

UCSF

UC San Francisco Previously Published Works

Title

Examination criteria and calibration procedures for prevention trials of the Early Childhood Caries Collaborating Centers

Permalink

<https://escholarship.org/uc/item/02p5r72d>

Journal

Journal of Public Health Dentistry, 75(4)

ISSN

0022-4006

Authors

Warren, John J
Weber-Gasparoni, Karin
Tinanoff, Norman
et al.

Publication Date

2015-09-01

DOI

10.1111/jphd.12102

Peer reviewed

Examination Criteria and Calibration Procedures for Prevention Trials of the Early Childhood Caries

Collaborating Centers

John J. Warren, DDS, MS¹, Karin Weber-Gasparoni, DDS, MS, PhD¹, Norman Tinanoff, DDS, MS², Terence S. Batliner, DDS³, Bonnie Jue, DDS⁴, William Santo, BA⁴, Raul I. Garcia, DMD⁵, Stuart A. Gansky, DrPH⁴, and the Early Childhood Caries Collaborating Centers Caries Outcome Working Group

¹University of Iowa College of Dentistry

²University of Maryland School of Dentistry

³University of Colorado-Anschutz Medical Campus, School of Dental Medicine, Denver

⁴University of California, San Francisco School of Dentistry

⁵Boston University, Henry M. Goldman School of Dental Medicine

Abstract

Objectives: To summarize diagnostic criteria and examiner training and calibration of the NIDCR-funded Early Childhood Caries Collaborating Centers (EC4) and report examiner calibration results from 2010-2014. The EC4 at Boston University, University of Colorado, and University of California San Francisco are performing randomized controlled early childhood caries (ECC) prevention trials with caries as the main outcome measure. *Methods:* The EC4 with University of Iowa consultants developed standardized tooth and tooth surface status examination criteria for use in field conditions, examiner training materials, and examiner calibration and re-calibration methodologies. Calibration and re-calibration were performed with 1-5 year old children in the San Francisco Mission District in which assessments from each examiner to be calibrated were compared to those from a single gold standard examiner from 2010-2014. Cohen's kappa statistic was used to determine inter-examiner agreement. *Results:* A total of 7 examiners were successfully (re)calibrated during that period, examining a total of 231 children. Overall unweighted Cohen's kappas for 10 surface conditions exceeded the criterion of 0.70. However, separate agreement for assessment of non-cavitated lesions, as in other studies, was lower. *Conclusions:* An experienced

multi-disciplinary and multi-institutional team was able to develop criteria and training materials to anticipate situations and field conditions the main trials would encounter. Examiners were successfully trained and (re)calibrated.

Key Words:

Dental Caries; DMF Index; Physical Examinations and Diagnoses; Observer Variation; Reproducibility of Results; Calibration; Bias (Epidemiology); Child, Preschool; National Institute of Dental and Craniofacial Research (U.S.)

Introduction

Systematic dental caries assessment began in the 1920s-1930s. Early attempts focused on characterizing the extent of caries experience in children and adolescents(1). In the late 1930s Klein, *et al.*(2) developed the DMF index, which sums the number of frankly decayed, missing and filled teeth (DMFT) or tooth surfaces (DMFS). When it was developed, there was virtually no effective caries prevention, prevalence was high and treatment was primitive by today's standards. The index was limited almost entirely to surveillance to characterize population caries burden with conservative criteria, to prevent caries over-estimation. These criteria were adapted for the primary dentition in the 1940s as the def index, and later the dmf index(3). Radike (4) formalized the DMF/dmf index in the 1960s, which the National Center for Health Statistics, National Institute of Dental Research (NIDR; now the National Institute of Dental and Craniofacial Research, NIDCR), and the National Health and Nutrition Examination Survey (NHANES) (5) later adopted.

While the DMF/dmf index provides important information, it has shortcomings (1), including that caries is assessed dichotomously, as present or absent on each tooth or tooth surface, with only frank cavitated lesions being considered and all else considered sound. Such criteria do not reflect contemporary understanding of caries as a chronic disease process over time, with levels ranging from very early lesions to destroyed crowns .

However, alternatives to the traditional DMF/dmf criteria have been developed to meet current scientific knowledge (6,7). One alternative is the D₁-D₄ criteria, which the World Health Organization (8,9) originally developed, characterizing surface level caries as initial non-cavitated caries (D₁), caries limited to cavitated enamel (D₂), dentinal caries (D₃), and caries with pulpal involvement (D₄). The British Association for the Study of Community Dentistry's examiner calibration methodology (10) adopted a modified version of those criteria. The International Caries Assessment and Detection System (ICDAS) (11) classifies caries severity using a 7-category ordinal scoring scale ranging from 0 (sound surface) to 6

(severe cavitation). The ICDAS criteria require teeth be thoroughly cleaned, viewed with an overhead exam light, and dried with compressed air during examinations.

Caries examinations in preschool children present challenges, related to lack of cooperation, frequent movement, and difficulty in keeping teeth dry. Moreover, primary teeth are smaller and lighter colored than permanent teeth, increasing difficulty in assessment. Thus, systems such as WHO and ICDAS are often impractical with preschool children because ICDAS requires compressed air and very careful tooth inspection to score individual tooth surfaces.

NIDCR Centers for Research to Reduce Oral Health Disparities

In 2008 the NIDCR funded three research centers on oral health disparities focusing on reducing early childhood caries (ECC): Boston University's Center for Research to Evaluate and Eliminate Dental Disparities (CREEDD), University of Colorado Denver's Center for Native Oral Health Research (CNOHR), and University of California, San Francisco's Center to Address Disparities in Children's Oral Health (CANDO). These three centers became known as the Early Childhood Caries Collaborating Centers (EC4), with each center conducting separate randomized controlled trials to prevent ECC in high risk populations.

With common focus on ECC and a shared NIDCR-funded Data Coordinating Center, the EC4 concluded that the same criteria to assess caries should be developed and used across EC4 studies. The EC4 sought to develop practical criteria for high-risk populations of young children which would yield high agreement among multiple examiners.

This purpose of this paper is to describe standardized criteria for identifying and recording caries in preschool children utilizing caries severity levels, as well as procedures for training, calibrating and re-calibrating clinical examiners. We also present 2010-2014 (re)calibration results. These criteria were

adapted from previously published criteria (12), and are being used by the NIDCR funded Early Childhood Caries Collaborating Centers (EC4).

Methods

Developing the EC4 Caries Criteria

The EC4 caries criteria were adapted from those in the Iowa Fluoride Study (12), designed to provide data on non-cavitated “white spot” demineralized lesions, similar to the WHO D1 lesions (8,9), but also consistent with traditional DMF or dmf criteria for more advanced lesions and filled surfaces. An EC4 Caries Outcome Working Group (COWG) developed the initial examination protocol and criteria suitable for field settings, without compressed air. University of Iowa consultants provided further structure and details for training and calibration.

Caries Criteria

A knee-to-knee examination (Figure 1) was used (13) for smaller children, generally younger than 4, while a portable pediatric dental chair was used for older children. Debris and plaque were removed with a toothbrush (Oral-B, Iowa City, Iowa, USA) for 1 minute, which also served to acclimate the child to the exam setting. The EC4 examiners used the Defend Mirrorlite system (Mydent International, Hauppauge, New York, USA) and a clip-on headlamp (Surgitel, Ann Arbor, Michigan, USA) affixed to an eyeshield, with the teeth dried with gauze but without compressed air. The examinations were visual only (i.e., no tactile examination with an explorer or periodontal probe) with no magnification beyond corrective eye glasses.

The criteria distinguished cavitated from non-cavitated lesions, but not lesion depth (i.e., enamel and dentinal caries as in the D₁-D₄ or (d₁-d₄) WHO system). In short, while our d₁ classification closely

matches that of the d_1 - d_4 system (8,9), the term d_{2+} used by the EC4 essentially combines the d_2 , d_3 and d_4 WHO classifications.

For examinations, a two pass system similar to NHANES (5) was used. First, tooth-level status codes were recorded for all teeth. When primary teeth were missing, the determination of reason for missing a tooth was left to the examiner's professional judgment, based on eruption and exfoliation patterns, patterns of missing teeth and/or disease in the individual, and input from parent /caregiver (when present) considered for consistency with the clinical findings. For children age 3 or younger, outside of clear trauma situations, all missing teeth were considered unerupted or lost due to caries, depending on tooth eruption patterns. For children age 5 or older, some children may have begun exfoliating anterior primary teeth, which were considered naturally exfoliated. (See Table 1 for tooth status and tooth surface status codes used).

d_1 (White Spot or Non-Cavitated) Lesions (Code W)

Pit and fissure $_1$ lesions often appeared as distinct chalky white enamel directly adjacent to or into a pit or fissure, but in contrast to permanent dentition, were less often visible. Pit and fissure d_1 lesions typically appeared stained light to medium brown, but could be dark brown. Although usually not a consideration in young children, d_1 lesions did not appear shiny which helps distinguish d_1 lesions from arrested lesions. This classification implied no clinically visible or irreversible enamel structure loss in pits and fissures. No evidence of undermining (darkened subsurface seen through adjacent enamel) was present(12).

Smooth surface d_1 lesions were usually observed as distinct chalky white lines close to soft tissue margins (i.e. White Spot Lesion). There was no clinically visible or irreversible enamel structure loss or enamel surface break(12). On approximal smooth surfaces, scores were based on direct vision. Small smooth surface d_1 lesions presented challenges in distinguishing true lesions from enamel defects or

sound surfaces. Thus, smooth surface d_1 lesions needed to meet a minimum threshold. For single surface d_1 lesions, the lesion needed to extend at least one-third of the distance across the surface as a d_1 lesion. Single surface lesions extending less than one-third of the way across the surface were scored as sound. If a lesion extended at least one-third of the surface and extended beyond the line angle onto a second surface, the second surface was also scored as having a d_1 lesion, regardless of the extent of the lesion on the second surface (i.e. second surface did not need to meet the one-third rule).

d_{2+} (Cavitated or Frank) Lesions (Code D)

Pit and fissure d_{2+} lesions also may have had distinct chalky white enamel adjacent to a pit or fissure. Typically, the color was medium to dark brown, but could range from light to dark brown. For d_{2+} lesion classification, visible enamel structure loss was required (12).

Smooth surface d_{2+} lesions often appeared as distinct chalky white enamel, usually close to the soft tissue margin. Demonstrable enamel structure loss was required (12). For approximal smooth surfaces, d_{2+} lesions were scored only after direct visible confirmation that a break in the proximal enamel surface was evident with or without evidence of undermining or discoloration under the marginal ridge.

Definitions of Tooth Surfaces, Restored Surfaces and Other Considerations

Criteria defining individual tooth surfaces and scoring restored tooth surfaces were similar to DMF/dmf criteria. Specifically, anterior incisal tooth surfaces were not scored. If a lesion or restoration was confined solely to the incisal edge, the nearest adjacent surface was given the incisal score.

For a lesion on a posterior or anterior tooth that extended beyond the line angle onto another surface, the other surface was also scored as affected. When the tooth crown was destroyed by caries and only the roots remain, all surfaces were scored as frank (d_{2+}) caries. A posterior tooth restoration

was required to extend at least 1mm beyond the line angle to be considered as involving the adjacent surface. However, a proximal restoration on an anterior tooth was not considered to involve the adjacent labial or lingual surface unless it extended at least one-third into the surface. The reason for this criterion was that tooth structure on adjacent surfaces must often be removed to provide access for the restoration of a proximal lesion on anterior teeth.

Sealants were not often used in the primary dentition, so most tooth-colored (white) restorative materials were likely to be composite resin restorations (code F). Thus, when the examiner was reasonably certain that a composite material was used as restoration in any part of the fissure, the surface was scored as a non-amalgam restoration (code F). If there was sufficient evidence that the fissure had a sealant (not a restoration), the surface was scored as sealed (Code S). In case of doubt, the more conservative call regarding filled surfaces due to caries (in the spirit of epidemiologic surveys such as NHANES) was made and the surface was coded as sealed (Code S); i.e. may underestimate filled surfaces due to caries but overestimate sealants. Note that some composite materials were difficult to see. When examining a restoration for recurrent caries, a defective restoration was not considered carious in the absence of definitive visual criteria for caries.

Hypoplastic or malformed teeth were scored like any other teeth as in NHANES. However, when such a tooth was restored solely for aesthetic reasons as reported by a parent/caregiver, those surfaces were scored as sound (Code K). If a hypoplastic tooth without caries was restored with a full crown, the surfaces were coded sound. Similarly, restorations placed due to trauma (usually unilaterally on anterior teeth) not due to caries as reported by a parent/caregiver were scored as sound. Fractured teeth without caries were also scored as sound. Non-vital teeth were scored in the same manner as vital teeth. Any restorations deemed present solely due to root canal access without caries were not recorded as restorations, but were coded as sound. Teeth and tooth surfaces that could not be scored (e.g.

examination not completed, excessive calculus, gingival hemorrhage or, in older children, orthodontic brackets) were scored with Code Z.

When more than one condition existed on a given tooth surface, only one call per surface was made. For example, when a tooth was both filled (Code A or Code F) and frankly decayed (Code D), or when a sealant (Code S) was present along with frank (d_{2+}) decay (Code D), the tooth surface was scored as “decayed” – frank caries took precedence over restorations and sealants. This hierarchy and additional rules for multiple conditions on the same tooth surface are enumerated in Table 1. Temporary restorations were scored in the same way as permanent restorations. Fractured restorations were scored as if the restorations were intact unless caries was present. If frank (d_{2+}) caries was found within or adjacent to the margins of a fractured or missing restoration, frank caries was scored only in the surfaces involved. Missing restorations were scored as if the teeth were sound (Code K) unless there was caries.

Training and (Re)Calibration Procedures

Examiner Training

For the training portion, examiner-trainees were provided with a manual describing study protocols and examination criteria, and guidance regarding examination of young children. Examiner-trainees also had access to a collection of photographic slides that illustrated caries criteria and explained the examination protocol. The examiner-trainees reviewed these materials independently and then successfully completed a 15-item criteria quiz by scoring at least 12 (80%) correct.

Following individual review and quiz completion, the examiner-trainees met together with an examiner trainer (JJW) who reviewed the photographs, described and explained the criteria, and answered examiner-trainees' questions. Following this 3-hour session, the examiner-trainees completed a simulated calibration exercise using a separate set of photographic slides. The examiner-trainees were

required to score at least 90% of the surface calls correctly to be considered as “passing” the simulated calibration.

The last training program section included two clinical components: (1) the gold standard examiner (GSE) conducted demonstration examinations which examiner-trainees observed. These familiarized the examiner-trainees with the exam protocol and scoring, and allowed the GSE to point out clinical findings and approaches to examining young children; and (2) practice examinations to allow examiner-trainees to become familiar with instruments, the protocol, and their data recorder, as well as to discuss findings with the GSE. The recorders used a custom designed, HIPAA-compliant, user-friendly, stable, flexible, Flash-based software package, CAries Research INstrument(14), to collect calibration data in XML format.

Examiner Calibration

Traditionally, examiner calibration relies on repeat examination of individuals by different examiners, but in conducting calibration with young children the individual child's cooperation may allow for only two to four examinations per child. Moreover, it may not be practical to conduct full-mouth examinations for each calibration subject. Thus, certain provisions were made in conducting calibration with young children. These provisions included having the gold standard examiner (GSE) always perform the first or second exam on each child, and recording each child's cooperation level using the Frankl scale (15,16). Also, half-mouth examinations were frequently conducted for younger or less cooperative children; the first examiner selected the half-mouth with fewer sound surfaces. In addition to these provisions, the examination sequence was varied, so the GSE was not always the first examiner for a given child. Similarly, the examination sequence was varied, so each examiner-trainee calibrated was sometimes first, second or third, and examination order was balanced for all the examiner-trainees

calibrated. Thus, a specific examiner rotation was generated and followed during the calibration procedures. Figure 2 depicts the rotation pattern for examiner calibration.

Based on the statistical power required to assess levels of agreement, each examiner-trainee examined a minimum of 13 subjects paired with the gold standard examiner (GSE). Given the unpredictable cooperation level in this age group, calibration required planning to examine more than 13 subjects. To prepare the examiner-trainees for various situations, recruited calibration participants reflected a diverse array of disease levels. Study staff invited patients of record meeting age and caries status category targets, based on clinic records from the prior two months, for the calibration; about 10% additional potential participants were invited to allow for no shows due to illness and other circumstances. One parent/guardian provided written informed consent in English or Spanish for child participation. In addition, study examiners provided written informed consent to participate. The child's parent/guardian received a \$20 grocery voucher as compensation for the time required to travel to the clinic and participate. The UCSF institutional review board, the Committee on Human Research, reviewed and approved procedures for this (re)calibration.

To be considered calibrated as an examiner for the EC4 studies, each examiner-trainee was required to meet certain thresholds of surface-level agreement with the GSE, as assessed by the unweighted Cohen's kappa statistic for nominally scaled tooth surface conditions (since tooth surface scores do not follow a monotonically ordered disease progression). For nominally scaled surface level scores (excluding d_1 or Code W calls), examiner-trainees were required to reach surface level Cohen's kappa values of 0.75 or greater. In addition, overall surface-level Cohen's kappa values for all the nominally scaled surface level scores (including d_1 or Code W) were required to be 0.70 or higher. Examiner-trainees were required to achieve these agreement levels for both measures to be considered calibrated. Initially, for d_1 lesions, examiner-trainees were required to reach surface-level kappa values of 0.40 or greater. This criterion was ultimately not used, as it was unclear whether a kappa this low

provided sufficient agreement and not all examiners reached this threshold for d_1 lesions. As described earlier, each examiner-trainee was required to examine a minimum of 13 subjects along with the GSE. After each examiner-trainee and GSE completed examinations on 13 subjects, kappa scores were calculated for each examiner. If an individual examiner-trainee met the aforementioned standards on both criteria, they were certified as standardized and calibration was considered complete. When an individual examiner-trainee did not meet these standards after the first 13 subjects, the examiner-trainee discussed discrepancies/problems with the GSE and then was required to examine at least 7 new subjects, with agreement re-assessed. If the individual examiner-trainee met the agreement standards with the addition of these next 7 subjects, he or she was certified as calibrated; if not, he or she completed 7 additional examinations with the GSE. In practice, calibration was re-assessed in half-days for logistical reasons, which tended to correspond roughly to 7 new participants.

Given the longitudinal nature of all the EC4 studies, it is necessary that all examiners be annually re-calibrated to maintain high examiner agreement levels and adhere to study criteria. . Thus, re-calibration was conducted annually, with plans for annual re-calibration for the duration of each study. Examiner re-calibration used the same criteria, examination procedures, and agreement standards as described above for initial calibration.

Behavior Management Approaches

Examiners received written and “hands-on” instruction on behavior management approaches including normal behavior of young children during examinations and specific behavioral management techniques, such as distraction techniques, positive reinforcement techniques and the “tell-show-do” technique. Children with low Frankl scores (15,16) who could not be managed to comply with 2 examinations were excluded from the calibration exercise.

Unweighted Cohen's Kappa statistics for nominally scaled categories were calculated using standard methods three ways: (i) overall (up to 10 categories since U, T, X, M affect all surfaces of the corresponding tooth), (ii) pooling non-cavitated d_1 (W) surface codes with sound (K) surface codes (up to 9 categories), and (iii) non-cavitated versus not non-cavitated (2 categories). Kappa statistics were also calculated only for more cooperative children with positive or very positive Frankl behavior scores (3 or 4).

Results

The initial training and calibration focused on calibrating 3 examiners – one from each of the EC4 centers – to a gold standard examiner (KWG). Subsequent re-calibration of examiners followed in 2011 to 2014. Also, 4 additional examiners were trained and calibrated for the EC4 centers. The training, calibration, and recalibrations occurred at the San Francisco Native American Health Center (NAHC). NAHC provided space for training and (re)calibration, and recruited 50 1-5 year old children annually.

Agreement with the GSE on caries diagnosis was assessed for each examiner-trainee using Cohen's kappa statistics at the d_1 and d_{2+} levels and at a combined "total" level. Table 2 shows demographics of the children participating. They were evenly split between boys and girls; most were Hispanic, between 2 and 5, and many races (based on parent report) were represented. Table 3 shows the cross-classified age by caries category status used for calibration recruitment; emphasis was placed on those with at least 1 cavitated or restored surface (column 3) or with non-cavitated lesions (column 2). Mean caries indices at the non-cavitated and cavitated thresholds in Table 4 show fairly consistent amounts of disease over time, with an overall average of about 2 cavitated surfaces (d_{2+s}), about 2 restored surfaces ($d_{2+fs} - d_{2+s}$), less than $\frac{1}{2}$ missing surface ($d_{2+mfs} - d_{2+fs}$), and about $\frac{3}{4}$ non-cavitated surface ($d_{2+s} - d_{1+s}$ or $d_{2+mfs} - d_{1+mfs}$).

Table 5 summarizes the number of duplicate examinations and Cohen's kappa statistics for each of the 7 examiners (labeled A-G) matched pairwise with the GSE for the initial calibration in 2010 and re-calibration in 2011-2014. Not all examiners participated each year. Kappa statistics met both criteria for each examiner each year, although three times examiners qualifying overall kappa excluded the first round of exams (indicated with asterisks). One examiner only saw 12 children but was classified as calibrated rather than conduct an additional day of calibration for 1 child, since 1 child would not reduce the kappa statistics below the thresholds already exceeded for that examiner. Kappa statistics for non-cavitated lesions versus not non-cavitated lesions were rather low (often below 0.5) indicating that, despite training and specific criteria, they could not be reliably detected. Over all the years, an individual examiner's kappas do not have a consistent pattern, appearing to increase or decrease slightly; however, examiners tended to need to perform fewer exams to meet the criteria during re-calibration than for initial calibration. Most children were cooperative (Frankl scores of 3 or 4) and thus, examiner agreement among all children and those who were more cooperative did not appear to differ much, increasing or decreasing slightly.

Discussion

As the results suggest, using these criteria and training and calibration protocols yielded high levels of inter-examiner agreement for cavitated (d_{2+}) caries and for combined cavitated and non-cavitated (d_1 and d_{2+}) caries in very young children. However, as expected (12), inter-examiner agreement was much lower, measured by kappa, for d_1 level lesions. This is similar to findings of low examiner agreement for the ICDAS scores of 1 and 2 (11).

There are several possible reasons for lower agreement for d_1 lesions, mostly due to very small lesion size and their sometimes subtle appearance. Such lesions can be difficult to detect under even ideal circumstances, but with very young children, vision is often compromised due to lack of

cooperation (i.e., the child may not open his/her mouth) and difficulty to achieve and maintain a dry examination field. In addition, young children often move extensively during examination, so carefully observing a particular tooth surface may not always be possible. It is also important to note the protocol did not allow using magnification but only using normal corrective eye glasses. Finally, while the criteria were carefully developed to have precise wording and training focused on identifying these lesions, this process is inherently difficult in young children. We incorporated certain “rules” such as requiring smooth surface d_1 lesions to extend across at least one-third of the surface and considering all stained pits and fissures as d_1 lesions; nonetheless, agreement for d_1 lesions was modest, suggesting that even greater efforts may be necessary to achieve higher agreement levels. (Subsequent training and calibration session results achieved higher agreement levels and involved few subjects, further suggesting a longer learning curve for d_1 lesions.)

Our findings also suggest variation in agreement level achieved among different examiners, which may have been due to differences in previous experience with very young children. Specifically, greater agreement was achieved and achieved more quickly among examiners who had extensive experience with young children than for examiners with less experience. Thus, future studies involving young children may wish to consider examiners who are more experienced with such children when planning studies.

While the EC4 calibration was in many ways similar to that employed for NHANES examinations (17), there were some differences. Since examiners visited San Francisco only annually, for three days, the time for training and calibration was compressed. Intra-examiner agreement assessments were not feasible. Examiners could remember results from prior exams in a period that was short. Repeating exams in, say, one month was neither logistically or financially feasible. Prior work by UCSF indicated that intra-examiner agreement in young children was quite high (intra-examiner agreement $\kappa=0.96$) (18); (intra-examiner agreement $0.71 \leq \kappa \leq 0.87$) (19). In addition, to save time on-site, some training was

done via distance learning, particularly prior to recalibration. Also, in contrast to other calibration protocols (17), some discussion between the GSE and examiner-trainees was allowed after all examinations were completed for a given participant to improve understanding and facilitate speedier (re)calibration.

While developing the protocol, the COWG members debated whether to include ancillary information from the accompanying parent/caregiver about the reason for missing teeth. The COWG's pediatric dentists convincingly reasoned symmetric eruption, exfoliation, and caries patterns could be distinguished from asymmetric trauma patterns. However, epidemiologic validation studies show that parent report of child's oral health is highly positively correlated with clinical dental exam status (e.g. 20). Moreover, parent recall bias is low for child's tooth extraction due to caries: in 6-9 year old New Zealand children 18.7% of parents/caregivers reported a child with a tooth extraction due to caries while dental exam and child report both showed 19.2% had an extraction due to caries for kappa=0.92 (21). Thus, if the information from the parent/caregiver about the reason for tooth loss was available, the COWG decided to include that information. The main limitation is that when the parent/caregiver does not accompany the child this additional information is not available, resulting in a situation where information was available for some children but not others.

Another issue that arose during training and calibration included problems with detecting tooth-colored sealants and restorations, and determining whether tooth-colored materials were sealants or restorations. The criteria were conservative, so that restorations were not recorded unless the examiner was certain, with surfaces recorded as either sound or as having a sealant present, as appropriate. As tooth-colored materials have proliferated in recent years, their detection and classification present challenges that did not exist previously, and will likely continue to be a future challenge. Additionally, radiographs present a tradeoff between the potential gain in interproximal caries detection versus increased complexity and costs with young children in field settings as well as increased - though small -

radiation exposure. Thus, radiographs were not part of the examination protocol since they could not be used in the main EC4 trials, so interproximal lesions were likely underestimated. Efforts should be made to develop better visualization methods for restorations and sealants, as well as interproximal lesions, including protocols for using magnification and transillumination in standardized caries exams. However, such protocols must be relatively simple, quick and inexpensive to implement and not complicate the already difficult task of examining very young children.

In some cases, a child's normal reaction of crying during the examination made recording findings more challenging, sometimes making it difficult for the recorder to hear the examiner. The calibration protocol emphasizes the importance of examiners reviewing their findings prior to dismissing calibration subjects to assure that accurate recording occurred. In this way, inaccuracies in recording can be avoided so that agreement assessments measure only differences between examiners.

Finally, in developing criteria, protocols and training/calibration strategies for EC4 studies of young children, there was a desire to make the criteria as simple as possible while collecting reliable data accurately capturing clinical status. Along the same lines, there was a desire to reduce equipment needs to accommodate examinations conducted in challenging field settings. Thus, the COWG deliberately decided to only record caries at the non-cavitated and cavitated levels without requiring compressed air drying of tooth surfaces for the examination protocol. This decision resulted in some trade-offs – a somewhat less detailed assessment than might be available from other systems (e.g., ICDAS), and perhaps a decreased ability to detect lesions and restorations without compressed air. However, the decision resulted in a pragmatic system amenable to use with very young children. In particular, while compressed air enhances examiners' ability to assess tooth surfaces, it can also be cumbersome to use and disconcerting for younger children. Another decision made was to utilize mostly young children for the calibration sessions, which had the drawback of limited cooperation among participants; however,

having older, more cooperative children in the calibration exercises would likely over-estimate examiner agreement relative to the planned studies' populations of younger children.

The EC4 studies focus on preventing early childhood caries (ECC). A necessary part of studying any disease is accurately measuring it. While certain provisions must be made for dealing with the young child population as described herein, our results demonstrate that through fairly simple criteria, along with extensive training and (re)calibration, it is possible to achieve acceptable inter-examiner agreement levels in very young children.

Acknowledgments

The authors thank Dr. Carolyn Brown who was Director of the NAHC Dental Clinic through 2014, as well as Karina Alcala and the other Native American Health Center Staff, for their willingness and enthusiasm to make these (re)calibrations possible. Most of all we thank the families who participated in the (re)calibrations.

This publication was made possible by award numbers U54DE019285, U54DE019275, and U54DE019259 from the National Institute of Dental and Craniofacial Research (NIDCR), a component of the US DHHS National Institutes of Health (NIH). CARIN was developed with initial support from award number R21DE08650 and U54DE014251 and refined with support from award number U54DE019285, both from NIDCR, a component of NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Draft manuals were developed and reviewed by members of the Early Childhood Caries Collaborating Centers Caries Outcome Work Group, which has its full roster listed on the website <http://www.oralhealthdisparities.org>.

The Initial Caries Outcome Working Group was comprised of: Jane Atkinson, Terry Batliner, Lisa Chung, Stuart Gansky, Raul Garcia, Steven Gregorich, Bill Henderson, Bonnie Jue, Ruth Nowjack-Raymer,

Martha Nunn, Valerie Orlando, Francisco Ramos-Gomez, Margaret Rasmussen, William Santo., Norman Tinanoff, John Warren, Karin Weber-Gasparoni, and Jane Weintraub.

References

1. Burt BA, Eklund SA. *Dentistry, Dental Practice and the Community*, 6th Ed. St. Louis: Elsevier-Saunders; 2005.
2. Klein H, Palmer CE, Knutson JW. Studies on dental caries: I. Dental status and dental needs of elementary school children. *Pub Health Rep.* 1938;53:751-65.
3. Gruebbel AO. A measure of dental caries prevalence and treatment service for deciduous teeth. *J Dent Res.* 1944;23:163-8.
4. Radike AW. Criteria for diagnosis of dental caries. In: Proceedings of the conference on the clinical testing of cariostatic agents. Chicago: American Dental Association; 1972. p. 87-8.
5. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National Health and Nutrition Examination Survey: Plan and operations, 1999–2010. National Center for Health Statistics. *Vital Health Stat.* 2013;1(56):1-37.
6. Pitts NB, Fyffe HE. The effect of varying diagnostic thresholds upon clinical caries data for a low prevalence group. *J Dent Res.* 1988;67: 592-96.
7. Pitts NB. Diagnostic tools and measurements – impact on appropriate care. *Community Dent Oral Epidemiol.* 1997;25:24-35.
8. World Health Organization. *Oral Health Surveys - Basic Methods*. 4th ed. Geneva: World Health Organization; 1997. 68 p.
9. World Health Organization. *A Guide to Oral Health Epidemiological Investigations*. Geneva: World Health Organization; 1979. 42 p.
10. Pine CM, Pitts NB, Nugent ZJ. 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 Dent Health.* 1997;14 Suppl 1: 18-29.
11. Ismail AI, Sohn W, Tellez M, Amaya A, Sen A, Hasson H, Pitts NB. The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community Dent Oral Epidemiol.* 2007;35:170-8.

12. Warren JJ, Levy SM, Kanellis MJ. Dental caries in the primary dentition: assessing prevalence of cavitated and non-cavitated lesions. *J Pub Health Dent*. 2002;62:109-14.
13. Nowak AJ, Warren JJ. Infant oral health and oral habits. *Pediatr Clin N America*. 2000;47:1043-66.
14. CARIN: CAries Research INstrument Software Package. Tech ID: 19030 / UC Case 2009-035-0. San Francisco: Regents of the University of California; c2009 [updated 2010; cited 2014 Aug 13]. Available from: <http://techtransfer.universityofcalifornia.edu/NCD/19030.html>.
15. Frankl SN, Shiere FR, Fogels HR. Should the parent remain with the child in the dental operator? *J Dent Child*. 1962;29:150-163.
16. Aartman IHA, van Everdingen T, Hoogstraten J, Schuurs AHB. Appraisal of Behavioral Measurement Techniques for Assessing Dental Anxiety and Fear in Children: A Review. *J Psychopath Behav Assess*. 1996;18(2):153-171.
17. Dye BA, Barker LK, Selwitz RH, Lewis BG, Wu T, Fryar CD, Ostchega Y, Beltran ED, Ley E. Overview and quality assurance for the National Health and Nutrition Examination Survey (NHANES) oral health component, 1999–2002. *Community Dent Oral Epidemiol* 2007; 35: 140–151.
18. Weintraub JA, Ramos-Gomez F, Jue B, Shain S, Hoover CI, Featherstone JD, Gansky SA. Fluoride varnish efficacy in preventing early childhood caries. *J Dent Res*. 2006; 85(2):172-6.
19. Ramos-Gomez FJ, Gansky SA, Featherstone JD, Jue B, Gonzalez-Beristain R, Santo W, Martinez E, Weintraub JA. Mother and youth access (MAYA) maternal chlorhexidine, counselling and paediatric fluoride varnish randomized clinical trial to prevent early childhood caries. *Int J Paediatr Dent*. 2012;22(3):169-79.
20. Talekar BS, Rozier RG, Slade GD, Ennett ST. Parental perceptions of their preschool-aged children's oral health. *J Am Dent Assoc*. 2005;136(3):364-72.
21. Jamieson LM, Thomson WM, McGee R. An assessment of validity and reliability of dental self-report items used in a National Child Nutrition Survey. *Commun Dent Oral Epidemiol* 2004; 32(1):49-54.

Table 1. Tooth, surface and hierarchy codes for examiner determinations of clinical conditions

Group	Code	Clinical Condition
Tooth Status Codes	P	Present (but not a tooth status below)
	R	Partially Erupted
	K	Sound (all surfaces)
	U	Unerupted
	M	Missing due to caries
	T	Missing due to trauma
	X	Missing due to exfoliation
	O	Missing due to other or unknown reasons
	C	Crown including stainless steel or other
	Z	Unable to score
Tooth Surface Status Codes	K	Sound
	D	Cavitated decayed (d ₂₊) lesion
	W	Demineralized (white spot or d ₁) lesion
	A	Filled surface - amalgam
	F	Filled surface - non-amalgam
	S	Sealed surface
	U	Unerupted surface
	Z	Unable to score
Precedence Hierarchy	D > A, F, W, S	Cavitated lesion takes precedence over any filled lesion, white spot lesion, or sealant
	A > F, W, S	Amalgam filling takes precedence over non-amalgam filling, white spot lesion, or sealant
	F > W, S	Non-amalgam filling takes precedence over white spot lesion, or sealant
	W > K, S	White spot lesion takes precedence over sound surface or sealant

Table 2. Child sociodemographics (gender, ethnicity, race, and age) by year and overall.

Measure	Level	2010	2011	2012	2013	2014	Overall
		(n=48)	(n=41)	(n=45)	(n=39)	(n=58)	(n=231)
		% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Gender	Female	31.3 (15)	61.0 (25)	48.9 (22)	53.8 (21)	50.0 (29)	48.5 (112)
	Male	66.7 (32)	36.6 (15)	46.7 (21)	46.2 (18)	46.6 (27)	48.9 (113)
	Missing	2.1 (1)	2.4 (1)	4.4 (2)	0 (0)	3.4 (2)	2.6 (6)
Ethnicity	Hispanic	91.7 (44)	82.9 (34)	82.2 (37)	89.7 (35)	69.0 (40)	82.3 (190)
	Non-Hispanic	0 (0)	14.6 (6)	11.1 (5)	7.7 (3)	24.1 (14)	12.1 (28)
	Unknown	8.3 (4)	2.4 (1)	6.7 (3)	2.6 (1)	6.9 (4)	5.6 (13)
Race	African-	6.3 (3)	2.4 (1)	6.7 (3)	7.7 (3)	12.1 (7)	7.4 (17)
	American /						
	Black						
	American Indian	0 (0)	14.6 (6)	6.7 (3)	2.6 (1)	1.7 (1)	4.8 (11)
	/ Native Alaskan						
	Caucasian /	0 (0)	14.6 (6)	35.6 (16)	20.5 (8)	12.1 (7)	16.0 (37)
	White						
	Other	0 (0)	26.8 (11)	26.7 (12)	46.2 (18)	50.0 (29)	30.3 (70)
>1 Race	0 (0)	22.0 (9)	4.4 (2)	0 (0)	3.4 (2)	5.6 (13)	
Unknown	93.8 (45)*	19.5 (8)	20.0 (9)	23.1 (9)	20.7 (12)	35.9 (83)	
Age	1	8.3 (4)	0 (0)	4.4 (2)	10.3 (4)	10.3 (6)	6.9 (16)
	2	18.8 (9)	0 (0)	17.8 (8)	15.4 (6)	17.2 (10)	14.3 (33)
	3	14.6 (7)	0 (0)	20.0 (9)	30.8 (12)	13.8 (8)	15.6 (36)
	4	27.1 (13)	0 (0)	31.1 (14)	20.5 (8)	37.9 (22)	24.7 (57)
	5	29.2 (14)	0 (0)	24.4 (11)	20.5 (8)	19.0 (11)	19.0 (44)
	Missing†	2.1 (1)	100 (41)	2.2 (1)	2.6 (1)	1.7 (1)	19.5 (45)

*Race data not available for 2010

†Age data not available for 2011

Table 3. Distribution of children by age group and caries status category* (N=231).

Age (years)	No decay (d₁₊mfs=0)	Only d₁ lesions (d₁>0 and d₂₊mfs=0)	At least 1 cavitated or filled surface (d₂₊fs>0)	Severe decay (2+ teeth each with 4+ dmf surfaces)	Total
	% (n)	% (n)	% (n)	% (n)	% (n)
<3	32.6 (16)	36.5 (18)	30.6 (15)	2.0 (1)	21.2 (49)
≥3	12.4 (17)	17.5 (24)	70.1 (96)	13.1 (18)	59.3 (137)
Missing†	33.3 (15)	13.3 (6)	53.3 (24)	2.2 (1)	19.5 (45)
Total	20.8 (48)	20.8 (48)	58.4 (135)	8.7 (20)	100.0 (231)

* Gold standard examiner determined

† Age data not available for 2011

Table 4. Primary tooth surface caries indices* by year and overall.

Caries Measure	2010 (n=48) Mean (SD)†	2011 (n=41) Mean (SD)	2012 (n=45) Mean (SD)	2013 (n=39) Mean (SD)	2014 (n=58) Mean (SD)	Overall (n=231) Mean (SD)
d ₁₊ mfs	5.6 (5.9)	4.3 (6.1)	4.8 (5.3)	5.6 (5.7)	5.3 (4.9)	5.2 (5.5)
d ₁₊ s	2.9 (3.7)	2.2 (3.2)	2.8 (3.3)	3.0 (3.0)	2.9 (3.1)	2.8 (3.2)
d ₂₊ mfs	4.6 (5.3)	3.6 (5.6)	4.7 (7.2)	4.5 (6.1)	4.7 (5.5)	4.5 (6.0)
d ₂₊ fs	4.4 (4.9)	3.5 (5.6)	4.4 (6.7)	4.0 (5.7)	4.7 (5.5)	4.2 (5.7)
d ₂₊ s	2.0 (3.2)	1.4 (2.9)	2.7 (5.8)	1.8 (2.7)	2.3 (3.7)	2.1 (3.8)

Gold standard examiner determined

† SD = standard deviation

Table 5. (Re)Calibration agreement of each examiner versus gold standard examiner, 2010-2014*; for all children and only more cooperative children (Frankl scores 3 or 4).

Year	Examiner	n	All			Frankl 3+			
			Overall† kappa (criterion n 0.70)	d ₁ □sound‡ kappa (criterion 0.75)	d ₁ /no d ₁ †† kappa	Overall† kappa (criterion n 0.70)	d ₁ □sound‡ kappa (criterion 0.75)	d ₁ /no d ₁ †† kappa	
2010	A	29	0.77	0.88	0.38	22	0.74	0.86	0.38
	B	17	0.73	0.82	0.54	9	0.70	0.75	0.55
	C	31	0.65§	0.80	0.16	20	0.62	0.78	0.12
2011	B	14	0.77	0.88	0.38	11	0.75	0.87	0.35
	C	17	0.71	0.79	0.20	10	0.74	0.83	0.21
	D	22	0.80	0.87	0.35	17	0.81	0.88	0.36
	E	13	0.84	0.90	0.16	9	0.87	0.94	0.18
2012	B	14	0.76	0.85	0.55	11	0.83	0.89	0.61
	C	23	0.67§	0.83	0.15	21	0.68	0.84	0.17
	D	15	0.78	0.89	0.46	14	0.78	0.89	0.46
	E	20	0.82	0.87	0.50	16	0.83	0.87	0.52
	F	18	0.75	0.86	0.40	16	0.67	0.79	0.40
2013	B	14	0.77	0.87	0.52	11	0.81	0.88	0.59
	C	14	0.75	0.90	0.20	11	0.69	0.85	0.20
	D	13	0.79	0.88	0.51	9	0.81	0.90	0.56
	E	14	0.76	0.83	0.49	11	0.77	0.83	0.52
	F	12	0.78	0.83	0.45	9	0.79	0.83	0.52
2014	B	14	0.83	0.89	0.46	10	0.87	0.94	0.37
	C	13	0.80	0.87	0.31	12	0.80	0.88	0.27
	D	13	0.73	0.85	0.43	12	0.72	0.84	0.42
	E	13	0.71	0.78	0.45	12	0.74	0.77	0.58
	F	25	0.68§	0.76	0.46	23	0.67	0.75	0.45
	G	19	0.69§	0.84	0.33	15	0.66	0.83	0.32

* Native American Health Center, San Francisco

† “Overall kappa” is unweighted Cohen’s kappa for 10 nominal tooth surface categories

‡ “d₁□sound” is unweighted Cohen’s kappa for 9 nominal tooth surface categories with non-cavitated caries (d1 or code W) recoded as sound (code K)

¶ “d₁/no d₁” is unweighted Cohen’s kappa for 2 tooth surface categories: non-cavitated caries (d1 or code W) versus any other status

§ Overall results shown; Final session results $k > 0.70$

- 12 children examined

Figure 1. Knee-to-knee dental examination position. Knee-to-knee is the most practical and comfortable way to position young children during the examination. Examiner and caregiver face each other with knees touching while caregiver holds child on lap on a straddle position with child facing caregiver. Caregiver carefully lays child on examiner's lap and holds the child's hands during the examination. Examiner stabilizes child's head to control movement for the child's safety and to allow a better examination



Figure 2. Example of examiner rotation scheme for calibrating three examiners with one Gold Standard Examiner (GSE)

Subject	Examiner	Subject	Examiner
Subject 1	<ol style="list-style-type: none"> 1. GSE* 2. Examiner 1 3. Examiner 2† 	Subject 2	<ol style="list-style-type: none"> 1. Examiner 2 2. GSE 3. Examiner 3†
Subject 3	<ol style="list-style-type: none"> 1. Examiner 1 2. GSE 3. Examiner 2† 	Subject 4	<ol style="list-style-type: none"> 1. GSE 2. Examiner 3 3. Examiner 1†
Subject 5	<ol style="list-style-type: none"> 1. GSE 2. Examiner 2 3. Examiner 3† 	Subject 6	<ol style="list-style-type: none"> 1. Examiner 3 2. GSE 3. Examiner 1†

* GSE = Gold Standard Examiner

† Third examination done when child cooperation permits; Four examinations could be performed with extremely cooperative children who usually were accustomed to dental exams.