Article published online: 2021-11-01

ORIGINAL ARTICLE

Reproducibility of an organoleptic method for halitosis assessment

Késsia Suênia Fidelis de Mesquita-Guimarães, Gabriela Cristina Santin[1], Camila Scatena[2], Antônio Luiz Rodrigues-Junior[3], Mônica Campos Serra[4]

Departments of Pediatric Dentistry and [4]Restorative Dentistry, School of Dentistry of Ribeirão Preto, University of São Paulo, Ribeirão Preto, [3]Department of Social Medicine, School of Medicine of Ribeirão Preto, University of São Paulo, SP, [1]Department of Dentistry, State Faculty of Maringá, Maringá, PR, [2]Department of Pediatric Dentistry, Faculty of Serra Gaucham, Caxias do Sul, RS, Brazil

Address for correspondence:
Dr. Késsia Suênia Fidelis de Mesquita Guimarães,
Av. do Café, s/n, Monte Alegre,
14040-904, Ribeirão Preto, SP, Brazil.
E-mail: kessiamesquita@gmail.com

ABSTRACT

Background: The organoleptic evaluation is considered the gold standard between evaluation methods of halitosis, but its main drawback is the difficulty of reproducibility. **Purpose:** The aim of this study was to evaluate the reproducibility of an organoleptic evaluation method using three levels of scores (0 = no odor, 1 = moderate odor, and 2 = strong odor) to increase reliability between researchers and clinicians. **Methods:** The evaluation was blindly conducted by two examiners previously calibrated by the Smell Identification Test and compliance in clinical trials. Statistical calculations were done with STATA® software version 9.0. **Results:** The degree of agreement between examiners was 82.5%, with estimated Kappa (κ =0.69), with substantial agreement. **Conclusion:** The scale used in this study by organoleptic method was effective and reproducible but must be repeated and compared to other methods for better consistency of results.

Key words
Breath, halitosis, judges, organoleptic

INTRODUCTION

Halitosis is a common condition that has an important negative factor in social communication.[1-3] Clinicians generally prefer to use the gold standard[1,3-6] organoleptic method[7] to evaluation because it is a practical sensory test,[8] has a low cost, and does not require specific equipment.[1]

Examiners can estimate the quality of the breath using common odor scale scores to assess the intensity of how pleasant or unpleasant is the odor.[7] Studies of agreement or reliability are performed to evaluate the error of a measurement.[9] Therefore, it is estimated that the higher Kappa, the greater the concordance and reliability or reproducibility of the results. The most widely used scale of halitosis,[10-13] proposed by Rosenberg *et al.*,[14] has six levels of evaluation, but the Kappa values are low.[15]

Researchers have sought to simplify the test and implement better ways of training, calibration, and standardization between examiners[7] for increased reliability and reproducibility. Although organoleptic tests with four levels of evaluation have been proposed[16] and Kappa values have increased,[16] to date, no study has been carried out using a scale with three levels.

The aim of this study was to evaluate the interexaminer reliability of an organoleptic evaluation method with a halitosis detection scale with three levels of scores.

Access this article online	
Quick Response Code:	Website: www.ejgd.org
	DOI: 10.4103/2278-9626.198600

How to cite this article: de Mesquita-Guimarães KS, Santin GC, Scatena C, Rodrigues AL, Serra MC. Reproducibility of an organoleptic method for halitosis assessment. Eur J Gen Dent 2017;6:9-13.

METHODS

Study design

A cross-sectional and exploratory study was conducted by quantifying the response variables for halitosis. The evaluation was performed by means of organoleptic or hedonic test, as a tool for data collection, resulting in a scaled measurement categorized with three scores (0 = no odor, 1 = moderate odor, and 2 = strong odor).

The study consisted of assessing the validity of the diagnostic technique, in which two examiners performed the observation of the object under study (halitosis), aiming at the analysis of reproducibility of clinical criteria.

Ethical aspects

The study was initiated after approval by the Ethics Committee of the School of Dentistry of Ribeirão Preto - FORP/USP (CAAE No 10755212.7.0000.5419) and the consent form was signed by all volunteers, guaranteeing the individual right of free choice. The principles of bioethics, autonomy, vulnerability, and beneficence/nonmaleficence were respected.

The study did not cause scratches or damages to participants. The volunteers received an oral hygiene kit (toothbrush, toothpaste, mouthwash, and dental floss) and guidelines about halitosis and oral care by an educational brochure developed for this purpose.

Calibration of judges – The Smell Identification Test

Before the clinical evaluation, the olfactory acuity was assessed using the Portuguese version of the Smell Identification Test (SIT), University of Pennsylvania (Sensonics Inc., Haddon Heights, NJ, USA),[17] by two examiners (KSFMG and GCS). Later, the interexaminer agreement would be assessed through Kappa test.

The olfactory test consists of four cards with ten pages. Each page contains a marquee at the bottom that, when scraped with a pencil, releases a scent corresponding to one of the four suggested odor options. Each examiner must check the box that most closely matches up for a total forty test odors.

In the test, the estimated Kappa (κ =0.96) was classified on the scale as an "almost perfect agreement,"[18] showing 97.5% of agreement between examiners.

Calibration by agreement

A calibration of agreement was performed by the examiners (KSFMG and GCS) using ten volunteers, not included in the survey. In cases of disagreement (20% of cases), a mutual consensus was sought from the discussion and interpretation of the examination criteria.

Selection of participants

The sample was determined by the convenience criterion, composed of forty volunteers of both genders (8 men and 32 women), aged between 18 and 40 years old (mean for men: 32.7 years old, mean for women: 29 years old) who attend the School of Dentistry of Ribeirão Preto, University of São Paulo, with powers of discernment to explicitly express informed consent to participate in the study. Volunteers who presented aspects that interfere directly or indirectly in the homogeneity of the sample such as the presence of diabetes, use of antibiotics, smoking, individuals with disabilities that affect decision-making, and those who at any time manifested the desire to no longer participate the project were excluded from the study.

Clinical evaluation of organoleptic test

Examiners conducted the research blindly, independently, and without communication to ensure that the opinion of one did not interfere with that of the other. A privacy screen was made (200 cm × 90 cm) with a plastic tube inserted through it (10 cm × 4.5 cm) to separate the volunteers and examiners. During recruitment, the volunteers were instructed to avoid eating spicy foods, onions, and garlic, using perfumed cosmetics, drinking alcohol, and using mouthwash for at least 24 h.[1,19]

On evaluation, the judges also avoided drinking coffee, tea, or juice, smoking, and using perfumed cosmetics.[19]

The volunteers were instructed to close the mouth for 3 min in an upright position and breathe through their nose.[14] While the volunteers exhaled slowly into one end of the tube, each examiner individually performed the evaluation on the other end and confidentially marked a record containing scores. Examiners evaluated volunteers in the same order.

The standardized forms had been previously encoded, stored, and sealed in brown envelopes to ensure that the study was triple-blind (volunteers, judges, and statistician).

Assessment of reproducibility

For scores, reliability analysis was used. By measuring the percentage from an array of responses of the examiners, Kappa statistic or Kappa coefficient (κ) was estimated, which expresses the level of agreement observed between judges beyond the level that would be expected by chance,[12] according to the following formula:

$$\kappa = \frac{C - C_0}{1 - C_0}$$

C is the correlation observed by the study and C_0 is the agreement expected by chance. The evaluation criterion of the estimated Kappa was established by Landis and Koch,[18]

allowing the interpretation of Kappa values [Table 1]. Statistical calculations were done with STATA® software version 9.0. (StataCorp LP, Texas, USA).

RESULTS

The clinical evaluation of halitosis obtained a degree of agreement of 82.5% between the examiners. The Kappa estimate (κ =0.69) [Table 2] was rated with "substantial agreement" on the scale.[18]

The examiners agreed that twenty volunteers (50%) had no oral malodor (grade 0), seven volunteers (17.5%) had moderate odor (grade 1), and six volunteers (15%) showed strong odor (grade 2) [Table 2].

DISCUSSION

So far, there is no uniformity in protocols for the diagnosis of halitosis.[1] The volatile sulfur-containing compounds (VSCs), using the halimeter equipment, are a relatively inexpensive and easy to use method[3] but cannot detect some important odorants[3,20] such as volatile short-chain fatty acids, polyamines, alcohols, phenyl compounds, alkanes, ketones, and nitrogen-containing compounds.[20] In addition, portable instruments that measure volatile compounds may fail in device sensitivity due to the contamination of the sensor, which requires periodic calibration to adjust for sensitivity loss.[14]

Gas chromatography is a highly reproducible and reliable method[20] that measures, through of the production of mass spectra, the concentration of VSCs in the saliva samples, tongue coating, or expired air.[3,20] However, it is expensive, as is most of the specific equipment,[21] making the technique applicable only in academic practice.[21]

Table 1: Landis and Koch Kappa Scale (1977)

κ	Degree of agreement
<0.00	Without agreement
0.00-0.20	Insignificant
0.21-0.40	Median
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Table 2: Frequency distribution of the responses of clinical assessments of halitosis, given by different judges

Judge A	Judge B			
	0	1	2	Total
0	20	2	0	22
1	5	7	0	12
2	0	0	6	6
Total	25	9	6	40

Studies show a significant correlation between these methods and organoleptic methods,[14,16,22] which are performed by the perception of oral odor by smell. Despite being a subjective method,[23] the organoleptic method is also referred to as being easy to perform,[4] is similar to the daily situation of the patient,[24] does not require equipment,[4] and has a low cost.[1] The main disadvantage is the low reproducibility and that it replies upon inter- and intra-examiner reliability.[24] However, this can be improved with training and calibration of examiners.[5]

Some studies did not report the calibration of their examiners[11,25] or do not specify how the calibration was performed.[4,13] Others cite calibration, by agreement,[22] the training of examiners,[12,16,26] or olfactory tests,[6,17] but do not calculate the Kappa.[4,6,13,17] Kappa shows the proportion of agreements beyond that expected by chance and ranges from "−1" to "+1," where "−1" means complete disagreement and "+1" is exact agreement in the readings.[27]

The Kappa is important to assess how often the exact same scores can be replicated,[15] and its absence makes difficult to interpret and to compare the results. In this study, calibration before the sensory evaluation was performed by two examiners with the olfactory SIT (Sensonics Inc., Haddon Heights, NJ, USA) and the Kappa was calculated for interexaminers (κ =0.96), showing high sensitivity to smell and degree of almost perfect agreement,[18] ensuring a highly reliable calibration between the two examiners.

Although one examiner is sufficient to perform the organoleptic test, it is highly recommended to have a second examiner to provide a second opinion or take measurements when the first is not present or when he or she presents fatigue.[28] Thus, in this research study, the organoleptic test was performed by two examiners, enabling the calculation of Kappa for interexaminers.

The number of scores used in sensory evaluation can also affect the results achieved. The more simplified for the scale scores, the greater the possibility of increasing the coefficient. Thus, scales with simplified levels could provide greater agreement, depending on the number of examiners, the number of scores, and if the Kappa calculation is for intra- or inter-examiners.

Using a simplified scale, with three levels of scores and two examiners, the degree of interexaminer agreement in the present study was 82.5% and κ =0.69, with a substantial agreement (0.61–0.80).[18] The degree of agreement found in this study corroborates the results of Oho *et al.*[16] although these researchers used four levels of scores and three examiners.

In studies whose patients are more likely to have different degrees of halitosis and that involve further

investigation, such as periodontal disease, plaque coating on the tongue, and smoking, a broader scale, as proposed by Oho *et al.*,[16] with four levels of scores will possibly be required. However, the most simplified scale proposed in this study, with three levels of scores, could be recommended in dental practices, in offices, and in pediatric dentistry (in which requires skill in the child's approach) since it has been shown to be reproducible, and it is easier to perform, especially if the professional is not calibrated and has little experience.

The aim of this study was not to infer results for the study population, nor to perform a dichotomic evaluation (yes or no) for the presence of halitosis, but to evaluate the performance of a simplified data collection instrument - a diagnostic halitosis technique - with a view to analysis of in reproducibility.

In some cases, the diagnosis of halitosis requires a careful investigation of the patient for prevention and treatment. Questions about frequency of dental visits, dental products used, presence and maintenance of dental hygiene of the oral cavity, time of occurrence of halitosis, psychological factors, dietary habits, smoking, snoring, and/or dry mouth symptoms should be included in the anamnesis.[28] In addition, there should be research in nonoral etiologies such as disorders of upper and lower respiratory tract, gastrointestinal tract disorders, systemic diseases, metabolic disorders, medications, carcinomas, and occurrence of stressful situations.[3]

The scale used in this study was effective but must be repeated and compared to other more specific and sophisticated methods to observe possible correlations to better the consistency of the results methods.

CONCLUSION

The present study agrees that the organoleptic method is effective for evaluation of halitosis. The scale with three levels of scores (0 = no odor, 1 = moderate odor, and 2 = strong odor) is reproducible and a simple alternative for daily assessment in the dental office. However, further studies should be performed comparing this scale with other methods to validate such an outcome.

Acknowledgments

The authors would like to thank National Council for Scientific and Technological Development (CNPq - N 309219/2009-4) for the research grant and financial support.

Financial support and sponsorship

This study was financially supported by National Council for Scientific and Technological Development (CNPq - N 309219/2009-4).

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Bollen CM, Beikler T. Halitosis: the multidisciplinary approach. Int J Oral Sci 2012;4:55-63.
2. Bornstein MM, Stocker BL, Seemann R, Bürgin WB, Lussi A. Prevalence of halitosis in young male adults: a study in swiss army recruits comparing self-reported and clinical data. J Periodontol 2009;80:24-31.
3. Bornstein MM, Kislig K, Hoti BB, Seemann R, Lussi A. Prevalence of halitosis in the population of the city of Bern, Switzerland: A study comparing self-reported and clinical data. Eur J Oral Sci 2009;117:261-7.
4. Apatzidou AD, Bakirtzoglou E, Vouros I, Karagiannis V, Papa A, Konstantinidis A. Association between oral malodour and periodontal disease-related parameters in the general population. Acta Odontol Scand 2013;71:189-95.
5. Greenman J, Lenton P, Seemann R, Nachnani S. Organoleptic assessment of halitosis for dental professionals – general recommendations. J Breath Res 2014;8:017102.
6. Iwanicka-Grzegorek E, Michalik J, Kepa J, Wierzbicka M, Aleksinski M, Pierzynowska E. Subjective patients' opinion and evaluation of halitosis using halimeter and organoleptic scores. Oral Dis 2005;11 Suppl 1:86-8.
7. Greenman J, Duffield J, Spencer P, Rosenberg M, Corry D, Saad S, *et al.* Study on the organoleptic intensity scale for measuring oral malodor. J Dent Res 2004;83:81-5.
8. Greenman J, El-Maaytah M, Duffield J, Spencer P, Rosenberg M, Corry D, *et al.* Assessing the relationship between concentrations of malodor compounds and odor scores from judges. J Am Dent Assoc 2005;136:749-57.
9. Kizhner V, Xu D, Krespi YP. A new tool measuring oral malodor quality of life. Eur Arch Otorhinolaryngol 2011;268:1227-32.
10. Laleman I, Dadamio J, De Geest S, Dekeyser C, Quirynen M. Instrumental assessment of halitosis for the general dental practitioner. J Breath Res 2014;8:017103.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
12. Mcdowell J. Measuring Health: A Guide to Rating Scales and Questionnaires. 3rd ed. New York: Oxford University Press; 1987. p. 3-748.
13. Morita M, Wang HL. Relationship of sulcular sulfide level to severity of periodontal disease and BANA test. J Periodontol 2001;72:74-8.
14. Rosenberg M, Septon I, Eli I, Bar-Ness R, Gelernter I, Brenner S, *et al.* Halitosis measurement by an industrial sulphide monitor. J Periodontol 1991;62:487-9.
15. Nachnani S, Majerus G, Lenton P, Hodges J, Magallanes E. Effects of training on odor judges scoring intensity. Oral Dis 2005;11 Suppl 1:40-4.
16. Oho T, Yoshida Y, Shimazaki Y, Yamashita Y, Koga T. Characteristics of patients complaining of halitosis and the usefulness of gas chromatography for diagnosing halitosis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2001;91:531-4.
17. Pereira MG. Epidemiology: Theory and Practice. Rio de Janeiro: Guanabara Koogan; 1995. p. 596.
18. Quirynen M, Dadamio J, Van den Velde S, De Smit M, Dekeyser C, Van Tornout M, *et al.* Characteristics of 2000 patients who visited a halitosis clinic. J Clin Periodontol 2009;36:970-5.
19. Dadamio J, Van Tornout M, Van den Velde S, Federico R, Dekeyser C, Quirynen M. A novel and visual test for oral malodour: First observations. J Breath Res 2011;5:046003.

20. Baharvand M, Maleki Z, Mohammadi S, Alavi K, Moghaddam EJ. Assessment of oral malodor: A comparison of the organoleptic method with sulfide monitoring. J Contemp Dent Pract 2008;9:76-83.
21. Rosenberg M, Kulkarni GV, Bosy A, McCulloch CA. Reproducibility and sensitivity of oral malodor measurements with a portable sulphide monitor. J Dent Res 1991;70:1436-40.
22. Rosenberg M, McCulloch CA. Measurement of oral malodor: current methods and future prospects. J Periodontol 1992;63:776-82.
23. Seemann R, Conceicao MD, Filippi A, Greenman J, Lenton P, Nachnani S, *et al.* Halitosis management by the general dental practitioner – Results of an international consensus workshop. J Breath Res 2014;8:017101.
24. Sopapornamorn P, Ueno M, Vachirarojpisan T, Shinada K, Kawaguchi Y. Association between oral malodor and measurements obtained using a new sulfide monitor. J Dent 2006;34:770-4.
25. Sterer N, Greenstein RB, Rosenberg M. Beta-galactosidase activity in saliva is associated with oral malodor. J Dent Res 2002;81:182-5.
26. Tanaka M, Anguri H, Nishida N, Ojima M, Nagata H, Shizukuishi S. Reliability of clinical parameters for predicting the outcome of oral malodor treatment. J Dent Res 2003;82:518-22.
27. van den Broek AM, Feenstra L, de Baat C. A review of the current literature on aetiology and measurement methods of halitosis. J Dent 2007;35:627-35.
28. Yaegaki K, Coil JM. Examination, classification, and treatment of halitosis; clinical perspectives. J Can Dent Assoc 2000;66:257-61.