8. This paper used ICDAS rather than the WHO DMFT, because (1) is characterised by several distinct stages of progression not captured by the WHO dmf/DMF assessment of caries lesions, (2) non-cavitated lesions are reversible and (3) disease management should be based on the stage of disease (Fejerskov and Kidd, 2008).

Comparing the *in vivo* and intra-oral photographs, this study showed that lesions detected *in vivo* were also detected using photographs, by the same examiners, with a trend towards greater sensitivity when using photographs. From a calibration perspective, it means that photographs are not missing out those lesions detected *in vivo*, which shows that while *in vivo* is the ideal calibration method, photographs may also be used without adversely affecting data quality. The photographs used in this analysis were selected for their clarity ensuring the absence of introduced artefacts such as light being reflected off the tooth surface. In the absence of artefacts and assuming that the photographs were valid, the better examination conditions and magnification when using photographs could be the reason for more caries being detected using this method. This finding raises the issue of validity of caries detection methods, particularly clinical (visual) versus photographs.

The visual and visual-tactile methods continue to be the standard for caries detection in the clinic. These have also been used as the gold standard in validation studies (Bader *et al.*, 2002, Gomez *et al.*, 2013). However, the strength of evidence to support the validity of this method is poor, calling into question its use as the gold standard (Bader *et al.*, 2002). When studying the validity of caries detection methods, histological sectioning of a tooth for visual evidence of caries activity is considered the most appropriate gold standard, even though it has limitations (Bader *et al.*, 2002). A recent study, comparing the validity of occlusal caries assessments using *in vivo* visual and intra-oral photographs, with histology as the reference standard, demonstrated that the use of intra-oral photographs was the more sensitive method (Boye *et al.*, 2012). Median sensitivity was 81.3% for the photographs compared with 65.6% for the *in vivo* method. Specificity for both methods was similar at 81.3% and 82.4%, respectively. However, these values may be overestimated since the visual assessment and photographs were performed on extracted teeth without the challenges of the *in vivo* oral examination. In addition, caries was assessed at the dentine level, which excluded early non-cavitated caries lesions, which may have also contributed to the high specificity estimates (Boye *et al.*, 2012). Nonetheless, the evidence shows that good quality photographs are more valid for caries detection. Further research is required to improve the strength of this evidence, particularly in relation to non-cavitated lesions.

For epidemiological caries studies and oral health programs focused on prevention, it is important to detect lesions before they become cavitated (Gussy *et al.*, 2015). In such cases, the use of intra-oral photographs to detect caries offers several advantages over the commonly used visual-tactile method: (1) greater sensitivity, (2) lower cost if a trained research assistant takes the photographs, (3) multiple examinations can be performed on the photographs without burdening the participants, (4) reduced participant anxiety as a result of the absence of the power dynamic a dental practitioner brings to the examination and, (5) a database of intra-oral photographs can be created to verify any discrepancies in the assessment, which can also be used for training and calibration of new examiners. These advantages will have to be balanced against disadvantages such as: the cost of the cameras, training and standardisation of personnel, participant compliance, accessibility to the tooth surfaces, photographs providing only a 2-dimensional picture and electronic data management.

A limitation of this study was the number of examiners. More than three examiners are generally recommended to achieve the highest score on a quality rating scale (Bader *et al.*, 2002). It should be noted that the use of the modified ICDAS method (no air drying) eliminates ICDAS caries code 1 on the smooth surfaces and lesions in the pits and fissures that would have only been visible after drying. This would cause an underestimation of the true population caries estimate. However, for this study on reliability estimates and systematic differences, the results would remain unchanged as the caries assessment protocol used was similar across detection methods.

Conclusions

It appears that good quality photographs alone may be used for training and calibration, without adversely affecting data quality. Intra-oral photography could be a more sensitive method for caries detection over *in vivo* visual examination, with promising application in epidemiology. More research is required to determine the most effective and efficient examiner calibration methods in oral health, including larger studies testing different combinations of calibration techniques.

References

Altarakemah, Y., Al-Sane, M., Lim, S., Kingman, A. and Ismail, A. I. (2013): A new approach to reliability assessment of dental caries examinations. *Community Dentistry and Oral Epidemiology* **41,** 309-16.

Aminabdi, N. A., Puralibaba, F., Erfanparast, L., Najafpour, E., Jamali, Z. and Adhami, S. E. (2011): Impact of Temperament on Child Behavior in the Dental Setting. *Journal of Dental Research* **5,** 119-22.

Bader, J. D., Shugars, D. A. and Bonito, A. J. (2002): A systematic review of the performance of methods for identifying carious lesions. *Journal of Public Health Dentistry* **62,** 201-13.

Boye, U., Foster, G. R., Pretty, I. A. and Tickle, M. (2013a): The views of examiners on the use of intra-oral photographs to detect dental caries in epidemiological studies. *Community Dental Health* **30,** 34-8.

Boye, U., Pretty, I. A., Tickle, M. and Walsh, T. (2013b): Comparison of caries detection methods using varying numbers of intra-oral digital photographs with visual examination for epidemiology in children. *BMC Oral Health* **13,** 6.

Boye, U., Walsh, T., Pretty, I. A. and Tickle, M. (2012): Comparison of photographic and visual assessment of occlusal caries with histology as the reference standard. *BMC Oral Health* **12,** 10.

Boye, U., Willasey, A., Walsh, T., Tickle, M. and Pretty, I. A. (2013c): Comparison of an intra-oral photographic caries assessment with an established visual caries assessment method for use in dental epidemiological studies of children. *Community Dentistry and Oral Epidemiology* **41,** 526-33.

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 (ICDAS-II) in epidemiological surveys and comparability with standard World Health Organization criteria. *Caries Research* **43,** 245-9.

Chattopadhyay, A. (2010): *Oral Health Epidemiology Principles and Practice,* Sudbury, Massachusetts, Jones and Bartlett.

De Silva-Sanigorski, A. M., Calache, H., Gussy, M., Dashper, S., Gibson, J. and Waters, E. (2010): The VicGeneration study--a birth cohort to examine the environmental, behavioural and biological predictors of early childhood caries: background, aims and methods. *BMC Public Health* **10,** 97.

Elfrink, M. E., Veerkamp, J. S., Aartman, I. H., Moll, H. A. and Ten Cate, J. M. (2009): Validity of scoring caries and primary molar hypomineralization (DMH) on intraoral photographs. *European Archives of Paediatric Dentistry* **10 Suppl 1,** 5-10.

Fejerskov, O. and Kidd, E. (eds.) (2008): *Dental Caries: The Disease and Its Clinical Management,* Oxford: Blackwell Munksgaard Ltd.

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

Gomez, J., Tellez, M., Pretty, I. A., Ellwood, R. P. and Ismail, A. I. (2013): Non-cavitated carious lesions detection methods: a systematic review. *Community Dentistry and Oral Epidemiology* **41,** 54-66.

Gussy, M., Ashbolt, R., Carpenter, L., Virgo-Milton, M., Calache, H., Dashper, S., Leong, P., De Silva, A., De Livera, A., Simpson, J. and Waters, E. (2015): Natural history of dental caries in very young Australian children. *International Journal of Paediatric Dentistry*.

Hallgren, K. A. (2012): Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol* **8,** 23-34.

Ismail Ai, S. W., Tellez M, Amaya a, Sen a, Hasson H, Pitts Nb (2007): The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community Dentistry and Oral Epidemiology* **35,** 170-178.

Landis, J. R. and Koch, G. G. (1977): The measurement of observer agreement for categorical data. *Biometrics* **33,** 159-174.

Litaker, M. S., Gordan, V. V., Rindal, D. B., Fellows, J. L. and Gilbert, G. H. (2013): Cluster Effects in a National Dental PBRN restorative study. *Journal of Dental Research* **92,** 782-7.

Maclure, M. and Willett, W. C. (1987): Misinterpretation and misuse of the kappa statistic. *American Journal of Epidemiology* **126,** 161-9.

Masood, M., Masood, Y. and Newton, J. T. (2015): The Clustering Effects of Surfaces within the Tooth and Teeth within Individuals. *Journal of Dental Research* **94,** 281-8.

Nelson, S., Eggertsson, H., Powell, B., Mandelaris, J., Ntragtakis, M., Richardson, T. and Ferretti, G. (2011): Dental examiners consistency in applying the ICDAS criteria for a caries prevention community trial. *Community Dental Health* **28,** 238-242.

Pine, C. M., Pitts, N. B. and Nugent, Z. J. (1997): British Association for the Study of Community Dentistry (BASCD) guidance on the statistical aspects of training and calibration of examiners for surveys of child dental health. A BASCD coordinated dental epidemiology programme quality standard. *Community Dental Health* **14,** 18-29.

Topping, G. V. A., Hally, J., Bonner, B. and Pitts, N. B. (2008): International Caries Detection and Assessment System (ICDAS) e-Learning Package. Interactive CD ROM and Web-based software. London: Smile-on.