ARQUIVOS EM ODONTOLOGIA

Dental caries surveys: considerations about reproducibility

Renato Pereira da Silva[1] | **Andréa Videira Assaf[2]** | **Aline Sampieri Tonello[3]** | **Gláucia Maria Bovi Ambrosano[4]** | **Silvia Helena de Carvalho Salles Peres[5]** | **Marcelo de Castro Meneghim[4]** | **Antonio Carlos Pereira[4]**

[1]Departamento de Nutrição e Saúde, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil

[2]Instituto de Saúde de Nova Friburgo, Universidade Federal Fluminense, Nova Friburgo, Rio de Janeiro, Brasil

[3]Departamento de Saúde Pública, Universidade Federal do Maranhão, São Luís, Maranhão, Brasil

[4]Departamento de Ciências da Saúde e Odontologia Infantil, Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas, Piracicaba, São Paulo, Brasil

[5]Departamento de Odontopediatria, Ortodontia e Saúde Coletiva, Faculdade de Odontologia de Bauru, Universidade de São Paulo, Bauru, São Paulo, Brasil

Aim: To evaluate the impact of the recording of teeth whose clinical conditions rarely vary (anterior teeth and third molars) in inter-examiner agreement measurements.

Methods: Clinical data from 56 schoolchildren, 12 years of age, previously collected by two examiners, according to the "Oral Health Surveys: basic methods" codes and criteria, were analyzed in the present study. The effects from including/excluding such teeth upon reproducibility were measured by general percentage agreement (GPA) and Kappa statistics (κ) performances.

Results: The exclusion of anterior teeth associated with the inclusion of third molars produced a decrease in GPA that was simultaneous to an increase in the weighted Kappa (nominal data) and simple (dichotomous data) values. The incorrect inclusion of third molars (GPA = 100%; κ = + 1) in the reproducibility measurement artificially increased the inter-examiner Kappa values.

Conclusion: The inclusion/exclusion of anterior teeth and third molars, seeking a more reliable agreement among examiners, can have a positive or negative impact on the measured reproducibility values. A clear warning about the impact of including third molars in the reproducibility measurement, in the 12 years old age group, should be performed in "Oral Health Surveys: basic methods" and similar manuals.

Uniterms: Dental caries. DMF index. Epidemiology. Reproducibility of results.

INTRODUCTION

Oral health surveys are important tools for planning and evaluation in public health. Reproducibility is essential in assuring reliable examinations in such surveys[1]. The general percentage of agreement (GPA) and Cohen's Kappa (k) statistics are the most frequently used tests in oral health surveys[1,2]. Cohen's Kappa statistics are calculated from square matrices 2x2, 3x3, 4x4, and so forth, and may be influenced by current development patterns of dental caries in samples[3-7]. Therefore, the prevalence of dental caries in oral health surveys deserves special attention, and a truly representative sample for reproducibility studies must cover the entire spectrum of dental caries[1,7].

In training and calibration processes, all teeth are examined. Their clinical conditions are registered in specific children or adult forms from the "Oral Health Surveys: basic methods"[1] or similar documents. However, teeth whose

clinical conditions rarely produce inter-examiner divergences, such as anterior teeth (due to a decreasing prevalence and polarization of dental caries)[8-11] and third molars (which very often receive code "8", according World Health Organization (WHO) diagnostic criteria[1], since they can remain impacted and asymptomatic throughout life)[12-14], may contribute to an overestimation of the intra- and inter-examiner agreement in epidemiological surveys. Primarily in prevalence populations of low dental caries (DMF-T below 2.6 for 12-year-old children and DMF-T below 8.9 for adults aged 35-44 years)[1], such a situation may produce biased results for both the reproducibility and validity of diagnostic tests in such surveys.

Therefore, this study sought to evaluate the impact of the recording of the clinical conditions of anterior teeth and third molars upon the performance of GPA and Kappa statistics, according to WHO diagnostic criteria[1], within an epidemiological setting.

MATERIAL AND METHODS

ETHICAL ASPECTS

This study was approved by the Research Ethics Committee from the Bauru Dental School of the University of São Paulo (USP) (Protocol #40/2006), Brazil. It was conducted according to Resolution 196/96 from National Health Council of the Brazilian Ministry of Health.

SAMPLE AND DENTAL CARIES EXAMINATION

The underlying premise of the present study stemmed from the first calibration session of the dental caries survey conducted in the Tagliaferro et al.[15] study. Regarding the calibration time, training and calibration sessions lasted 28 hours and were conducted by an expert examiner: a) Theoretical discussion about the applied variables, codes, and examination criteria: 4 hours; b) Practical discussion: 8 hours; c) Calibration process itself: 14 hours (2 sessions with an interval of 1 week); d) Final calibration discussion: 2 hours. In this session, the examiners asked the following questions: a) How to proceed in relation to teeth 18, 28, 38, and 48, at 12-years-old, given that all of the third molars were coded as "8" during the Kappa calculation? b) The anterior teeth in this age group are generally firm. Do they need to be registered during the calibration process?

After such questions, the researchers sought to develop the present study, based on the use of data produced during the calibration process from the Tagliaferro et al.[15] study, seeking to assess inter-examiner agreement results, considering different simulations including/excluding anterior teeth and third molars. The calibration process methodology from Tagliaferro et al.[15] followed the WHO[1,16] recommendations.

WHO[1,16] defined that at least 25 individuals must be examined for reproducibility studies. In the present study, previous clinical data from 56 randomly selected 12-year-old schoolchildren, presenting a large spectrum of the prevalence of dental caries and belonging to 05 public schools of the municipality of Bauru, in the state of São Paulo, Brazil, were analyzed. Thus, a total of 672 anterior (upper and lower incisors and canine teeth) and 224 upper and lower third molars were included in this data set.

All WHO recommendations regarding the environmental and organizational conditions for the examination were followed. Examinations were always performed under natural light, using the tactile visual method, with a flat #5 mouth mirror, CPI/ball point probes, and prior tooth brushing. In addition, codes/criteria of clinical dental conditions were recorded in WHO caries forms by previously trained recoding staff[1,16].

DATA ANALYSIS

The reproducibility analysis focused on inter-examiner agreement measured by GPA and Kappa statistics. Although data dichotomization is inappropriate for reproducibility studies[3,6,17,18], this procedure was exclusively performed in an attempt to understand the impact of inclusion/exclusion of anterior teeth and third molars upon the reproducibility values. Therefore, the simple Kappa was used for dichotomous data, while the weighted Kappa was used for non-dichotomous data (original data). For data dichotomization, the WHO codes[1] ("1. Caries", "2. Filled, with caries", and "4. Missing due to caries") were pooled as "Code 1: CARIES", while the other WHO codes ("0. Sound", "3. Filled, no caries", "5. Missing for any other reason", "6. Fissure sealant", "7. Fixed dental prothesis abutment, special crown, or venner/implant", "8. Unerupted tooth (crown)/unexposed root", and "9. Not recorded") were pooled as "Code 0: NO CARIES". All WHO codes (not only those used to compose the DMF-T index) are very important to reproducibility analyses in studies of dental caries[1,16].

For statistical purposes, a GPA value of "100%", associated with a variable Kappa value of "0", for agreement expected by chance, or of "+1", for perfect agreement, was assumed for the third molars. The inter-examiner agreement was evaluated "tooth by tooth" by GPA and Kappa. The GPA and Kappa values found for each tooth were used to calculate the mean values of the GPA and Kappa statistics.

The One Sample T-test ($\alpha = 0.05$), assuming that the sample came from an approximately normal distribution, was performed to compare GPA and Kappa means. Data analysis was performed in the IBM® SPSS® Statistics.

RESULTS

Approximately 98% of the examined anterior teeth (n = 660) were caries-free. The inter-examiner agreement, including all teeth (except third molar teeth), showed GPA = 87.07% and Kappa = 0.53. The other inter-examiner agreement values are simulation results for the present study purposes (Table 1).

Table 1 - GPA and Kappa mean values

	GPA (%)	p-value	Kappa	p-value
*Non-dichotomous (original data; 3rd molar not considered)**	*87.07*	*Test value*	*0.53*	*Test value*
Non-dichotomous; 3rd molar included (GPA = 100%; κ = 0)	88.68	0.509	0.35	0.005
Non-dichotomous; 3rd molar included (GPA = 100%; κ = +1)	88.68	0.509	0.69	0.005
Non-dichotomous; Anterior teeth excluded	78.64	0.020	0.56	0.531
Non-dichotomous; Anterior teeth excluded; 3rd molar included (GPA = 100%; κ = 0)	82.91	0.214	0.45	0.237
Non-dichotomous; Anterior teeth excluded; 3rd molar included (GPA = 100%; κ = +1)	82.91	0.214	0.65	0.048
Dichotomous (original data; 3rd molar not considered)	*88.91*	*Test value*	*0.46*	*Test value*
Dichotomous; 3rd molar included (GPA = 100%; κ = 0)	90.29	0.492	0.32	0.018
Dichotomous; 3rd molar included (GPA = 100%; κ = +1)	90.29	0.492	0.63	0.010
Dichotomous; Anterior teeth excluded	82.08	0.022	0.52	0.358
Dichotomous; Anterior teeth excluded; 3rd molar included (GPA = 100%; κ = 0)	85.66	0.241	0.41	0.502
Dichotomous; Anterior teeth excluded; 3rd molar included (GPA = 100%; κ = +1)	85.66	0.241	0.61	0.029

* One sample T-test, $\alpha = 0.05$.

As observed in Table 1, while the dichotomization increased the GPA values, the Kappa values decreased in all data settings. The inclusion of third molars to the original examined data setting increased the GPA for both dichotomized or non-data settings.

The exclusion of anterior teeth decreased the GPA value for both dichotomous and non-dichotomous data settings. When associated with the third molar inclusion, under a perfect agreement (κ = +1), the Kappa values are artificially increased for dichotomous and non-dichotomous data settings. Such results meet statistical significance (Table 1).

DISCUSSION

As is well-known, at 12 years of age, all permanent teeth (except third molars) have already erupted. Therefore, it seems logical to use adult forms[1] to register their clinical conditions. However, that situation may lead to

incorrect results of reproducibility examinations. Although reproducibility issues have been studied extensively, a specific impact of tooth recording that rarely varies has yet to be performed.

Reproducibility is necessary, although not sufficient; it is a condition for the validity of examination methods. Both GPA and Kappa statistics are recommended by the WHO for this very purpose[1]. The GPA is the simplest method to measure inter-examiner agreement. In general, its values are higher than Kappa values[4]. In the present study, the intra-examiner agreement analysis results were not used, since the interest outcome is easier to be checked in inter-examiner agreement measurement among different examiners. An acceptable GPA value ranges from 85-90%[4,16], similar to all GPA values found in the present study (Table 1).

Nevertheless, the WHO warns that Kappa statistics are more adequate when the prevalence of dental caries is low in the population under study[1], when conducted by measuring inter-examiner agreement beyond chance[5]. However, Kappa statistics are highly influenced by disease prevalence in the sample[5,6,17]. The prevalence of dental caries in the state of São Paulo, Brazil, is low (DMF-T index for 12-years-old = 1.90)[2]. Acceptable simple and weighted Kappa values are above 0.60 and 0.65, respectively[1,2,17].

The disagreements for the clinical condition diagnosis of anterior teeth (due to a decreasing prevalence and polarization of dental caries)[8-11] and third molars (erupted at 17-20 years of age but may remain impacted and asymptomatic throughout life)[12-14] are almost nonexistent at 12-years of age[12,17]. Anterior teeth are very important to the physiology, phonation, esthetics, and sociability of individuals. Consequently, such teeth are very important to oral health professionals and DMF-T index composition. Any change that involves these teeth will artificially modify reproducibility. If a perfect agreement (κ = +1) is considered, the Kappa value increases for all settings[18]. These results meet statistical significance (Table 1). Thus, to exclude them from the samples in training and calibration processes is unthinkable. Third molars, however, present no clinical importance to oral health services in the short and mid-term analyses[13,14].

As seen in Table 1, the inclusion of the third molars (considering κ = +1) increased the inter-examiner agreement in comparison to the original data setting, even when the anterior teeth are excluded from the analysis. That situation is likely to occur, especially among inexperienced oral health staff, since it is not established as

a caution in the "Oral Health Surveys: basic methods"[1,16]. The clinical forms presented in "Oral Health Surveys: basic methods" (4th edition)[16] are unique for children and adults. Although its 5th edition[1] presents specific forms for children and adults, a clear warning about the impact of the incorrect recording of the clinical conditions of third molars in epidemiological surveys is not made. Both calibration manuals, "SB Brasil 2010"[19] and "SB São Paulo 2015"[20], two broad Brazilian dental caries surveys, provide spreadsheets that record third molar clinical conditions so as to illustrate the reproducibility calculation method as well.

The main limitation of this study refers to its inadequate external validity. The recommendation to exclude third molars from the analysis was tested only for 12-year-old schoolchildren. The effect of this exclusion (seeking to achieve a more reliable reproducibility) must be tested in other age groups, in older individuals with erupted third molars and a higher prevalence of dental caries, including an experience of dental caries in both third molars and anterior teeth.

CONCLUSION

The presented study attempt to achieve better and more reliable reproducibility results in dental caries epidemiological surveys, through the exclusion of clinical conditions of anterior teeth and the recording of third molars. The exclusion of anterior teeth and the inclusion of third molars have a direct impact upon reproducibility values. However, neither practice is recommended, as both produced biased reproducibility outcomes. More specifically, the impact of recording third molars should be highly discouraged (and avoided) in the training sessions and at calibration stages for dental caries epidemiological surveys to be applied with 12-year-old subjects. The "Oral Health Surveys: basic methods" and similar documents should provide this warning clearly.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

ORCID

Renato Pereira da Silva https://orcid.org/0000-0003-3893-969X

Andréa Videira Assaf https://orcid.org/0000-0001-7071-7988

Aline Sampieri Tonello https://orcid.org/0000-0002-8402-9112

Gláucia Maria Bovi Ambrosano https://orcid.org/0000-0001-5950-6370

Silvia Helena de Carvalho Salles Peres https://orcid.org/0000-0003-3811-7899

Marcelo de Castro Meneghim https://orcid.org/0000-0003-2673-3627

Antonio Carlos Pereira https://orcid.org/0000-0003-1703-8171

REFERENCES

1. World Health Organization (WHO). Oral health surveys: basic methods. 5th ed. Geneva: WHO Press; 2013.

2. Pereira AC, Vieira V, Frias AC. Pesquisa estadual de saúde bucal: relatório final. Águas de São Pedro: Livronovo, 2016.

3. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measur. 1960;20(1):37-46.

4. Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-9.

5. Sim J, Wright CC. The Kappa statistics in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3):257-68.

6. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360-3.

7. Tonello AS, Silva RP, Assaf AV, Ambrosano GMB, Peres SHCS, Pereira AC, et al. Inter-examiner agreement dental caries epidemiological surveys: the importance of disease prevalence in the sample. Rev Bras Epidemiol. 2016;19(2):272-9.

8. Narvai PC, Frazão P, Roncalli AG, Antunes JLF. [Dental caries in Brazil: decline, polarization, inequality and social exclusion]. Rev Panam Salud Pública. 2006;19(6):385-93. [Portuguese].

9. Batchelor, PA, Sheiham A. Grouping of tooth surfaces by susceptibility to caries: a study in 5-16 year-old children. BMC Oral Health. 2004;4(1):1-6.

10. Oulis C, Theodorou M, Mastrogiannakis T, Mamai-Chomata E, Polychronopoulou A, Athanasouli T. Oral health status and treatment needs of the Hellenic Population - a pathfinder survey - proposals for improvement. Hellenic Stomatological Review. 2009;53:97-120.

11. Carvalho JC. Caries process on occlusal surfaces: evolving evidence and understanding. Caries Res. 2014;48(4):339-46.

12. Ahlqwist M, Gröndahl HG. Prevalence of impacted teeth and associated pathology in middle-aged and older Swedish women. Community Dent Oral Epidemiol. 1991;19(2):116-9.

13. Song F, O'Meara S, Wilson P, Golder S, Kleijnen J. The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth. Health Technol Assess. 2000;4(15):1-55.

14. Mettes TG, Nienhuijs MEL, van der Sanden WJM, Verdonschot EH, Plasschaert AJM. Interventions for treating asymptomatic impacted wisdom teeth in adolescents and adults. Cochrane Database Syst Rev. 2005;18(2):CD003879.

15. Tagliaferro EPS, Meneghim MC, Ambrosano GMB, Pereira AC, Sales-Peres SHC, Sales-Peres A, et al. Distribution and prevalence of dental caries in Bauru, Brazil, 1976-2006. Int Dent J. 2008;58(2):75-80.

16. World Health Organization (WHO). Oral health surveys: basic methods. 4th ed. Geneva: WHO Press; 1997.

17. Peres MA, Traebert J, Marcenes W. Calibration of examiners for dental caries epidemiological studies. Cad Saúde Pública. 2001;17(1):153-9.

18. Donner A, Eliasziw M. Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. Biometrics. 1994;50(2):550-5.

19. Brasil. Ministério da Saúde. SAS/SVS/Departamento de Atenção Básica. Coordenação Nacional de Saúde Bucal. SB Brasil 2010: Pesquisa Nacional de Saúde Bucal. Manual de Calibração de Examinadores. Brasília: Ministério da Saúde; 2009. Disponível em: http://www.saude.gov.br/bucal.

20. Secretaria Estadual da Saúde do Estado de São Paulo. Faculdade de Odontologia de Piracicaba - UNICAMP. SB São Paulo 2015: Pesquisa Estadual de Saúde Bucal. Manual de Calibração de Examinadores. São Paulo, Piracicaba, SP: A Secretaria; 2015. Disponível em: https://w2.fop.unicamp.br/sbsp2015/down/manual_de_calibracao_SBSP_2015.pdf.

Inquéritos de cárie dentária: considerações sobre a reprodutibilidade

Objetivo: Avaliar o impacto do registro de dentes anteriores e terceiros molares, cujas condições clínicas pouco variam, sobre as mensurações de concordância inter-examinadores.

Métodos: Dados clínicos de 56 escolares de 12 anos de idade, previamente coletados por 2 examinadores, segundo os códigos e critérios expressos no "Levantamentos em Saúde Bucal: métodos básicos" foram analisados neste estudo. Os efeitos da inclusão/exclusão destes dentes sobre a concordância inter-examinadores foram mensurados calculando-se a porcentagem geral de concordância (GPA) e estatística Kappa (κ).

Resultados: A inclusão de terceiros molares aumentou a GPA para dados dicotomizados ou não. A exclusão de dentes anteriores diminuiu a GPA para dados dicotomizados ou não. Quando associada à inclusão de terceiros molares, sob perfeita concordância ($\kappa = +1$), os valores de Kappa foram artificialmente aumentados tanto para dados dicotomizados quanto não-dicotomizados.

Conclusão: A inclusão/exclusão de dentes anteriores e/ou de terceiros molares, no sentido de se evidenciar melhor as discordâncias entre examinadores, podem impactar positiva ou negativamente sobre a fidedignidade da reprodutibilidade mensurada. Uma advertência clara sobre o impacto da inclusão dos dentes terceiros molares no cálculo da reprodutibilidade, para a faixa etária de 12 anos, deveria estar presente no "Levantamentos em Saúde Bucal: métodos básicos" e manuais semelhantes.

Descritores: Cárie dentária. Índice CPO. Epidemiologia. Reprodutibilidade dos testes.