



Published in final edited form as:

*J Public Health Dent.* 2012 ; 72(2): 172–175. doi:10.1111/j.1752-7325.2012.00315.x.

## Examiner Reliability of Fluorosis Scoring: A Comparison of Photographic and Clinical Examination Findings

Noemi Cruz-Orcutt, DDS<sup>1</sup>, John J. Warren, DDS, MS<sup>1</sup>, Barbara Broffitt, MS<sup>1</sup>, Steven M. Levy, DDS, MPH<sup>1,2</sup>, and Karin Weber-Gasparoni, DDS, PhD<sup>3</sup>

<sup>1</sup>Department of Preventive & Community Dentistry, The University of Iowa, Iowa City, IA 52242

<sup>2</sup>Department of Epidemiology, The University of Iowa, Iowa City, IA 52242

<sup>3</sup>Department of Pediatric Dentistry, The University of Iowa, Iowa City, IA 52242

### Abstract

**Objective**—To assess and compare examiner reliability of clinical and photographic fluorosis examinations using the Fluorosis Risk Index (FRI) among children in the Iowa Fluoride Study (IFS).

**Methods**—The IFS examined 538 children for fluorosis and dental caries at age 13 and obtained intra-oral photographs from nearly all of them. To assess examiner reliability, duplicate clinical examinations were conducted for 40 of the subjects. In addition, 200 of the photographs were scored independently for fluorosis by two examiners in a standardized manner. Fluorosis data were compared between examiners for the clinical exams and separately for the photographic exams, and a comparison was made between clinical and photographic exams. For all 3 comparisons, examiner reliability was assessed using kappa statistics at the tooth level.

**Results**—Inter-examiner reliability for the duplicate clinical exams on the sample of 40 subjects as measured by kappa was 0.59, while the repeat exams of the 200 photographs yielded a kappa of 0.64. For the comparison of photographic and clinical exams, inter-examiner reliability, as measured by weighted kappa, was 0.46. FRI scores obtained using the photographs were higher on average than those obtained from the clinical exams. Fluorosis prevalence was higher for photographs (33%) than found for clinical exam (18%).

**Conclusion**—Results suggest inter-examiner reliability is greater and fluorosis scores higher when using photographic compared to clinical examinations.

### Keywords

Dental fluorosis; adolescent; reproducibility of results

### Introduction

Several epidemiological indices have been used to describe the clinical appearance of dental fluorosis (1,2). The choice of one of these fluorosis indices in a particular study or survey depends on the study's purpose (2), particularly researchers' desire to clearly assess different levels of fluorosis (2). It has also been shown that fluorosis is more apparent when teeth are dry, but it is difficult to accurately standardize tooth dryness (2,3). Because of this,

achieving high reliability in fluorosis examinations is difficult, and a few studies have suggested that use of photographs may help to optimize reliability (3–5).

Cochran et al. (3) assessed the reproducibility of a standardized photographic technique for recording fluorosis using the Thylstrup and Fejerskov (TF) and Developmental Defects of Enamel (DDE) indices in seven European countries. The results showed that intra- and inter-examiner agreement, as assessed by the kappa statistic, ranged from 0.32 to 0.70. Overall, when the teeth were not dried prior to photographs being made, the study found the reliability values to be slightly higher than when the teeth were dried (3). In addition, when photographic transparencies of wet teeth were examined using the TF index, 60% were defined as normal, but this dropped to 31% when the teeth viewed in the transparencies were dry (3).

Clinical examination data and photographic data were used to assess examiner reliability in a study of 49 children in Brazil (4). The results showed that inter-examiner kappa statistics ranged from 0.46 to 0.67, and that fluorosis prevalence was higher with the clinical examinations (49%) than with the photographic examinations (37%); however, no separate assessment of reliability was made for the photographic examinations, and it was unclear what criteria were used to assess fluorosis in this study (4).

Lastly, a study conducted in China by Wong et al (5) among 257 10- to 12-year-old children, assessed the level of agreement for the DDE index between clinical and standardized photographic examination, as well as intra-examiner agreement levels for both types of examination. At the tooth-level, kappa statistics ranged (depending on specific photographs used) from 0.61 to 0.91 for individual examiners' agreement between clinical and photographic examinations. For intra-examiner reliability at the tooth level, kappa statistics ranged from 0.73 to 0.95 for the photographic examinations, and 0.73 to 0.89 for the clinical examinations. The authors reported that the prevalence of fluorosis was slightly higher with the photographic evaluation than with the clinical examination (5).

These studies provide limited support for the use of photographs as a reliable means of scoring dental fluorosis, and provide somewhat conflicting results. For example, Wong et al (5) found higher prevalence of fluorosis with photographic examinations, but Martens et al (4) found lower prevalence, while Cochran noted prevalence of fluorosis in photographic examinations depended on the level of dryness of the teeth (3). However, none of these studies have assessed reliability for both photographic examinations and clinical examinations, and also compared the two methods. In addition, while these studies have included different indices, none have evaluated the reliability of using photographs for the Fluorosis Risk Index (FRI). Based on these limitations, the present study was conducted to assess and compare examiner reliability of clinical and photographic fluorosis examinations using the FRI among children in the Iowa Fluoride Study.

## Methods

The Iowa Fluoride Study (IFS) is an ongoing longitudinal study that has followed a birth cohort to assess how fluoride exposures, dietary and other factors have affected dental caries and dental fluorosis development (6,7). Examinations for dental caries and dental fluorosis were conducted when the children were approximately age 5, age 9 and age 13. The fluorosis exam results at age 13 (n=538) are the focus of these analyses.

The Iowa Fluoride Study cohort was recruited over a three-year period, so that in order to examine the children at approximately the same ages, examinations took place over an extended period, with individual examinations scheduled at the subject's convenience. To accommodate this sporadic examination schedule, the IFS employed two trained examiners

(examiners #1 and #2), with duplicate examinations (n=40) scheduled when possible throughout the examination period to assess inter-examiner reliability. The results of these duplicate examinations are reported here. No intra-examiner reliability assessments were conducted for the clinical examinations.

Clinical examinations for dental fluorosis were completed using the Fluorosis Risk Index (FRI), and were done with minimal drying of the teeth, as per the FRI protocol (8,9). The FRI is a method of scoring the buccal surface of each tooth by dividing these surfaces into 4 zones: the incisal edge, and incisal, middle and cervical thirds. Each zone is scored according to FRI criteria as: 0 = no fluorosis, 1 = questionable, 2 = definitive fluorosis and 3 = severe fluorosis (8,9). Thus, each tooth receives four scores – one for each zone.

Clinical photographs of the maxillary incisor teeth were obtained immediately after the IFS clinical examination for fluorosis and caries. As a result of the use of compressed air during the caries examination (which followed the separate fluorosis examination) the teeth had been systematically dried at the time of the photographs. All photographs were made with a Nikon digital camera and macro lens designed for intra-oral photography. All photographs were taken using the same light source, f-stop, 1:2 magnification and an automatic flash. The photographs were later downloaded into a PC-based computer system for storage and viewing. After all of the IFS examinations were complete, the photographs were viewed and 238 were included based on the subject having had the clinical examination completed by examiner #1, and not having current orthodontic treatment. From these, 24 were excluded due to lost or poor quality (e.g., excessive glare, over- or under-exposure) photographs, and 14 excluded due to missing, rotated or partially erupted maxillary incisors. A third trained examiner (examiner #3), who did not participate in the clinical exams, scored these 200 photographs, as did one of the original clinical examiners (examiner #1). In addition, intra-examiner reliability of photographic scoring was assessed by repeat scoring of 61 of the 200 photographs by examiner #3. All photographic scoring was done using the same computer and monitor in the same darkened room.

For these analyses, the clinical exam and photographic exam scores for each of the buccal surface zones (incisal edge, incisal, middle and cervical thirds) for all four maxillary incisor teeth were scored using FRI. Thus, each individual subject had 16 FRI zone scores. From these scores, the most involved score for each tooth was selected to assess tooth-level agreement (i.e., 4 scores per subject). These tooth-level FRI scores were compared between examiners and between the clinical examination scoring and photographic scoring, as well as intra-examiner (Examiner #3) evaluation for the photographic scoring. Percentage agreement and kappa statistics were used to assess reliability at the zone, tooth and person levels, with only the tooth-level data reported here due to the similarity of results. Specifically, simple kappa values were based on whether the most affected zone was scored as definitive fluorosis (FRI score of 2 or 3) or as no fluorosis/questionable (FRI score of 0 or 1) and were computed for assessments of reliability for the clinical examination and the photographic examination. For the comparison between clinical and photographic examination, weighted kappa values were generated in order to utilize the full range of FRI scores.

## Results

Tooth level inter-examiner reliability for the duplicate clinical examination scoring (n=40) is shown at the top of Table 1. Reliability for the clinical examinations was lower ( $\kappa=0.59$ ) than for scoring using the photographs ( $\kappa=0.64$ , middle section of Table 1, n=200). The kappa statistic for intra-examiner reliability of the photographic scoring by the single examiner (Examiner #3, n=61) was 0.71, as shown on the bottom part of Table 1. While the

differences in reliability were relatively small, the kappa values for the clinical examinations fell into the range of “moderate” (0.41 to 0.60) agreement, while the kappa values for the photographic examinations were considered as “substantial”(0.61 to 0.80) agreement, as defined by Landis and Koch (10).

Intra-examiner reliability between the photographs and clinical exams from the single examiner (Examiner #1, Table 2) was found to have a weighted kappa of 0.46, with the FRI scores obtained from the photographs being higher on average (mean=0.74) compared to those obtained clinically (mean=0.38). In addition, as shown in Table 2, the proportion of zones with FRI scores of 2 or greater was higher for the photographic examination (21%) than for the clinical examination (8%). Person-level fluorosis prevalence, defined as having two or more teeth with FRI scores of 2 or more, was higher for the photographic examinations (33%) compared to the clinical examinations (18%) (data not shown).

## Discussion

The study had two main findings: that using photographs to score fluorosis resulted in somewhat improved inter-examiner reliability when compared to clinical scoring (Table 1), and that the photographic examinations (after the teeth were dry) produced higher FRI scores on average and higher prevalence of fluorosis than did the clinical examinations where the teeth were not dried (Table 2). Neither of these findings were surprising. For example, Wong, et al (5), reported slightly higher *intra*-examiner reliability for photographic examinations than for clinical examinations, which is similar to the findings for inter-examiner reliability in the present study. The finding that fluorosis prevalence was higher upon photographic examination than for clinical examination is also consistent with the Wong, et al (5) study, and with the study reported by Cochran, et al (3), which reported higher fluorosis prevalence when the teeth were dry. In the present study, the clinical examinations were done with minimal drying of the teeth, but the teeth had been systematically dried prior to the photographs being taken. However, this latter finding is in contrast to the findings of a study conducted in Brazil which found higher prevalence upon clinical examination than for photographic examination (4). The authors of that study suggested that the difference in prevalence between the types of examination may have been due to different methods of drying the teeth –use of gauze for the clinical examination and “natural” drying for the photographic examination. In addition, that study used different examiners for the clinical and photographic examinations which also may have accounted for some of the differences.

The differences in examiner reliability and fluorosis prevalence obtained between the clinical and photographic examinations suggests that there may be trade-offs when choosing a means of assessing dental fluorosis. Clearly, thoroughly drying the teeth may produce better reliability, it is questionable whether fluorosis scored under these conditions is meaningful clinically, since the teeth normally exist in a moist environment. In contrast, clinically measuring fluorosis under “wet” conditions may miss some of the subtleties of fluorosis and may not detect its very mildest forms. Thus, choosing whether to include a photographic assessment in a particular study may depend on whether esthetics/clinical relevance or in a more detailed assessment of the biological condition is the study’s primary purpose.

While the study was able to assess reliability of both clinical and photographic examinations with a large number of subjects, it also had some limitations. First, while the teeth had been thoroughly dried just prior to obtaining the photographs (as part of a clinical caries exam using compressed air), we did not standardize the length of time the teeth were dried as was done in other studies (3,4). In addition, while all of the photographs were taken in a similar

manner in terms of magnification, f-stop and lighting, the focal distance or angles were not strictly standardized. Lastly, the IFS cohort is from a limited geographic area, is of generally higher socioeconomic status, and while fluorosis prevalence was moderate, most of the fluorosis was mild or very mild, so that this study was not able to assess reliability across a wide range of fluorosis severity.

In conclusion, results from this study suggest that photographic examination of maxillary incisors for dental fluorosis results in somewhat greater levels of examiner reliability than can be obtained through clinical examination. Given that photographs of the maxillary incisors are easily obtained, and that most studies of fluorosis and fluorosis risk have focused on the maxillary incisors, the study adds further evidence that photographic evaluation of fluorosis may be useful. However, the overestimation of prevalence compared to that obtained without full drying of the teeth must also be considered.

## Acknowledgments

This study was supported by NIH Grant R01-DE09551, and by a University of Iowa Dental Student Research Grant.

## References

1. Burt, BA.; Eklund, SA. Dentistry, Dental Practice and the Community, Chapter 22 - Dental Fluorosis. St. Louis: Elsevier Saunders; 2005. p. 287-93.
2. Rozier RG. Epidemiologic indices for measuring the clinical manifestations of dental fluorosis: overview and critique. *Adv Dent Res.* 1994; 8:39–55. [PubMed: 7993559]
3. Cochran JA, Ketley CE, Sanches L, Mamai-Homata E, Oila A-M, Arnadottir IB, van Loveren C, Whelton HP, O'Mullane DM. A standardized photographic method for evaluating enamel opacities including fluorosis. *Community Dent Oral Epidemiol.* 2004; 32 (Suppl 1):19–27. [PubMed: 15016113]
4. Martins CC, Chalub L, Lima-Arsati YB, Almeida Pordeus I, Martins Paiva S. Agreement in the diagnosis of dental fluorosis in central incisors performed by a standardized photographic method and clinical examination. *Cad Saude Publica.* 2009; 25:1017–24. [PubMed: 19488486]
5. Wong HM, McGrath C, Lo ECM, King NM. Photographs as a means of assessing developmental defects of enamel. *Community Dent Oral Epidemiol.* 2005; 33:438–46. [PubMed: 16262611]
6. Levy SM, Warren JJ, Broffitt B, Kanellis MJ. Associations between dental fluorosis of the permanent and primary dentitions. *J Pub Health Dent.* 2006; 66:180–5. [PubMed: 16913244]
7. Hong L, Levy SM, Broffitt B, Warren JJ, Kanellis MJ, Wefel JS, Dawson DV. Timing of fluoride intake in relations to development of fluorosis on maxillary central incisors. *Community Dent Oral Epidemiol.* 2006; 34:299–309. [PubMed: 16856950]
8. Pendrys D. The Fluorosis Risk Index: a method for investigating risk factors. *J Public Health Dent.* 1990; 50:291–8. [PubMed: 2231522]
9. Pendrys DG, Katz RV. Risk of enamel fluorosis associated with fluoride supplementation, infant formula, and fluoride dentifrice use. *Am J Epidemiol.* 1989; 130:1199–1208. [PubMed: 2589311]
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–74. [PubMed: 843571]

**Table 1** Tooth-Level Inter-examiner and Intra-examiner Agreement from Clinical Examination and Photographs for Fluorosis of the Maxillary Incisors.

Examiner #1 (clinical exam n=40)	Examiner #2 (clinical exam n=40) *			
	No fluorosis (FRI=0,1)	Definitive fluorosis (FRI=2,3)	Agreement	Kappa
No fluorosis	114	13	87%	0.59
Definitive fluorosis	7	20		
Examiner #1 (photographs n = 200)	Examiner #3 (photographs n=200)			
	No fluorosis (FRI=0,1)	Definitive fluorosis (FRI=2,3)	Agreement	Kappa
No fluorosis	552	79	87%	0.64
Definitive fluorosis	27	142		
Examiner #3 First assessment (photographs n=61)	Examiner #3 second assessment (photographs n =61)			
	No fluorosis (FRI=0,1)	Definitive fluorosis (FRI=2,3)	Agreement	Kappa
No fluorosis	139	10	83%	0.71
Definitive fluorosis	32	63		

\* Note that some teeth were not scored during the clinical examination because they were missing or incompletely erupted.

**Table 2**  
 Distribution of Tooth-Level FRI Scores and Intra-examiner Agreement of Clinical vs. Photographic Examination (Examiner #1)

	Clinical Examination FRI scores (n=200)				Total
	0	1	2	3	
0	368	13	3	0	384 (48%)
1	160	79	8	0	247 (31%)
2	34	79	50	0	163 (20%)
3	0	0	4	2	6 (1%)
Total	562 (70%)	171 (21%)	65 (8%)	2 (<1%)	800

Photographic Mean FRI score = 0.74; Clinical Mean FRI score = 0.38

62% agreement

Weighted kappa = 0.46