# Accepted manuscript

Rodrigues, L. G., Sampaio, A. A., Cruz, C. A. G., Vettore, M. V. & Ferreira, R. C. (2022). A systematic review of measurement instruments for oral health assessment of older adults in long-term care facilities by nondental professionals. Gerodontology, 2022, 1-13. https://doi.org/10.1111/ger.12648

Published in: Gerodontology

DOI: https://doi.org/10.1111/ger.12648

AURA: https://hdl.handle.net/11250/3066208



Available: 31. July 2023

# A systematic review of measurement instruments for oral health assessment of older adults in long-term care facilities by nondental professionals

Lorrany Gabriela Rodrigues[1], Aline Araújo Sampaio[1], Carlos Antonio Gomes da Cruz[1], Mario Vianna Vettore[2], Raquel Conceição Ferreira[1]

[1]Faculty of Dentistry, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[2]Department of Health and Nursing Sciences, University of Agder, Kristiansand, Norway

Short Title: Oral health assessment of older adults in long-term care facilities

Correspondence:
Raquel Conceição Ferreira
Department of Social and Preventive Dentistry
Faculty of Dentistry/Federal University of Minas Gerais
6627 Presidente Antonio Carlos Av.
Belo Horizonte, Minas Gerais, Postal code, Brazil
Tel: +553134092412
E-mail: raquelcf@ufmg.br

Keywords: Oral health. Long term care. Older people. Psychometrics. Geriatric Assessment

**ABSTRACT**

Background: Regular oral health assessment of older adults living in Long-term Care Facilities (LTCF) can improve their oral health. Aims: This study aimed to systematically review studies describing the development of instruments employed by nondental professionals to assess the oral health of older adults in LTCF and to evaluate their measurement properties. Material & Methods: Electronic searches were conducted in the MEDLINE (PubMed), Embase, Web of Science, Scopus, and LILACS databases. Measurement properties of the identified instruments were evaluated using the Consensus-based Standards to select health Measurement Instruments (COSMIN) checklist. Studies assessing at least one measurement property (validity, reliability, or responsiveness) of instruments used to assess oral health of older adults living in LTCF by nondental professionals were considered. The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) checklist was used to evaluate the quality of evidence. Results: Fifteen studies reporting measurement properties of seven instruments were selected. The ohr-interRAI and the OHAT were considered to have sufficient content validity, with high and moderate evidence quality, respectively. OHAT, BOHSE, and DHI showed acceptable results on reliability but with very low quality of evidence. DHI and OHSTNP also showed acceptable results for criterion validity, but, with low quality of evidence, insufficient or unclear results were observed for the remaining measurement properties. Studies evaluating the validity criteria of BOHSE and hypotheses testing of ohr-MDS were considered to have high evidence quality. Discussion and Conclusion: The ohr-interRAI can be provisionally recommended for use until further evidence is provided. Further methodologically rigorous studies are needed to assess the measurement properties of the existing instruments.

**KEYWORDS** geriatric assessment, long-term care, older people, oral health, psychometrics

## 1 INTRODUCTION

Assessing the oral health of older adults living in Long-Term Care Facilities (LTCF) should be part of their health care, as a relevant strategy to prevent oral and systemic diseases such as aspiration pneumonia.[1] Measurement instruments have been developed to assess the oral health of older adults living in LTCF by caregivers and nursing professionals; these include the Oral Health Assessment Tool (OHAT), Minimum Data Set (MDS), Dental Hygiene Registration (DHR), Oral Health Screening Tool for Nursing Personnel (OHSTNP), and oral health-related section of the interRAI (ohr-interRAI).[1–5] The use of valid and reliable instruments is essential for identifying oral health problems, allowing daily oral health care planning, and timely referral to dental services when necessary,[3,5–9] although there is evidence that some may have methodological flaws.[8,9]

Three previous systematic reviews analysed instruments for assessing the oral health of institutionalised[9] and noninstitutionalised[8,10] older adults by nondentist professionals. Of those, only one review evaluated the measurement properties of the instruments using a validated assessment tool, such as the Consensus-based Standards to select health Measurement Instruments (COSMIN) and Grading of Recommendations, Assessment, Development, and Evaluation (GRADE). Such tools aim to improve the selection of the best instrument for both research and clinical practice.[9] The previous reviews did not focus on specific instruments for older adults living in LTCF. Additionally, an updated review of the instruments is timely, since new studies have been published.

Implementing best practices in LTCF, such as oral health assessment, should be supported by evidence-based information and must be appropriate for the context. This review can identify the most suitable instrument for oral health evaluation of institutionalised older adults by nondentist professionals, indicate the need for additional studies assessing the measurement properties of the existing instruments, or recommend developing and testing an additional instrument. This study aimed to systematically review the literature to describe instruments used to assess the oral health of older adults living in LTCF by nondental professionals and to evaluate their measurement properties.

## 2. MATERIALS AND METHODS

This systematic review was developed and reported according to the Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA)[11] and COSMIN checklist for systematic reviews.[12] The protocol of this systematic review was registered on the PROSPERO database (Ref: CRD42020191479) and previously described.[13] The review questions were What instruments have been developed and tested to assess the oral health of older adults living in LTCF by nondental professionals, and Do these instruments present acceptable measurement properties? The inclusion criteria were the instrument must aim to assess oral health of older adults living in LTCF (assessed population); the instrument must be applied by nondental professionals (target population), such as

caregivers and nurses; the study must describe the development of the instrument and/or the assessment of at least one of the following measurement properties: validity, reliability, or responsiveness and published in English, Spanish, or Portuguese. No restrictions for the year of publication or publication status were imposed. This review included epidemiologic studies, comprising validation and observational studies reporting measurement properties of the instruments used to assess oral health of older adults living in LTCF by nondental professionals.

### 2.1 Search strategy

Electronic searches were conducted in the databases MEDLINE (PubMed), Embase, Web of Science, Scopus, and LILACS (Latin American and Caribbean Literature in Health Sciences). The reference lists of included articles were also screened for additional relevant research (Appendix S1). Screening and eligibility assessment was performed independently by two trained reviewers. They evaluated the relevant articles according to abstract and then the full text, respectively. Disagreements between the reviewers were resolved by consensus. Two reviewers performed data extraction independently using a prepiloted spreadsheet developed for this purpose. A third reviewer was consulted in case of disagreements. Detailed information on the methods is found in other publications.[13]

### 2.2 Quality assessment

Figure 1 presents the steps to assess the methodological quality and grading the quality of evidence. Two trained reviewers independently evaluated the quality of the included studies using the COSMIN checklist.[12] The COSMIN aims to improve the selection of Patient-Reported Outcome Measures (PROMs) instruments in research and clinical practice by using specific tools to select the most appropriate one. Despite focusing on PROMs, this method can also be adapted for other purposes, such as instruments for measuring clinical outcomes.[12] According to COSMIN, the three domains, which must be evaluated to assess the quality of the instrument are reliability, validity, and responsiveness. The definitions of measurement properties and their respective domains can be found in the following link: https://cosmin.nl/wp-content/uploads/COSMINdefinitions-domains-measurement-properties.pdf. [12]

The Risk of Bias checklist was used to assess the methodological quality of the included studies. The checklist contains standards related to design requirements and preferred statistical methods to assess the measurement properties in 10 boxes: with one box for PROM development and nine for measurement properties: content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. The measurement properties evaluated in a study were used to determine which box should be completed. A standard COSMIN Excel spreadsheet was used for each box. A four-point rating scale was assigned for each study as follows: “very good,” “adequate,” “doubtful,” or “inadequate.” The overall rating of the quality of each study was determined considering the lowest rating of any standard for each box.[12]

The measurement properties of each study were rated as sufficient (+), insufficient (−), or indeterminate (?) against the adapted criteria for good measurement properties[12] (Appendix S2). The overview Tables were used to assess whether the results of two or more articles using the same instrument were consistent with each other to obtain the pooled result against the criteria for good measurement properties. When possible, the results from different studies related to the same measurement property were statistically pooled through meta-analysis using a random effects model in STATA, version 16.0.

### 2.3 Grading the quality of evidence

After pooling the evidence according to the measurement property for each instrument and rating the pooled result against the criteria for good measurement properties, the quality of the evidence was graded using a modified version of the GRADE approach for systematic reviews that classifies the certainty of the findings as high, moderate, low, or very low. Four factors of the GRADE approach were used to evaluate measurement properties: risk of bias, inconsistency, imprecision, and indirectness.[14] The GRADE approach was used to downgrade evidence when there were concerns about the quality of evidence. The starting point is always the assumption that the obtained pooled result is of a "high" quality level. The quality of evidence can be subsequently downgraded by one or two levels per factor to moderate, low or very low evidence if there is imprecision (low sample size) or serious or very serious risk of bias, inconsistency, indirectness. The quality of evidence can even be downgraded by three levels when the evidence is based on only one inadequate study (extremely serious risk of bias).[13,14]

### 2.4 Formulating recommendations

The recommendation must consider interpretability (degree to which one can assign qualitative meaning that is, clinical or commonly understood connotations, to a quantitative score or change in scores) and feasibility (ease of application of the instrument in the intended context of use). Finally, to reach an evidence-based and fully transparent recommendation for use in an evaluative application, the instruments were categorised into three categories, with respect to the construct of interest and study population according to COSMIN.

A Instruments with evidence of sufficient content validity (any level) AND at least low-quality evidence for sufficient internal consistency. These instruments can be recommended for use and results obtained with these instruments are trustworthy.

B Instruments categorised not under A or C. These instruments have the potential to be recommended for use, but they require further research to assess their quality.

C Instruments with high quality evidence for an insufficient measurement property. These instruments should not be recommended for use.

When only instruments categorised as "B" are found in a review, the one with the best evidence for content validity could be provisionally recommended for use until further evidence is provided.[12]

## 3 RESULTS

The PRISMA flow diagram of the systematic literature search is presented in Figure 2. A total of 2673 articles were identified, of them, 15 were included in the systematic review.

### 3.1 Oral health assessment instruments' characteristics

Table 1 shows the characteristics of the included studies. The studies were of older adults living in LTCF or nursing homes in several countries.

Table 2 shows the characteristics of the seven instruments used to evaluate the oral health of older adults living in LTCF by nondental professionals. The target population was composed of nursing staff and caregivers and most of the instruments was applied through clinical examination. One study was based on interviews and observations, with a recall period of 3 days before the assessment. In that study, the ohr-interRAI instrument was applied by interview and observation during meals or inspection of the mouth.[5] English was the original language of most instruments. The OHSTNP was translated and applied in Japanese. The OHAT was translated into Portuguese,[15] Italian,[16] Turkish,[17] and German[18] languages. All versions were included in this review because different measurement properties were measured. Only the OHSTNP was developed to be used by nursing personnel for screening oral health and functional aspects of oral health without previous training.[4]

Most instruments use a three-point response scale (healthy, changes and, unhealthy)[1,3,4,15–24] accompanied by criteria that define each response option. For example, the Tongue category of OHAT is defined as "healthy" (normal, moist roughness, pink); "changes" (patchy, fissured, red, coated) and "unhealthy" (a patch that is red and/or white, ulcerated, swollen). Images are used in the OAS to exemplify each oral health condition evaluated. The instruments using dichotomous response scales[5,25,26] require the professional to choose "yes" or "no" according to a predefined parameter that defines the oral health condition. For example, the category "Has broken, fragmented, lose or otherwise nonintact natural teeth (damaged teeth)" from ohr-interRAI describes a condition of damaged teeth and the presence of any alteration was considered sufficient to refer the older adults to a dental surgeon. Some instruments take into account the sum of the scores, but they do not use those scores to make any recommendations.

The items of the instruments are described in Table 3. The instrument DHI[3] is based only on the presence or absence of dental biofilm on natural teeth. The OHSTNP[4] was the most complete instrument for assessing oral health, including tongue protrusion, cheek puffing test, articulation, oral intake, and coughing during meals besides the items presented in

others instruments. Pairs of teeth in chewing position are evaluated by BOHSE[21,22] and OAS.[23] The BOHSE also considers the lymph nodes.[21,22]

The results of the rating system based on COSMIN Risk of bias are presented in Appendix S3. The pooled results, overall rating, and quality of evidence are also reported (Appendices S3, S4, and S5). The measurement property "Criterion validity" was the most investigated among the studies, while no study has assessed Cross-cultural validity/measurement invariance or Responsiveness (Figure 3). Eight studies using 6 instruments evaluated the content validity and/or instrument development. The ohr-interRAI and OHAT were considered to have sufficient content validity. The former was judged to be of high evidence quality and the latter to be of moderate evidence quality (Table 4).

Only the OHAT, BOHSE, and DHI had sufficient (+) data on reliability, but with very low quality of evidence. DHI and OHSTNP also had sufficient (+) data on Criterion validity, but with low quality of evidence. The remaining data on measurement properties were considered to be insufficient or indeterminate. The quality of evidence was High for studies that evaluated criterion validity for the BOHSE, and hypotheses testing with the ohr-MDS (Table 5). The pooled Spearman's correlation coefficient for Criterion validity of BOHSE was 0.56 (95% CI 0.46–0.67) (Figure 4).

The recommendation level for an evaluative application of the instruments was B (OHAT, DHI, OAS, OHSTNP, and Ohr-interRAI) or C (BOHSE and ohr-MDS).

## 4 DISCUSSION

This systematic review examined the quality of the available evidence of the measurement properties of instruments used to assess the oral health of older adults living LTCF by nondental professionals. The ohr-interRAI, OHAT, DHI, OAS instruments have the potential to be recommended for use, but they require further research to assess the quality. However, among these, the ohrinterRAI was considered to be the most appropriate instrument, because it was the only instrument categorised as "B" that showed a high-quality evidence of sufficient content validity.[12] Overall, the instruments are similar in respect of assessment methods and similar categories, with easy clinical understanding and capable of generating interpretable findings, resulting in a decision to refer the older adults to a dentist or to improve oral hygiene care. These instruments were considered to have adequate length and were considered feasible in the context of LTCF. However, the application of most instruments needs training if applied by nondentist professionals.

A crucial aspect of assessing the quality of evidence for most studies was the downgrade regarding the risk of bias, due to the existence of few studies of good individual quality

on the same instrument and the imprecision of their findings as a result of small sample sizes.

Content validity is considered the most important measurement property because it should be clear that the items of the instrument are relevant, comprehensive, and comprehensible. In this review, a very low quality of evidence of insufficient content validity was observed for the majority of the instruments (OHSTNP, OAS, BOHSE, and DHI). The included studies showed a serious risk of bias, and the main reason was the lack of questions about the instruments for nondental professionals. This finding may be explained by a lack of consensus on the methods to assess content validity, resulting in large variability among studies.[27] Another hypothesis is that this measurement property was not considered because the evaluated instruments measure clinical outcomes and do not measure a PROM. However, it is necessary to assess whether all items are relevant and encompass all the required aspects to evaluate the oral health conditions among older adults. Additionally, the presented concepts must be understandable by nondental professionals who will apply the instrument in older adults.

The standards of construct validity (hypothesis testing) for the ohr-MDS were scored as very good on risk of bias evaluation, reaching a high quality of evidence.[26] Studies assessing the properties of criterion validity of the ohr-MDS did not use adequate statistical methods.[19,25] The Ohr-InterRAI is a newer version of the ohr-MDS. Although the ohr-InterRAI has a sufficient overall rating and high-quality evidence for good content validity, it has very low quality of evidence for reliability, measurement error and criterion validity.[20] Only one study of very low evidence quality reported reliability. Although this study had an adequate sample size, it ranked as inadequate in the evaluation of bias risk due to the nonreporting of the time between the test–retest resulting extremely serious risk of bias.[20] Furthermore, eight items of the instrument showed low kappa coefficients (below 0.70) (criteria for good measurement property), with a lower value for “Gums”.20 Measurement error was sufficient because the percentage of agreement was above 80% for all items. In this same study, the criterion validity was indeterminate because the sensitivity and specificity measures, which are recommended for the dichotomous scale, were not calculated.[20]

There was only very low-quality evidence of sufficient reliability but insufficient (Kappa >0.79) measurement error (% agreement < 0.80 for the BOHSE. These two measurement properties were evaluated by the same study,[21] which demonstrated an extremely serious risk of bias because the time interval was inadequate. A time interval between administrations of about 2 weeks is often considered appropriate,[12] i.e., long enough to prevent recall bias and short enough to ensure that the patient's oral health status has not changed. For criterion validity, the pooled estimate demonstrated an insufficient Spearman correlation coefficient (correlation with gold standard 100), and absence of indirectness. The high-quality evidence of insufficient results suggested the nonrecommendation of BOHSE for evaluative application. The low correlation between

scores by dentists and nondental professionals indicates the need for training the nursing team to achieve consistent performance against the gold standard.

The BOHSE was considered a difficult and length instrument. The OHAT was developed from the simplification of this instrument for more practical use.[1] This was the instrument in which more measurement properties were evaluated. A moderate quality of evidence of indeterminate structural validity was demonstrated by a study of the Italian version of the OHAT.[16] The authors did not provide the results of confirmatory factor analysis as determined by criteria for good measurement. The study showed adequate precision because the sample size was 368. However, there was a serious risk of bias because only one study with adequate overall rating is available (moderate quality of evidence). Another moderate-quality evidence demonstrated insufficient internal consistency of OHAT.[15] This study showed a very good rating on the risk of bias evaluation, and the level of evidence was downgraded to moderate because of the imprecision (n < 100. The Cronbach's alpha was 0.40, below the cutoff of 0.70 which indicates good measurement properties. However, it can be argued whether structural validity and internal consistency are relevant measurement properties for the OHAT. Structural validity refers to the degree to which the instrument scores adequately reflect the dimensionality of the construct to be measured[12,16] and internal consistency refers to the degree of interrelatedness among the instrument items. The first question to evaluate these properties is "Does the scale consists of effect indicators, i.e., is it based on a reflective model?". In a reflective model, all items of the instrument indicate the same underlying construct. Therefore, the items are expected to be highly correlated and interchangeable. Its counterpart is a formative model, in which the combination of the items forms the construct. The items do not need to be correlated. Then, structural validity and internal consistency are not relevant for items of formative models. It is believed that the oral health items of the evaluated instruments are not necessarily correlated, as the same individual may have an unhealthy condition along with a healthy one to determine oral health.

There was low evidence for sufficient reliability (ICC = 0.78) and measurement error (% agreement >0.80) of the OHAT. The short test–retest time interval was the reason why the study showed inadequate overall rating, corresponding to an extremely serious risk of bias, downgrading the level of evidence.[1] Lowquality evidence (imprecision; sample size <50 demonstrated insufficient criterion validity.[1,18,21] Regarding risk of bias, although one of the studies showed inadequate overall rating (correlation between continuous scores of oral health scale obtained by nondental professionals and the scale obtained by the dentist—gold standard—was not used as statistical method), two studies of very good quality were available, resulting in the absence of risk of bias for grading the quality of evidence.[1,18] These two studies used different statistical strategies (ICC and Pearson correlation) and thus a pooled estimate was not calculated. However, the two measures were insufficient when compared with the cutoff of ≥0.70 (good measurement property). Besides, the intraclass correlation coefficient (ICC) is a preferred metric for reliability in a test–retest study design.[28,29] The moderate quality of evidence demonstrated the sufficient construct validity of a Turkish version of the OHAT.[17] The level of evidence

was downgraded to a moderate level because the only study was considered with a serious risk of bias (the authors did not report whether the necessary assumptions for ICC were met).

Relying on insufficient sample sizes was a common problem in evaluating instruments' properties. Some instruments obtained sufficient data, such as, internal consistency of the OAS and criterion validity of the OHSTNP and the DHI, but the small sample sizes used in these studies resulted in classifying them as with low or very low quality of evidence due to imprecision.

The reliability, measurement error and criterion validity properties of instruments can benefit from adequate training of nondental professionals. Most tools require additional training prior the administration by the nursing staff and caregivers.[1,3,5,15–17,19,21–23,25,26]

Although some studies reported that this is perceived as an additional workload on the part of the team that is already overloaded,[21,22] the training can improve the professional abilities to detect dental needs.[1] Strategies such as using oral condition photographs can help caregivers recognise healthy and unhealthy oral health.20 Similarly, video training may increase oral health knowledge.[20] Furthermore, attendance at regular training sessions is needed to ensure a lasting effect.[29]

This review demonstrates the need to improve the methodological quality of the studies in respect of sample size determination and selection of appropriate statistical tests according to the property to be measured. Additionally, the studies on content validity must evaluate the relevance and comprehensiveness of the instrument items from the perspective of the target population and professionals from all relevant disciplines. The studies on reliability must define appropriate intervals for the test–retest and evaluate the appropriate measurement properties according to the type of model: reflexive or formative. Standardisation of methods and methodological rigour are necessary to obtain pooled measures for the same instrument based on high-quality evidence. Moreover, check lists should be used to report study findings so that all the necessary information is available for a proper judgement of the available quality of evidence. In addition, some terms of measurement properties are not used consistently in the included studies.[9] Therefore, it is recommended that future studies on the development of measurement instruments follow a standardised methodology, such as the COSMIN standards, to design the studies and report the results.

## 5 CONCLUSION

Given the importance of defining the valid and reliable instruments for oral health assessment of older adults, our findings suggest the need to improve the methodological aspects of the studies and carry out further investigation of the measurement properties of the existing instruments that have potential. The ohr-interRAI seems to be the most

suitable instrument. The use of these instruments in the assessment of oral health, both on admission of residents and over periodic evaluations, is essential to assess the quality of care for elders and therefore to integrate oral health care into the overall care planning.

## AUTHOR CONTRIBUTIONS

Lorrany Gabriela Rodrigues and Raquel Conceição Ferreira contributed to conception and design of the review. Lorrany Gabriela Rodrigues, Carlos Antonio Gomes da Cruz, Raquel Conceição Ferreira, and Aline Araújo Sampaio contributed to acquisition, qualitative analysis and interpretation of the data. Lorrany Gabriela Rodrigues and Carlos Antonio Gomes da Cruz drafted the article. Lorrany Gabriela Rodrigues, Carlos Antonio Gomes da Cruz, Mario Vianna Vettore, Raquel Conceição Ferreira, and Aline Araújo Sampaio contributed to the revision of the manuscript and final approval of the manuscript to be published.

## ACKNOWLEDGEMENTS

The authors thank the Federal University of Minas Gerais (UFMG) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for making this work possible. Raquel Conceição Ferreira received financial support from FAPEMIG, Brazil (Fundação de Amparo à Pesquisa do Estado de Minas Gerais - Programa Pesquisador Mineiro - PPM-00603-18). This study was financed in part by CAPES – Brasil – Finance Code 001.

## CONFLICT OF INTEREST

The authors declare no conflict of interests.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## ORCID

Lorrany Gabriela Rodrigues https://orcid. org/0000-0002-9531-5392

Aline Araújo Sampaio https://orcid.org/0000-0002-8704-5994

Carlos Antonio Gomes da Cruz https://orcid. org/0000-0001-8835-7650

Mario Vianna Vettore https://orcid.org/0000-0001-6299-4432

Raquel Conceição Ferreira https://orcid. org/0000-0001-8897-9345

## REFERENCES

1. Chalmers JM, King PL, Spencer AJ, Wright FAC, Carter KD. The Oral health assessment tool - validity and reliability. Aust Dent J. 2005;50(3):191-199.

2. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, Nonemaker S. Reliability estimates for the minimum data set for nursing home resident assessment and care screening (MDS). Gerontologist. 1995;35(2):172-178.

3. Fjeld KG, Eide H, Mowe M, Hove LH, Willumsen T. Dental hygiene registration: development, and reliability and validity testing of an assessment scale designed for nurses in institutions. J Clin Nurs. 2017;26(13–14):1845-1853.

4. Tsukada S, Ito K, Stegaroiu R, Shibata S, Ohuchi A. An oral health and function screening tool for nursing personnel of long-term care facilities to identify the need for dentist referral without preliminary training. Gerodontology. 2017;34(2):232-239.

5. Krausch-Hofmann S, De Almeida MJ, Declerck D, et al. The oral health-related section of the interRAI: evaluation of test content validity by expert rating and assessment of potential reasons for inaccurate assessments based on focus group discussions with caregivers. Gerodontology. 2019;36(4):382-394.

6. Wårdh I, Jonsson M, Wikström M. Attitudes to and knowledge about oral health care among nursing home personnel - an area in need of improvement. Gerodontology. 2012;29(2):e787-e792.

7. Kossioni AE, Hajto-Bryk J, Janssens B, et al. Practical guidelines for physicians in promoting Oral health in frail older adults. J Am Med Dir Assoc. 2018;19(12):1039-1046.

8. Chalmers JM, Pearson A. A systematic review of oral health assessment by nurses and carers for residents with dementia in residential care facilities. Spec care. Dent. 2005;25(5):227-233.

9. Everaars B, Weening-verbree LF, Jerkovi K, et al. Measurement properties of oral health assessments for non-dental healthcare professionals in older people: a systematic review. BMC Geriatr. 2020;20(1):4.

10. Thapa R, Chimoriya R, Arora A. The development and psychometric properties of oral health assessment instruments used by non-dental professionals for nursing home residents: a systematic review. BMC Geriatr. 2021;21(1):1-16.

11. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and metaanalyses: the PRISMA Statement. PLoS Med. 2009;6(7):e1000097.

12. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-1157.

13. Rodrigues LG, Vettore MV, Figueiredo IL, Sampaio AA, Ferreira RC. Oral health assessment of older adults living in long-term care facilities by non-dental professionals: a systematic review protocol of instruments. BMJ Open. 2021;11:e050289.

14. Schünemann HJ, Brozek J, Guyatt G, Oxman A. GRADE Handbook - Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach. The GRADE work Group; 2013. https://gdt.gradepro.org/app/handbook/handbook. html

15. de Mello ALSF, Zimermann K, Gonçalves LHT. Nurse's assessment of oral health of elderly people: OHAT validity and reliability. Rev Gaucha Enferm. 2012;33(2):36-44.

16. Finotto S, Bertolini G, Camellini R, et al. Linguistic-cultural validation of the oral health assessment tool (OHAT) for the Italian context. BMC Nurs. 2020;19(1):7.

17. Ercan Şahin N, Anne Jablonski R. Psychometric properties of the oral health assessment tool. Cukurova Med J. 2019;44(1):513-520.

18. Klotz AL, Zajac M, Ehret J, Hassel AJ, Rammelsberg P, Zenthöfer A. Development of a German version of the Oral health assessment tool. Aging Clin Exp Res. 2020;32(1):165-172.

19. Arvidson-Bufano UB. Nurses' oral health assessments of nursing home residents pre- and post-training: a pilot study. Spec Care Dent. 1996;16(2):58-64.

20. Krausch-Hofmann S, Tran TD, Janssens B, et al. Assessment of oral health in older adults by non-dental professional caregivers—development and validation of a photograph-supported oral health–related section for the interRAI suite of instruments. Clin Oral Investig. 2020;25(6):3475-3486.

21. Kayser-Jones J, Bird WF, Paul SM, Long L, Schell ES. An instrument to assess the oral health status of nursing home residents. Gerontologist. 1995;35(6):814-824.

22. Lin CY, Jones DB, Godwin K, Knebl JA, Niessen L. Oral health assessment by nursing staff of Alzheimer's patients in a long-termcare facility. Spec Care Dent. 1999;19(2):64-71.

23. Yanagisawa S, Nakano M, Goto T, Yoshioka M, Shirayama Y. Development of an oral assessment sheet for evaluating older adults in nursing homes. Res Gerontol Nurs. 2017;10(5):234-240.

24. Blank LW, Arvidson-Bufano UB, Yellowitz JA. The effect of nurses' background on performance of nursing home resident oral health assessments pre- and post-training. Spec Care Dent. 1996;16(2):65-70.

25. Hoben M, Poss JW, Norton PG, Estabrooks CA. Oral/dental items in the resident assessment instrument - minimum data set 2.0 lack validity: results of a retrospective, longitudinal validation study. Popul Health Metr. 2016;14:36.

26. Fayers PM, Machin D. Quality of Life: the Assessment, Analysis and Interpretation of Patient-Reported Outcomes. 2nd ed. Wiley; 2013:568.

27. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30-46.

28. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Cambridge University Press; 2011:338.

29. Jansson MM, Syrjälä HP, Ohtonen PP, Meriläinen MH, Kyngäs HA, Ala-Kokko TI. Effects of simulation education on oral care practices – a randomized controlled trial. Nurs Crit Care. 2017;22(3):161-168.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

FIGURE 1 Flowchart of methodological quality evaluation and grading the quality of evidence

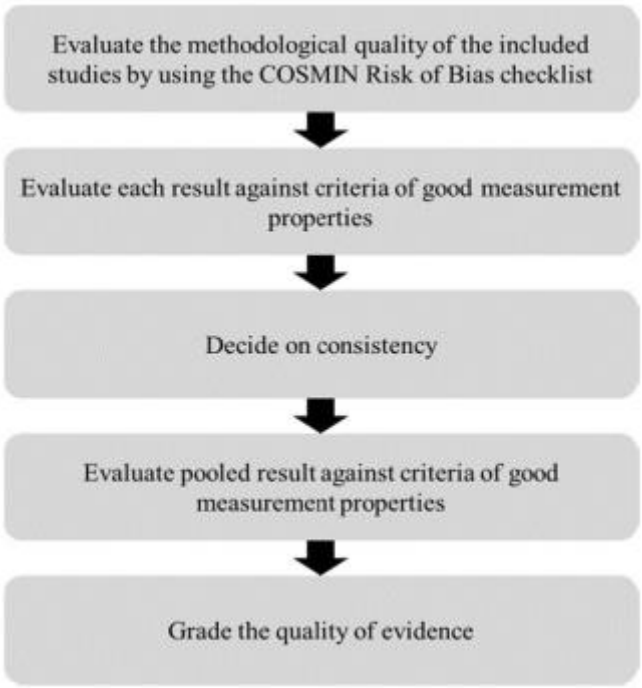


FIGURE 2 PRISMA flow diagram

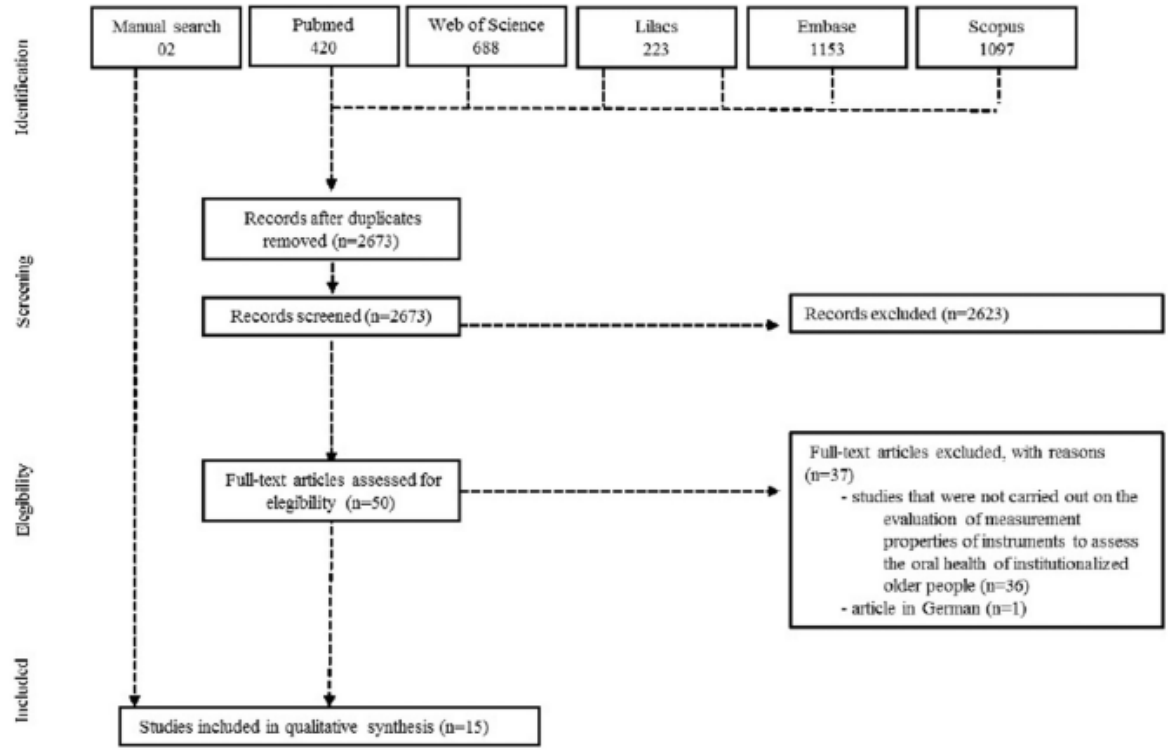

TABLE 1 Characteristics of the included studies

| Authors, date of publication | Country | Setting | Participants | | | |
|---|---|---|---|---|---|---|
| | | | Sample size | Response rate | Age years mean (SD), range | Sex (% females) |
| Chalmers et al., 2005 | Australia | Residential care homes | 455 | 85.20% | 82.1 years | Not informed |
| Mello et al., 2012 | Brazil | Long-term care facility | 50 | Did not report | 60–98 years | Not informed |
| Klotz et al., 2019 | Germany | Nursing home | 18 | 100% | 82.4 (10.3) | 61.1% |
| Sahin et al., 2019 | Turkey | Nursing home | 100 | 71.42% | 65–74 years (41%)<br>75–84 years (45%)<br>85 and over (14%) | 49% |
| Finotto et al., 2020 | Italia | Long-term care facility | 368 | Did not report | Not informed | Not informed |
| Arvidson-Bufano et al., 1996 | USA | Long-term care facility | 50 | Did not report | 65–97 years | 74% |
| Blank et al., 1996 | USA | Long-term care facility | 50 | Did not report | Not informed | Not informed |
| Hoben et al., 2016 | Canada | Nursing homes | 2711 | Not applicable | Assessment 1: 84.4 (8.9) Assessment 2: 85.4 (8.9) Assessment 3: 86.0 (8.8) Assessment 4: 86.6 (8.4) Assessment 5: 86.2 (8.7) Assessment 6: 82.3 (13.1) | Assessment 1: 68.2% Assessment 2: 68.2% Assessment 3: 71.1% Assessment 4: 76.2% Assessment 5: 80.7% Assessment 6: 87.5% |
| Krausch-Hofmann et al., 2019 | Belgium | Long-term care facility | Not applicable[a] | Not applicable | Not applicable | Not applicable |
| Krausch-Hofmann et al., 2020 | Belgium | Long-term care facilities | 260 | 100% | 86.3 (7.3) | 76.9% |
| Kayser-Jones et al., 1995 (465) | USA | Nursing homes | 100 | 71.50% | 82 years | 50% |
| Lin et al., 1999 (536) | USA | Long-term care facility | 68 | 100% | Did not report | 76% |
| Fjeld et al., 2016 | Norway | Nursing home | 41 | 100% | Not informed | Not informed |
| Tsukada et al., 2016 | Japan | Long-term care facility | 57 | 70.21% | 85.9; CI 84.0–87.8 years | 87.71% |
| Yanagisawa et al., 2017 | Japan | Nursing homes | 188 | Did not report | Not informed | Not informed |

a Study does not present evaluation with the elderly population.

TABLE 2 Characteristics of the included instruments used to evaluate the oral health of older adults living in LTCF by nondental professionals

| Author-date | Oral health assessment | Target population (n) | Mode of administration | Recall period | Language | Training need | Scale | Scoring |
|---|---|---|---|---|---|---|---|---|
| Chalmers et al., 2005 | OHAT | Personnel care attendants registered nurses, enrolled nurses, nurse assistants (n = did not report) | Exam registration | NA | English | Yes | Three-point response scale (healthy, changes, and unhealthy) | If any category receives score 1 or 2, please organise for a dentist to examine the resident |
| Mello et al., 2012 | OHAT | Nurses (n = did not report) | Exam registration | NA | Portuguese | Yes | Three-point response scale (healthy, changes, and unhealthy) | If any category receives score 1 or 2, please organise for a dentist to examine the resident |
| Klotz et al., 2019 | OHAT | Caregivers (n = 4) | Exam registration | NA | German | Yes | Three-point response scale (healthy, changes, and unhealthy) | Sum of responses ranging from 0 to 16. If any category receives score 1 or 2, please organise for a dentist to examine the resident |
| Sahin et al., 2019 | OHAT | Nurses (n = 2) | Exam registration | NA | Turkish | Yes | Three-point response scale (healthy, changes, and unhealthy) | Sum of responses ranging from 0 to 16: 0 = very healthy to 16 = very unhealthy; If any category receives score 1 or 2, the professionals should be provided an exam by a dentist |
| Finotto et al., 2020 | OHAT | Nurses (n = 258) | Exam registration | NA | Italian | Yes | Three-point response scale (healthy, changes, and unhealthy) | Sum of responses ranging from 0 to 16. If any item scores 1 or 2, a visit to a dentist must be arranged |
| Arvidson-Bufano et al., 1996 | ohr-MDS | Nurses (n = 18) | Exam registration | NA | English | Yes | Three-point response scale (healthy, changes, and unhealthy) | Indication for further evaluation and/or treatment, is positive if any one of the six items is ticked |
| Blank et al., 1996 | ohr-MDS | Nurses (n = 18) | Exam registration | NA | English | Yes | Dichotomous scale (yes or no) | Indication for further evaluation and/or treatment is positive if any one of the six items is ticked |
| Hoben et al., 2016 | ohr-MDS | Nurses (n = did not report) | Exam registration | NA | English | Yes | Dichotomous scale (yes or no) | Did not report |

| Author-date | Oral health assessment | Target population (n) | Mode of administration | Recall period | Language | Training need | Scale | Scoring |
|---|---|---|---|---|---|---|---|---|
| Krausch-Hofmann et al., 2019 | ohr-interRAI | Caregivers (n = 23) | Interview and observation during meals or inspection of the mouth | 3 days prior to the assessment | Flemish-Dutch | Yes | Dichotomous scale (yes or no) | Did not report |
| Krausch-Hofmann et al., 2020 | ohr-interRAI | Caregivers (n = 4) | Exam registration | NA | Flemish-Dutch | Yes | Three-point response scale (healthy, changes and, unhealthy) | Interventions were recommended if one or more of the items were rated as nonacceptable |
| Kayser-Jones et al., 1995 | BOHSE | Registered nurses (RN), licensed vocational nurses (LVNs) e Certified nursing assistants (CNAs) (n = 16) | Exam registration | NA | English | Yes | Three-point response scale (healthy, changes and, unhealthy) | The final score was the sum of the scores from the 10 categories and ranged from 0 (very healthy) to 20 (very unhealthy) |
| Lin et al., 1999 | BOHSE | RNs, LVNs e CNAs (n = 16) | Exam registration | NA | English | Yes | Three-point response scale (healthy, changes, and unhealthy) | The final score was the sum of the scores from the 10 categories and ranged from 0 (very healthy) to 20 (very unhealthy). If the answer was "yes" for "need for treatment", the nurses were asked to choose a treatment action: within 2 months, 2 weeks, or emergency |
| Fjeld et al., 2016 | DHI | Nurses (n = 1) | Exam registration | NA | Did not report | Yes | Three-point response scale (healthy, changes, and unhealthy) | Add score for upper and lower jaw: 0 = Continue as usual; 1 = Check for deterioration and pay attention to difficult areas; 2–4 = Dental hygiene needs to improve |
| Tsukada et al., 2016 | OHSTNP | Caregivers (n = 12) | Exam registration | NA | English/ Japanese | No | Three-point response scale (healthy, changes, and unhealthy) | 0 to 24 |
| Yanagisawa et al., 2017 | OAS | Caregivers (n = 45) | Exam registration | NA | Did not report | Yes | Three-point response scale (healthy, changes, and unhealthy) | 0 to 18 |

TABLE 3 Items used in oral health assessment according to instruments

| Items of the instruments | OHAT | ohr-MDS (Hoben) | Ohr-interRAI 2020 | BOHSE | DHI | OHSTNP | OAS |
|---|---|---|---|---|---|---|---|
| Lips | ✓ | | | ✓ | | ✓ | |
| Tongue | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Gums and tissues | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Saliva | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| Natural teeth | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Denture | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| Oral cleanliness | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Dental/mouth pain | ✓ | ✓ | ✓ | | | | |
| Chewing problems | | ✓ | ✓ | ✓ | | | ✓ |
| Lymph nodes | | | | ✓ | | | |
| Articulation/Jaw opening | | | | | | ✓ | ✓ |
| Coughing | | | | | | ✓ | |
| Breath | | | | | | | ✓ |

FIGURE 3 Measurement properties of the instruments used for oral health evaluation of older adults assessed by each study. Green + symbol indicates an assessed measurement property.

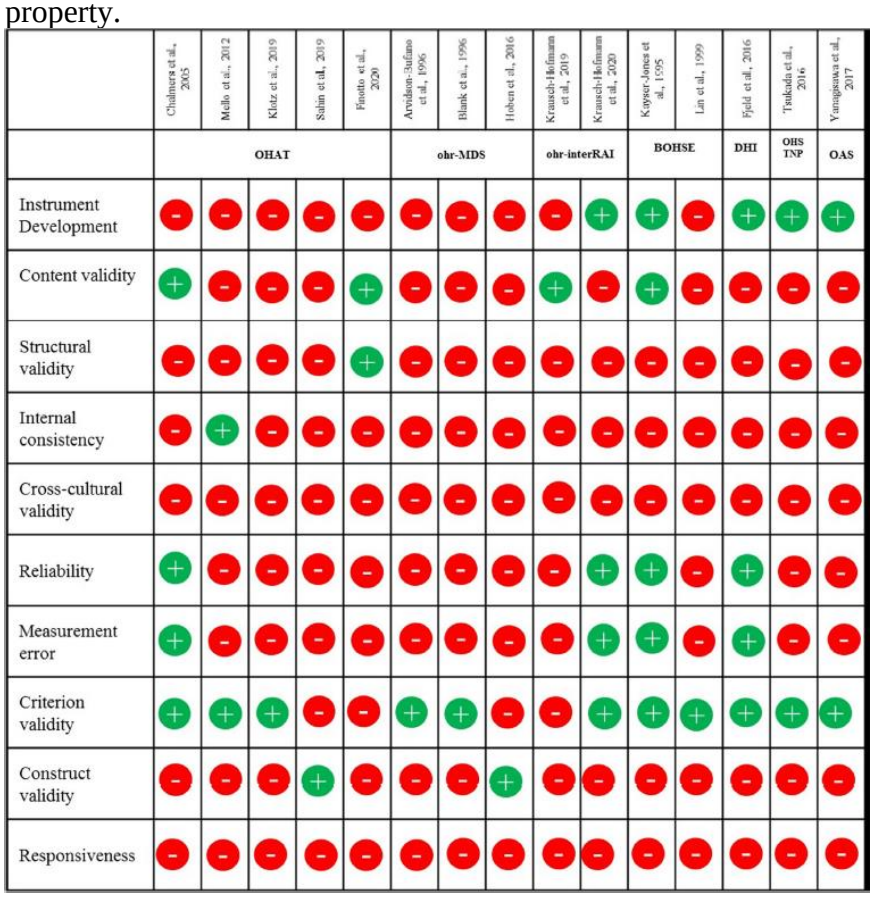

TABLE 4 Criteria content validity overall rating, and the grading of the quality of evidence

| Instrument | Overall rating | Quality of evidence |
|---|---|---|
| *OHAT* | | |
| Relevance | Sufficient | Moderate |
| Comprehensiveness | Sufficient | Moderate |
| Comprehensibility | Sufficient | Moderate |
| *DHI* | | |
| Relevance | Insufficient | Very Low |
| Comprehensiveness | Insufficient | Very Low |
| Comprehensibility | Insufficient | Very Low |
| *BOHSE* | | |
| Relevance | Insufficient | Very Low |
| Comprehensiveness | Sufficient | Very Low |
| Comprehensibility | Insufficient | Very Low |
| *OAS* | | |
| Relevance | Insufficient | Very Low |
| Comprehensiveness | Insufficient | Very Low |
| Comprehensibility | Insufficient | Very Low |
| *OHSTNP* | | |
| Relevance | Insufficient | Very Low |
| Comprehensiveness | Insufficient | Very Low |
| Comprehensibility | Insufficient | Very Low |
| *Ohr-interRAI* | | |
| Relevance | Sufficient | High |
| Comprehensiveness | Sufficient | High |
| Comprehensibility | Sufficient | High |

TABLE 5 Summary of Findings (SoF) per measurement property, including the pooled or summarised results of the measurement properties, overall rating (i.e., sufficient (+)/insufficient (−)/inconsistent (±)/indeterminate (?)), and the grading of the quality of evidence (high, moderate, low, and very low)

| Measurement properties | Summary or pooled result | Overall rating | Quality of evidence |
|---|---|---|---|
| *OHAT* | | | |
| Structural validity | Kaiser–Meyer–Olkin = 0.867; Bartlett's sphericity test ($\chi 2 = 814.64$, df = 36, $P < .000$)* | Indeterminate | Moderate |
| Internal consistency | Cronbach's alpha: 0.40 | Insufficient | Moderate |
| Reliability | ICC = 0.78; range of Kappa by item = 0.51 to 0.71 | Sufficient | Very low |
| Measurement error | % agreement + = 74.4 to 96.6 | Sufficient | Very low |
| Criterion validity | Inconsistent results: Pearson correlation (range): −0.1 to −0.94; ICC = 0.602 (95% CI 0.282 to 0.811) | Insufficient | Low |
| Construct validity—hypotheses testing | Oral health condition was significantly different between groups: the top group was comprised of the highest 27% of the scores whilst the bottom group was comprised of the lowest 27% of the scores. Significant t test ($P < .05$) except for dental pain | Sufficient | Moderate |
| *ohr-MDS* | | | |
| Criterion validity | Does not report | Indeterminate | Very low |
| Construct validity—hypotheses testing | Associations among oral/dental problems and other resident characteristics with mouth pain, dental problems and periodontal problems over time using General Estimating Equation Models. Severe underdetection of oral/dental problems and lack of association of well-known oral health predictors with oral/dental problems suggest validity problems of ohr-MDS | Insufficient | High |
| *ohr-InterRAI* | | | |
| Reliability | Simple Kappa by item = 0.45 to 0.93 | Insufficient | Very low |
| Measurement error | % agreement + = 80.0 to 96.4 | Sufficient | Very low |
| Criterion validity | Inconsistent: studies used different statistical strategies (ICC and Pearson correlation) | insufficient | Very low |
| *BOHSE* | | | |
| Reliability | Simple Kappa by item 0.79 to 0.88 | Sufficient | Very low |
| Measurement error | % agreement + = 50.5 to 98.0 | Insufficient | Very low |
| Criterion validity | Meta-analysis Pooled Spearman's correlation coefficient: 0.56 (CI 0.46 to 0.67) | Insufficient | High |
| *DHI* | | | |
| Reliablity | Simple Kappa by instrument 0.80 | Sufficient | Very low |
| Measurement error | Mic not defined | Indeterminate | Very low |
| Criterion validity | Spearman's correlation coefficient: 0.78 | Sufficient | Low |
| *OHSTNP* | | | |
| Criterion validity | Spearman's correlation coefficient: 0.81 to 0.84 | Sufficient | Low |
| *OAS* | | | |
| Internal consistency | Cronbach's alpha: 0.72 | Sufficient | Moderate |
| Criterion validity | ICC: 0.54 to 0.89 | Insufficient | Low |

Abbreviations: ICC, intraclass correlation coefficient; MIC, minimal important change.
*Exploratory factor analysis.

FIGURE 4 Forest plot displaying a random-effects meta-analysis of Spearman's correlation coefficient for Criterion validity of BOHSE