

Guglielmo Campus (Orcid ID: 0000-0002-8573-485X)

Markovic Dejan (Orcid ID: 0000-0002-4297-6744)

Vukovic Ana (Orcid ID: 0000-0002-6880-6153)

## Multilevel Calibration procedure for the oral health national multicentre survey in primary teeth.

### **Running head**

Multilevel calibration on primary dentition

Dejan Markovic<sup>a</sup>,

Ana Vukovic<sup>a</sup>,

Ivan Soldatovic<sup>b</sup>,

Tamara Peric<sup>a</sup>,

Biljana Kilibarda<sup>c</sup>,

Ruxandra Sava Rosianu<sup>d</sup>,

Guglielmo Campus<sup>e</sup>.

### **Statement of Author Contributions**

*Dejan Markovic:* Contributed to conception, design, drafted and critically revised the manuscript, study design, manuscript writing, review final revision of the manuscript. *Ana Vukovic:* Contributed to conception, design, drafted and critically revised the manuscript, study design, manuscript writing, review final revision of the manuscript. *Ivan Soldatovic:* Contributed to review the literature, design, data acquisition, statistical analysis and interpretation, critically revised the manuscript, review final revision of the manuscript. *Tamara Peric:* Contributed to conception, design, drafted and critically revised & editing the manuscript. *Biljana Kilibarda:* Contributed to conception, statistical interpretation, manuscript writing, critically revised the manuscript. *Ruxandra Sava Rosianu:* Contributed to conception, design, drafted and critically revised & editing the manuscript. *Guglielmo Campus:* Contributed to conception, design, formal analysis, methodology, data and statistical analysis, drafted and critically revised the manuscript.

<sup>a</sup>Department of Pediatric and Preventive Dentistry School of Dental Medicine, University of Belgrade, Serbia; [dejan.markovic@stomf.bg.ac.rs](mailto:dejan.markovic@stomf.bg.ac.rs), [ana.vukovic@stomf.bg.ac.rs](mailto:ana.vukovic@stomf.bg.ac.rs), [tamara.peric@stomf.bg.ac.rs](mailto:tamara.peric@stomf.bg.ac.rs)

<sup>b</sup>Department of Statistics and Bioinformatics, School of Medicine, University of Belgrade, Belgrade, Serbia [soldatovic.ivan@gmail.com](mailto:soldatovic.ivan@gmail.com)

<sup>c</sup>Institute of Public Health of Serbia "Dr Milovan Jovanovic Batut", Belgrade, Dr Subotica 5, 11000 Belgrade, Serbia [kilibarda\\_b@batut.org.rs](mailto:kilibarda_b@batut.org.rs)

<sup>d</sup>Department of Preventive, Community Dentistry and Oral Health, Faculty of Dentistry, University of Medicine and Pharmacy "Victor Babes" Timisoara, Romania; [sava-rosianu.ruxandra@umft.ro](mailto:sava-rosianu.ruxandra@umft.ro)

<sup>e</sup>Department of Pediatric and Preventive Dentistry, Dental Clinic, University of Bern, Switzerland; [guglielmo.campus@zmk.unibe.ch](mailto:guglielmo.campus@zmk.unibe.ch) (ORCID <https://orcid.org/0000-0002-8573-485X>)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1111/ipd.13072](https://doi.org/10.1111/ipd.13072)

**Corresponding author:**

Prof. Guglielmo Campus  
Department of Pediatric and Preventive Dentistry, Dental Clinic  
University of Bern,  
Freiburgstrasse 7, CH-3010 Bern, Switzerland  
Tel +41 0798441703  
Email [guglielmo.campus@zmk.unibe.ch](mailto:guglielmo.campus@zmk.unibe.ch)

**Acknowledgments**

The survey was done within the Program for Oral Health Improvement in Children and Youth of the Republic of Serbia, approved by the Government of Serbia, under the Ministry of Health (program number 1802, project number 4015). The authors would like to express sincere appreciation to the Borrow Foundation, WHO Country Office for Serbia, WHO Collaborating Center for Epidemiology and Public Health at University of Milan, and the team from the Prevention of Non-communicable Diseases (NCDs) department, WHO Headquarters, Geneva (Dr. Benoit Varenne). We gratefully acknowledge the kind help of pediatric dentists, oral hygienists and assistants from site locations in collecting all the necessary data for the nationwide database: Dr. Aleksic Tanja, Dr. Bogdanovic Natasa, Dr. Bozic Zorica, Dr. Drazic Jovanka, Dr. Djurkovic Marija, Dr. Egeric Dragana, Gocanin Jasmina, Dr. Hamzagic Naida, Prof. Dr. Igić Marija, Dr. Ikodinovic Natasa, Dr. Ivosevic Jelena, Dr. Javor Milica, Dr. Jeremic Marko, Prof. Dr. Kanjevac Tatjana, Dr. Koncar Sanda, Kreculj Ksenija, Kruškonja Lidija, Dr. Labovic Nebojša, Dr. Lapcevic Jasmina, Dr. Ljajic Jasera, Dr. Ljiljana Bozovic, Dr. Maksimovic Ivana, Dr. Madic Katarina, Dr. Marjanovic Eleonora, Dr. Mijatovic Sanja, Dr. Milenkovic Sonja, Dr. Milic Marija, Prof. Dr. Milojkovic Zoraida, Dr. Paunovic Mirjana, Dr. Perovic Marina, Petkovic Tijana, Prof. Dr. Petrovic Bojan, Dr. Petrovic Dusan, Dr. Radjen Vanja, Dr. Sekula Lovrov Violeta, Dr. Skrijelj Maida, Dr. Stevanovic Marko, Dr. Stevic Dušan, Dr. Svetlana Milic, Dr. Takić Biljana, Dr. Untenecker Katarina. The authors sincerely acknowledge hard work of Drs Valentina Stojic, and Marko Aksentijevic, Lana Milenkovic and Jovana Tajejvic (who were 5<sup>th</sup>-year students of the School of Dental Medicine, University of Belgrade, and helped during piloting, collecting all data and database entries).

**Number of Tables:** 2

**Number of Figures:** 1

**Word count:** 4683

**Statements***Statement of Ethics*

As written in Methods section, the parents of all children that participated as subjects in the calibration sessions provided written informed consent before the survey. Furthermore, the study was approved by the Ethical Committee of the University of Belgrade, School of Dental Medicine (36/10) and the Government of the Republic of Serbia.

*Conflict of Interest Statement*

The authors have no conflicts of interest to declare.

*Funding Sources*

The authors declare that no funding sources were received for the study.

*Data Availability Statement*

Data are available online in institutional digital repository of the University of Belgrade, School of Dental Medicine: <https://smile.stomf.bg.ac.rs/>.

## Multilevel Calibration procedure for the oral health national multicentre survey in primary teeth.

### Abstract

*Background:* Early childhood caries (ECC) requires systematically collected and standardized data.

*Aim:* To describe a novel multilevel calibration procedure in primary dentition.

*Design:* Calibration method involved two calibration levels: the first ( $L_1$ ) involved inter-examiner agreement between three main investigators, the group leaders (GLs) in the following level; the second level ( $L_2$ ) involved three groups of 11 pediatric dentists and inter-examiner agreement assessment according to the GLs in each group. The study sample consisted of 650 primary teeth surfaces in eight children (mean age  $6.56 \pm 2.22$  years). Surface-by-surface percent agreement, tooth-by-tooth percentage agreement, Cohen's kappa and Fleiss Kappa statistics were used to calculate inter-examiner reliability. Statistical analysis was performed with SPSS 27.0.

*Results:* Surface-by-surface agreement regarding ICDAS<sub>epi</sub>-merged revealed almost perfect agreement (>90.00%) both on  $L_1$  and  $L_2$ . Kappa values and ranges showed good agreement both at  $L_1$  (Overall Kappa=0.95) and  $L_2$  (Overall Kappa=0.98) and almost perfect consistency was detected between GLs at  $L_1$  (>91.30%) and substantial agreement at  $L_2$  (>85.00%). All examiner at  $L_2$  showed almost perfect positive agreement (sensitivity=96.77-100%) when detecting presence of dental plaque.

*Conclusions:* The calibration procedure appeared feasible prior to organizing multicenter epidemiological oral health survey in large population groups of preschool children, with higher number of examiners.

**Keywords:** Early Childhood Caries, Calibration, Reliability, Epidemiology.

## Introduction

Early childhood caries (ECC) prevention strategies require assessing prevalence and risks of the disease<sup>1</sup>, as clear insight into prevalence and risk assessment data enables adequate upstream approaches and appropriate preventive interventions designed according to populational needs<sup>2,3</sup>. Subsequently, more available, systematically collected and methodologically standardized ECC data are needed globally<sup>2</sup>.

World Health Organization (WHO) Basic Method for oral health surveys<sup>4</sup> and International Caries Detection and Assessment System (ICDAS)<sup>5</sup> are standardized data collection methods that are recommended and frequently used in dental public health and epidemiological surveys. Applying these guidelines allows avoiding biases when scoring oral disease and producing valid results. However, if the study design involves more than one examiner, careful method protocol needs to ensure that all examiners agree with each other regarding range of diagnostic criteria. Inconsistency between examiners when scoring oral disease usually happens due to differences in visual or tactile performance, the level of knowledge and training, fatigue, or interest in the study<sup>6</sup>. These differences increase and are more obvious if the study is multicentered and involves several examiners. Although WHO precisely described guidelines for calibration procedure<sup>4</sup>, these goals are sometimes hard to achieve in the field. Moreover, the review of the literature confirmed that data on the examiner reliability, reproducibility and consistency are poorly reported in oral health surveys<sup>7</sup>. A special challenge might represent organization of calibration and training session(s) in the survey involving several examiners performing oral health assessments in multiple sites and having toddlers and preschoolers as subjects. According to authors' knowledge no calibration procedure involving clinical examination of patients with primary teeth prior to oral health epidemiological multicenter survey has been described and published in English language.

The aim of this study was to validate and analyze feasibility and success of multilevel calibration method for primary dentition that was used for the first time in the national epidemiological multicenter oral health survey. This study described calibration procedure and the proposed method before the epidemiological oral health survey in the field setting.

## Method

The parents of all children that participated as subjects in the calibration sessions provided written informed consent before the survey. Furthermore, the study was approved by the Ethical Committee of the University of Belgrade (36/10, date of issue: June 19<sup>th</sup>, 2019) and the Government of the Republic of Serbia. The calibration sessions were organized during August and September 2019.

### *Epidemiological oral health survey design, setting and sampling*

The WHO guideline for oral health surveys suggested at least 12 randomly chosen sites in seven locations to obtain a representative sample: two urban and two suburban sites in the capital, two urban and two suburban sites in two large towns and four sites in four rural areas<sup>4</sup>. According to the National Law on the territorial organization, the Serbia is divided into four statistical territorial units corresponding to Nomenclature of territorial units for statistics (NUTS) level 2 (regions) of the *European Union principles and methodology*<sup>8,9</sup>. Moreover, according to the same law from 2007, the official division on settlement types in Serbia include “urban” and “other types of settlements”. Also, among total of 28 cities in the country, only six (Belgrade, Novi Sad, Nis, Pozarevac, Vranje, Uzice) are divided into urban and suburban areas<sup>10</sup>.

Having in mind all these local characteristics, the investigators locally adapted sampling in a way to obtain nationally representative sample, and used stratified sampling design: **1.** In the first step the cities that have both urban and peri-urban areas and other settlements were randomly selected; the sample frame was stratified into four NUTS level 2 regions of the country. **2.** Further stage of random sampling involved list of nurseries and kindergartens within the selected cities and settlements. The sample frame was stratified into urban, suburban, and other (usually considered rural) areas. **3.** In the final stage of the selection, the participants from each nursery/kindergarten were randomly chosen by selecting children’s groups within kindergartens at each site. The sample frame was stratified into age groups as Serbian nurseries involve children one to three years of age, and children attending kindergartens are divided into the groups depending on age: kindergarten groups (3-6 years old), and preschoolers (6-7 years old)<sup>11</sup>.

Moreover, the survey setting involved more than needed 12 randomly chosen locations in four Serbian NUTS 2 regions – aiming more locations than needed would enable adequate sample size. Finally, oral health survey was designed to involve 21 randomly chosen locations. Other than three main investigators who are University teachers, there were 33 licensed pediatric dentists having at least 5-year experience and working in dental offices in the national primary health care centers in chosen locations who participated in the survey. All the examiners were recruited after definite location confirmation and agreement of local authorities. They were contacted and asked to be involved as examiners, so they also participated in the calibration process.

### *Training of the future examiners*

Training sessions were held in Belgrade, during three workshops, and all 36 of team members attended. Since it is expected for the examiners to change their diagnostic criteria during time, the first two training workshops were done before calibration sessions and before the survey started, but the third one was organized a couple of weeks after the beginning of the survey.

Training workshops involved six hours of didactical education. The most experienced researcher and the project leader was set as the benchmark examiner, he/she provided lectures on diagnostic criteria and caries clinical measurements. Moreover, training in small groups involved video and photo case presentations, case analysis, and panel discussions. During case presentations the whole team was divided into groups – each group presented and discussed with others cases of children having healthy teeth and different of the disease. During panel discussion the whole team reached the agreement on diagnostic criteria and study design. All theoretical discussions were guided so the examiners would have enough confidence to solve any issue in the field according to the diagnostic criteria and caries clinical measurements set by benchmark examiner<sup>6</sup>. This part of the training was specially considered important since all questions about coding system and scoring specific conditions were solved before going out in the field. The written material on intraoral examination protocol, diagnostic criteria and codes was sent to each team member after the second training session and 4 weeks before calibration session so they could have enough time to prepare.

The team members agreed that the most recommended and feasible way to perform intraoral examinations in children with primary teeth would be in the field (nurseries and kindergartens). Therefore, intraoral examination of children with primary dentition during calibration sessions were not carried out in the dental office, but in the separate room using the chair, good natural light, dental mirror, and gauze for drying the teeth surfaces (dental probe was not used). Also, radiographs were not used to assess the caries experience. Trained examiners performed clinical examinations after supervised brushing the children's teeth with fluoridated toothpaste. Assessment of oral hygiene was done before brushing the teeth<sup>12</sup>.

#### *Calibration procedure*

The calibration procedure was coordinated by the WHO Collaborating Center for Epidemiology and Community Dentistry. Multilevel calibration procedure involved two calibration levels:

1. The first level ( $L_1$ ) of calibration involved agreement assessment between three main investigators who are University teachers and pediatric dentists with at least 5-year experience in the field of

pediatric dentistry (Fig. 1). They were assigned as group leaders (GLs) in the next level of calibration. The assessed agreement between all three GLs was presented as overall value while the range of Kappa values between GLs pairs was presented as the range in brackets.

2. Second level ( $L_2$ ) involved dividing all examiners into three main groups of 11 pediatric dentists according to their location. In each group of 11 examiners, calibration session was performed, and agreement assessment was calculated between each examiner in the group and one of the GLs who was responsible for that group. Also, overall value involved presenting overall agreement in each group between all raters, while the range of Kappa values between GL and each examiner within group was presented as the range in brackets. Total  $L_2$  value presented agreement between three groups at  $L_2$ .

The study sample consisted of 650 primary teeth surfaces in eight pre-selected children aged four to ten years (Range: 4.3 to 9.9 years) attending public kindergartens and schools in Serbia. Mean age of children with full primary dentition was  $4.6 \pm 0.3$  years, and mean age of children with mixed dentition was  $8.5 \pm 1.2$  years. Examined children had total 130 present primary teeth, nine primary teeth were extracted due to caries complications, and 21 primary teeth were not present due to exfoliation or trauma. Children were chosen by the main investigators using convenience sampling technique, recruiting subjects from the pool of patients who already attended regular dental checkup and presented with full range of dental conditions that will be registered during calibration sessions. During calibration process at  $L_1$  all 3 GLs clinically examined 180 primary tooth surfaces in two children presenting with 36 teeth (100 primary teeth surfaces in one child with full primary dentition and 80 primary teeth surfaces in one child with mixed dentition). Total of 470 primary teeth surfaces in six children presenting with 94 primary teeth were clinically examined at  $L_2$  in the following manner: 1. Group 1 members involving examiners number 4 to 14 and GL1 examined 160 primary teeth surfaces in 2 children presenting with 32 primary teeth (20 primary teeth in one child with full primary dentition and 12 primary teeth in one child with mixed dentition); Group 2 members involved examiners number 15 to 25 and GL2 – they examined 160 primary teeth surfaces in two children presenting with 32 primary teeth (19 primary teeth in one child with full primary dentition and 13 primary teeth in one child with mixed dentition); and Group 3 members were examiners number 26 to 36 and GL3 – they examined 150 primary teeth surfaces in two children with 30 primary teeth (20 primary teeth in one child with full primary dentition and 10 primary teeth in one child with mixed dentition). Each primary tooth was examined 3 times at  $L_1$  (by each GL), and 12 times in each group at  $L_2$  (by the GL and 11 examiners in each group).



In order to assure stability of the results one of GLs from L<sub>1</sub> and one of examiners from each group from L<sub>2</sub> performed test-retest on the same patient a week after initial calibration.

#### *Diagnostic criteria and variables*

The modified Oral Health Assessment Form for Children was used as dental record<sup>4</sup> during calibration sessions. Trained examiners performed following assessments: ICDASepi-merged<sup>13,14</sup>, fillings, missing teeth due to caries complications<sup>4</sup> and the level of oral hygiene using the modified and simplified Silness & Loe plaque index (sPI) adapted to field settings (0 – no plaque, 1 – thin plaque, and 2 – thick plaque)<sup>12</sup>. In order to analyze the agreement between examiners we transformed PI into binomial variable: plaque absent, and plaque present.

Clinical examination involved registering different stages of caries disease using ICDAS system with merged codes adapted to epidemiological field settings without air drying including only lesions that could be clinically visible – ICDASepi-merged<sup>12</sup>. Caries experience including missing and filled primary teeth surfaces was registered using standard method – dmfs, while “d” part included ICDASepi-merged system (depiMEmf)<sup>12,15</sup>. In order to analyze the agreement between examiners we transformed ICDASepi-merged into binomial variable: agreement achieved, and agreement not achieved. Achievement of agreement was defined as identical diagnosis made by the examiner as the one made by the benchmark.

#### Statistical analysis

Statistical analysis was performed with SPSS Version 27.0 statistic software package. Surface-by-surface percent agreement, tooth-by-tooth percentage agreement, Cohen’s kappa and Fleiss’ Kappa statistics were used to calculate inter-examiner reliability. Sensitivity (Sn) was calculated using percentage agreement regarding presence of the disease between benchmark and other raters at L<sub>1</sub> and between GLs and other examiners at L<sub>2</sub> within each group. Specificity (Sp) was calculated as a percentage agreement between previously mentioned raters at both levels regarding presence of healthy teeth. Both Sn and Sp were presented as range between each group examiner vs. group leader. Percent of agreement between rater pairs, as well as percent of full agreement between all 3 raters was assessed at L<sub>1</sub>. More detailed analysis at L<sub>2</sub> involved



examining the range of Kappa and percent of agreement between GLs and examiners in each group, as well as full agreement between all 3 groups presented as Total in Tables 1 and 2.

The strength of Kappa agreement was determined as follows: <0.51 slight, 0.51–0.60 fair, 0.61–0.70 acceptable, 0.71–0.80 moderate, 0.81–0.90 substantial, >0.90 almost perfect<sup>6</sup>.

## Results

Each of main investigators in  $L_1$  examined 180 primary teeth surfaces (100 primary teeth surfaces in child with full primary dentition and 80 primary teeth surfaces in child with mixed dentition) assessing  $d_{epiME}mf$ , and the level of oral hygiene. Caries experience at  $L_1$  of the examined children with primary and mixed dentition was  $dmfs=9.67\pm 0.57$  (range=9-10), and  $dmfs=11.67\pm 0.57$  (range=11-12), respectively.

Examiners in  $L_2$  totally examined 470 primary teeth surfaces: *Group 1*: 100 primary teeth surfaces in child with full primary dentition and 60 primary teeth surfaces in child with mixed dentition ( $dmfs=11.50\pm 0.80$  (range=10-12), and  $dmfs=12.83\pm 0.39$  (range=12-13), respectively); *Group 2*: 95 primary teeth surfaces in child with full primary dentition and 65 primary teeth surfaces in child with mixed dentition ( $dmfs=19.83\pm 0.39$  (range=19-20), and  $dmfs=12.17\pm 0.84$  (range=11-13), respectively); *Group 3*: 100 primary teeth surfaces in child with full primary dentition and 50 primary teeth surfaces in child with mixed dentition ( $dmfs=10.92\pm 0.29$  (range=10-11), and  $dmfs=66.0$  (range=66-66), respectively).

In total, during 4 calibration sessions 650 primary teeth surfaces were examined, and observed average prevalence of caries on primary teeth surfaces was 25.8%.

Table 1 reveals inter examiner surface-by-surface agreement regarding dental health. The obtained results revealed almost perfect agreement (>93.75%) on both  $L_1$  and  $L_2$ . Kappa ranges showed better agreement at  $L_1$  only between 3 GLs (Kappa Range=0.92-0.98) compared to agreement between examiners and group leaders at  $L_2$  in each group (G1 Kappa Range=0.91-1, G2 Kappa Range=0.85-0.98, and G3 Kappa Range=0.95-1.00). Also, the obtained sensitivity range, revealed almost perfect consistency between GLs at  $L_1$  (>91.30%) and substantial agreement between raters and groups at  $L_2$  (>82.60%) regarding detecting the disease (Table 1). Calibration

process at both levels, revealed perfect agreement between all raters when detecting healthy primary teeth (specificity) (Table 1). The distribution of the of ICDAS<sub>epi</sub> merged scores according to scores according to group leaders (GLs) at Level1 (L1) is presented in the supplementary file (Table 1S page 2).

Further analysis presented in Table 2 revealed inter-examiner tooth-by-tooth agreement regarding sPI. The distribution of the of Distribution of simplified plaque index (sPI) scores according to group leaders (GLs) at Level1 (L1) is displayed in the supplementary file (Table 2S page 6). The level of agreement between GLs at L<sub>1</sub> revealed perfect agreement and 100% sensitivity and specificity. However, when assessing other examiners at L<sub>2</sub>, analysis showed substantial level of agreement (Kappa=0.89) in the second L<sub>2</sub> group. The distribution of the of ICDAS<sub>epi</sub> merged scores according to scores according to the different group leaders (GL1, GL2, GL3) at Level2 (L2) is presented in the supplementary file (Table 3S page 7, Table 4S page 13 and Table 5S page 19). Also, wider overall range when detecting absence of dental plaque (Specificity=60.00-100.00%) was revealed in the third L<sub>2</sub> group (Table 2). All examiner at L<sub>2</sub> showed almost perfect positive agreement (Sensitivity=96.77-100%) when detecting presence of dental plaque (Table 2). Although Kappa, sensitivity and specificity showed wider range values when assessing inter-rater agreement on PI, the median (50<sup>th</sup> percentile) strongly suggested 100% agreement both at L<sub>1</sub> and L<sub>2</sub>. The distribution of the of simplified plaque index (sPI) scores according to the different group leaders (GL1, GL2, GL3) at Level2 (L2) is presented in the supplementary file (Table 6S page 25, Table 7S page 27 and Table 8S page 29).

## Discussion

According to authors' knowledge, the calibration method described in this paper is a novel approach to achieve local adaptation of standardized recommendations. The described multilevel calibration scheme showed acceptable level of agreement and the method is more feasible to organize in multicenter surveys. This might be helpful especially in circumstances where standard calibration procedure would be complicated to organize due to recourse restraints, multiple sites, and young children as subjects.

Calibration processes that are precisely described so they could be faithfully reproduced before oral health epidemiological surveys are scarce<sup>6,16</sup>. Making reliable data collection during epidemiological studies is of utmost importance since this enables analyzing trustworthy data and providing robust results interpretation. This is especially important when performing national epidemiological oral health survey because these results precede designing of adequate prevention intervention tailored according to population needs. Calibration method should be integral part of the report and should include following necessary aspects: a detailed description of the survey method preferably standardized, information on the materials and setting used (such as probe usage and type of probe, light conditions, radiographs, and information on teeth cleaning before examination), the precise description of the threshold used for detection of lesions, and the examiner characteristics (calibration, reliability testing and reporting)<sup>7</sup>.

The WHO *“Oral Health Surveys: Basic Methods”*<sup>4</sup> suggested calibration method where each examiner should independently examine at least 20 subjects. Although, WHO method was generally the most often used in more than half of published surveys, poor adherence to these guidelines was observed regarding the measurement and the reliability of measurement<sup>7</sup>. In the National Oral Health Survey (NOHS), the most feasible way to organize calibration session with preschool children as subjects, was to lower down the number of subjects yet trying not to affect the quality of agreement.

It was decided to modify the calibration method to multilevel procedure with lower number of the patients yet providing enough cases, teeth surfaces to exam and calculations to involve all stages of oral disease that needed to be diagnosed. All patients for the calibration sessions were pre-selected to represent the full range of conditions expected to be assessed in the present survey. Moreover, the patients that were preselected for the calibration sessions were conveniently chosen from the pool of cooperative patients who already attended regular dental checkups in the same center were each session was organized, so the attending primary health care pediatric dentist was able to prepare both the child and the parents in advance. In the case of the survey in our country, having in mind that 36 examiners were involved, according to the WHO calibration method it would be necessary to perform 720 examinations in total if each participant would perform intraoral exam in each of proposed at least 20 subjects. This is why

dividing examiners in two levels, and then in three groups in level 2 and having 12 examinations in each child seemed more feasible. Especially having in mind that besides lowering the number of subjects to four children with full primary dentition and four children with mixed dentition (two children at  $L_1$  and six children at  $L_2$ ), the total number of surfaces that were examined ( $n=650$ ) involving full range of disease stages enables success.

The calibration procedure as well as the future SNOHS was designed to use simplified ICDAS<sub>epi</sub>-merged<sup>12,17</sup>. Caries experience including missing and filled primary teeth surfaces was registered using standard method – dmfs, while “d” part included ICDAS<sub>epi</sub>-merged system (d<sub>epi</sub>MEmf)<sup>12,15</sup>. According to the literature, the inspection and visual evaluation of lesion size and extent showed good overall performance especially when validated and recognized scoring system was used<sup>18</sup>. Moreover, visual assessment of tooth surface by observing lesion location, plaque accumulation, roughness or radiance might be used to develop decisions regarding lesion activity<sup>19</sup>.

Although the results of the present survey showed successful level of agreement, it must be emphasized that lower agreement in  $L_2$ , wider range of percentages in  $L_2$  compared to  $L_1$  could be explained with the higher number of examiners. The wider range of agreement for negative response for plaque index scoring, might be linked to the negative result observed only in 3 examiners, indicating that plaque index scores might be over diagnosed. Besides, the use of probe in a dental office setting with air drying and artificial light would tackle these issues. Moreover, as the examiners participation was at voluntary bases, this might be considered as a systematic bias between raters as all the examiners were pediatric dentists in primary public health care centers. The efforts to address potential sources of bias in the present survey involved randomly chosen locations and cites and afterwards seeking examiners participation.

In order to have the most possible extensive coverage and reliable results, the examiners were instructed to supervise toothbrushing and dry the primary teeth surface with gauze and note if any non-cavitated lesions were observed during the survey.

The present survey implemented previously described checklist of methods’ aspects to be included in reports of surveys assessing caries experience, suggesting the most important strength of the present study. Further strengths involve that primary health care dentists who

are responsible for the everyday preventive and surgical dental care in the field were involved in the calibration procedure, they were acknowledged and considered important part of the team. Although presented results showed novel approach and methodological strength, there are some limitations that must be considered. Although enough cases, examined surfaces and calculations ensured range of diagnostic decisions, higher number of subjects in the calibration session indicates more successful training of the examiners. According to the literature review didactic and training sessions improve homogeneity of the results, especially if there are two calibration sessions<sup>6</sup>. But, the aim of this survey was to present more simple way to perform calibration, so the authors and examiners did not conduct oral health assessments in children before and after training sessions. On the other hand, future, more detailed research could be designed to include these assessments. Also, the fact that different patients were examined in  $L_1$  compared to  $L_2$  might be considered as limitation of the study. But, the purpose of the study was to assess agreement between teachers in  $L_1$ , and to compare this with the agreement of each teacher with his group of examiners in  $L_2$ . The possible limitation of the study might be the field setting and decision not to use dental probe, potentially resulting in underscoring smooth white spots or even enamel breakdowns and missing light dental plaque. However, this protocol was favored as the most feasible for epidemiological oral health survey involving nationally representative sample of preschool children. Also having in mind that the subjects that were used in calibration session were conveniently preselected cooperative children, it is possible that the subjects in different culture and/or in the survey itself might be less cooperative. However, this is why field setting during the survey is so important since the nursery and kindergarten represent already familiar setting to the children and the presence of daycare teachers would be crucial. Moreover, it would be advisable, as the authors previously described in the Method section, to aim more locations and more subjects than needed so adequate sample size could be achieved in case of drop off. Repeating this study method in different regions of the world and in different cultures would be helpful in terms of comparing results. Finally, the examiners were trained to record the activity of the lesions during training sessions; however, they were not calibrated on the activity of the lesions. The consensus regarding lesion activities was found and most of the examiners

agreed during training sessions and case study discussion. However, calibrating the examiners for lesion activity might be considered for future surveys.

Calibration studies “correct for biases, thus allowing the different sources of information to be brought together”<sup>20</sup>. This study focuses on the main issue arising when designing oral health multicenter epidemiological study – organizing difficulties regarding calibration procedures that lead to poor adherence to already existing standardized guidelines on the measurement and the reliability of the measurement. The main reason for the success of presented protocol was the possibility to find compromises when organizing calibration sessions resulting in adaptation of the process to local circumstances. More studies using the same method in different cultural settings having multiple sites are needed to confirm feasibility and soundness of the proposed calibration protocol.

Agreement between  $L_1$  raters (teachers) and teachers and their examiners in  $L_2$  as well as raters between both levels was satisfactory. The proposed calibration procedure appeared feasible prior to organizing multicenter epidemiological oral health survey in large population groups of preschool children, with higher number of examiners. More research is needed using this method in different regions.

### Why this paper is important to paediatric dentists

- In planning a survey, calibration of the examiners is mandatory and this paper provide a new methodology for multicenter survey.
- This paper aims to raise the epidemiological and dental public health knowledge and awareness among pediatric dentists.
- Gathering relevant data on ECC is of utmost importance before planning preventive intervention, and this paper provides reproducible method and calibration procedure specially for lower recourse settings.

### References

1. Hirsch GB, Edelstein BL, Frosh M, Anselmo T. A simulation model for designing effective interventions in early childhood caries. *Prev Chronic Dis*. 2012;9:E66. doi:10.5888/pcd9.110219
2. el Tantawi M, Folayan MO, Mehaina M, et al. Prevalence and data availability of early childhood caries in 193 united nations countries, 2007–2017. *Am J Public Health*. 2018;108(8). doi:10.2105/AJPH.2018.304466
3. Markovic D, Soldatovic I, Vukovic R, Peric T, Campus GG, Vukovic A. How Much Country Economy Influences ECC Profile in Serbian Children—A Macro-Level Factor Analysis. *Front Public Health*. 2019;7. doi:10.3389/fpubh.2019.00285
4. Petersen PE, Baez RJ. *Oral Health Surveys: Basic Methods*. 5th ed. World Health Organization; 2013.
5. Pitts N. “ICDAS”—an international system for caries detection and assessment being developed to facilitate caries epidemiology, research and appropriate clinical management. *Community Dent Health*. 2004;21(3):193-198.
6. Castiglia P, Campus G, Solinas G, Maida C, Strohmenger L. Children’s Oral Health in Italy: Training and Clinical Calibration of Examiners for the National Pathfinder about Caries Disease. *Oral Health Prev Dent*. 2007;5(4):255-261.
7. Agbaje JO, Lesaffre E, Declerck D. Assessment of caries experience in epidemiological surveys: A review. *Community Dent Health*. 2012;29(1):14-20. doi:10.1922/CDH\_2668-Olubanwo06
8. Gavrilovic D. Districts and region in the Republic of Serbia [document in Serbian]. Serbia SO of the R of, ed. Published online 2018. <http://publikacije.stat.gov.rs/G2018/Pdf/G201813045.pdf>
9. Eurostat. *Statistical Regions in the European Union and Partner Countries: NUTS and Statistical Regions 2021*. Publications Office of the European Union; 2020. Accessed October 25, 2022. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-20-092>
10. The National Assembly of Serbia. *The Law on Territorial Organization of the Republic of Serbia*. The Official Gazette of the Republic of Serbia. 2020;32(09):1-3. Accessed October 25, 2022.



- <http://www.pravno-informacioni-sistem.rs/SlGlasnikPortal/eli/rep/sgrs/skupstina/zakon/2007/129/1/reg>
11. Vujic S, Baronijan H, Baucal A. Investing in Early Childhood Education in Serbia. UNICEF; 2012. Accessed October 25, 2022. [https://www.unicef.org/serbia/sites/unicef.org.serbia/files/2018-08/Booklet\\_Investing\\_in\\_Early\\_Childhood\\_Education\\_in\\_Serbia.pdf](https://www.unicef.org/serbia/sites/unicef.org.serbia/files/2018-08/Booklet_Investing_in_Early_Childhood_Education_in_Serbia.pdf)
  12. Martignon S, Cortes A, Gómez SI, et al. How Long does it Take to Examine Young Children with the Caries ICDAS System and how do they Respond? *Braz Dent J.* 2018;29(4):374-380. doi:10.1590/0103-6440201801949
  13. Martignon S, Cortes A, Gamboa LF, et al. Effectiveness of the ICCMS caries management system for children: a 3-year multicentre randomised controlled trial. *Acta Odontol Scand.* Published online February 15, 2022:1-12. doi:10.1080/00016357.2022.2038263
  14. Evans RW, Feldens CA, Phantunvanit P. A protocol for early childhood caries diagnosis and risk assessment. *Community Dent Oral Epidemiol.* 2018;46(5):518-525. doi:10.1111/cdoe.12405
  15. Pitts NB, Ekstrand KR, ICDAS Foundation. International Caries Detection and Assessment System (ICDAS) and its International Caries Classification and Management System (ICCMS) - methods for staging of the caries process and enabling dentists to manage caries. *Community Dent Oral Epidemiol.* 2013;41(1):e41-52. doi:10.1111/cdoe.12025
  16. Lesaffre E, Mwalili SM, Declerck D. Analysis of Caries Experience Taking Inter-observer Bias and Variability into Account. *J Dent Res.* 2004;83(12):951-955. doi:10.1177/154405910408301212
  17. Evans RW, Feldens CA, Phantunvanit P. A protocol for early childhood caries diagnosis and risk assessment. *Community Dent Oral Epidemiol.* 2018;46(5):518-525. doi:10.1111/cdoe.12405
  18. Gimenez T, Piovesan C, Braga MM, et al. Visual Inspection for Caries Detection. *J Dent Res.* 2015;94(7):895-904. doi:10.1177/0022034515586763
  19. Kidd E. The implications of the new paradigm of dental caries. *J Dent.* 2011;39:S3-S8. doi:10.1016/j.jdent.2011.11.004
  20. Plummer M, Clayton D, Kaaks R. Calibration in multi-centre cohort studies. *Int J Epidemiol.* 1994;23(2):419-426. doi:10.1093/ije/23.2.419

## Legends

**Fig. 1.**

Schematic design of multilevel calibration process.

Supplementary material

**Table 1.** Agreement between examiners at Level1 (L<sub>1</sub>) and Level2 (L<sub>2</sub>) regarding ICDASePi-merged scores.

Observations	Number of surfaces	Measurements	At surface status
<i>(inter-examiner agreement assessment between three main investigators) L<sub>1</sub></i>			
G1, G2, G3 Leaders (2 patients examined – 36 teeth)	180	Agreement	177/180 (98.33%)
		G1 Leader vs G2 Leader	Sn=95.65% / Sp=100.00%
		G1 Leader vs G3 Leader	Sn=91.30% / Sp=100.00%
		Kappa Overall (range)	0.95 (0.92-0.98) <sup>a</sup>
<i>(inter-examiner agreement assessment between all examiners) L<sub>2</sub></i>			
G1 Leader with his group (Examiners number 4-14 – 2 patients – 32 teeth)	160	Agreement	156/160 (97.50%)
		Sn	85.00-100.00% <sup>b</sup>
		Sp	100.00-100.00% <sup>b</sup>
		Kappa Overall (range)	0.96 (0.91-1.00) <sup>a</sup>
G2 Leader with his group (Examiners number 15-25 – 2 patients – 32 teeth)	160	Agreement	150/160 (93.75%)
		Sn	82.60-95.65% <sup>b</sup>
		Sp	100.00-100.00% <sup>b</sup>
		Kappa Overall (range)	0.91 (0.85-0.98) <sup>a</sup>
G3 Leader with his group	150	Agreement	144/150 (96.00%)

<i>(Examiners number 26-36 – 2 patients – 30 teeth)</i>		Sn	91.48-100.00% <sup>b</sup>
		Sp	100.00-100.00 <sup>b</sup>
		Kappa Overall (range)	0.98 (0.95-1.00) <sup>a</sup>
<b>Total inter-examiner agreement assessment between all examiners L<sub>2</sub></b>		Agreement	450/470 (95.74%)
		Sn	95.55-100.00% <sup>b</sup>
		Sp	100.00-100.00% <sup>b</sup>
		Kappa Overall (range)	0.98 (0.97-1.00) <sup>a</sup>

G1 = group 1, G2 = group 2, G3 = group 3

Agreement – number of surfaces with the same diagnosis divided by total number of surfaces examined

Sn – Sensitivity; Sp – Specificity

<sup>a</sup>Overall Kappa within whole group including group leader and all examiners (in the brackets: range of Kappa values between benchmark and other group leaders at L<sub>1</sub> and/or between each group examiner vs. group leader at L<sub>2</sub>)

<sup>b</sup>Sn/Sp range between each group examiner vs. group leader

**Table 2.** Agreement between examiners at Level1 (L<sub>1</sub>) and Level2 (L<sub>2</sub>) for simplified plaque index (sPI) scores.

Observations	Number of teeth (n)	Kappa overall (range)	Agreement at surface/status (%)	Sp (%)	Sn (%)
<i>(inter-examiner agreement assessment between three main investigators) L<sub>1</sub></i>					
G1, G2, G3 Leaders	36	1.0 (1.0-1.0) <sup>a</sup>	36/36 (100.00%)	100.00- 100.00% <sup>b</sup>	100.00- 100.00% <sup>b</sup>
<i>(inter-examiner agreement assessment between all examiners) L<sub>2</sub></i>					
G1 Leader with his group ( <i>Examiners number 4-14 – 2 patients – 32 teeth</i> )	32	0.92 (0.90-1.00) <sup>a</sup>	31/32 (96.87%)	85.71- 100.00% <sup>b</sup>	100.00- 100.00% <sup>b</sup>
G2 Leader with his group ( <i>Examiners number 15-25 – 2 patients – 32 teeth</i> )	32	0.89 (0.81-1.00) <sup>a</sup>	27/32 (84.37%)	80.00- 100.00% <sup>b</sup>	88.00- 100.00% <sup>b</sup>
G3 Leader with his group ( <i>Examiners number 26-36 – 2 patients – 30 teeth</i> )	30	0.92 (0.67-1.00) <sup>a</sup>	24/30 (80.00%)	60.00- 100.00% <sup>b</sup>	100.00- 100.00% <sup>b</sup>

**Total inter-examiner agreement  
assessment between all examiners**

94

0.92 (0.80-1.00)<sup>a</sup>

82/94 (87.23%)

75.00-  
100.00%<sup>b</sup>

96.77-  
100.00%<sup>b</sup>

**L<sub>2</sub>**

---

*Agreement – number of surfaces with the same diagnosis divided by total number of surfaces examined*

*Sn – Sensitivity; Sp – Specificity*

<sup>a</sup>Overall Kappa within whole group including group leader and all examiners (in the brackets: range of Kappa values between benchmark and other group leaders at L<sub>1</sub> and/or between each group examiner vs. group leader at L<sub>2</sub>)

<sup>b</sup>Sn/Sp range between each group examiner vs. group leader

