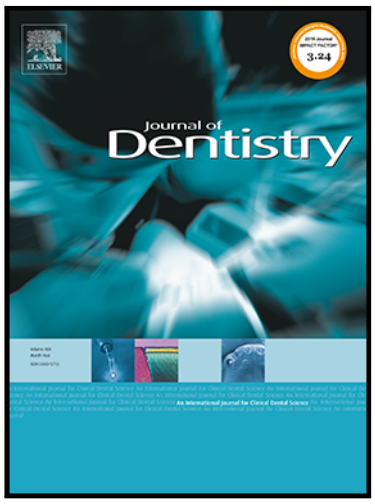

Diagnostic accuracy of on-scan assessments compared to clinical assessments using a periodontal probe for detecting gingival recession: A cross-sectional study

N.E West , M Wright , S Daly , R.G Newcombe , M Davies , M Kuralt , N.X West

PII: S0300-5712(26)00177-6
DOI: https://doi.org/10.1016/j.jdent.2026.106504
Reference: JJOD 106504

To appear in: *Journal of Dentistry*

Received date: 20 November 2025
Revised date: 8 January 2026
Accepted date: 10 January 2026

Please cite this article as: N.E West , M Wright , S Daly , R.G Newcombe , M Davies , M Kuralt , N.X West , Diagnostic accuracy of on-scan assessments compared to clinical assessments using a periodontal probe for detecting gingival recession: A cross-sectional study, *Journal of Dentistry* (2026), doi: https://doi.org/10.1016/j.jdent.2026.106504

Diagnostic accuracy of on-scan assessments compared to clinical assessments using a periodontal probe for detecting gingival recession: A cross-sectional study

West N.E[a], Wright M[a], Daly S[a] Newcombe R.G[b], Davies M[a], Kuralt M[c], West N.X[a]
[a] Periodontology, Clinical Trials Unit, Bristol Dental School, Lower Maudlin Street, Bristol, UK
[b] Institute of Primary Care and Public Health, Cardiff University, Cardiff, UK
[c] 3 Shape, Copenhagen, Denmark.

Author job titles and email addressed

Corresponding author:		
[a]Natasha E. West	Clinical research fellow	natasha.west@bristol.ac.uk
Co-authors:		
[a]Matthew Wright	Clinical research fellow	Ig22564@bristol.ac.uk
[a]Sinead Daly	Clinical research fellow	s.daly@bristol.ac.uk
[b]Robert Newcombe	Professor of Biostatistics	newcombe @cardiff.ac.uk
[a]Maria Davies	Clinical Research Manager	maria.davies@bristol.ac.uk
[c]Marko Kuralt	Principal Clinical Research Manager	marko.kuralt@3shape.com
[a]Nicola X. West	Professor/Honorary Consultant in Restorative Dentistry	n.x.west@bristol.ac.uk

Short title: Diagnostic accuracy of on-scan assessments of gingival recession

Corresponding author

Natasha E. West
Restorative Dentistry (Periodontology)
Clinical Trials Unit
Bristol Dental School
Lower Maudling Street, BRISTOL BS1 2LY
Tel: (0117) 342 9638
Facs: (10117) 342 4100

Key words: gingival, recession, intraoral, scan, digital, clinical, accuracy

Corresponding author statement:
I confirm that the submitted work is original, not currently under consideration for publication elsewhere, and is in compliance with all rules stipulated by the Journal of dentistry.

Acknowledgements

None

Trial registration:

This trial was registered in the ISRCTN registry (ISRCTN62574906). ISRCTN - ISRCTN62574906: Investigating how assessing gingival recession on 3D scans of teeth compares to dentists' standard clinical assessment

Abstract

Aim: Compare the diagnostic accuracy of measuring gingival recession (GR) using a digital ruler on intraoral scans with clinical measurements using a UNC-15 probe. **Methods:** A prospective, observational, examiner-blinded, single-arm, cross-sectional, single-site study in adults aged 18+ with minimum of 20 teeth. All participants underwent an intraoral scan. Clinical GR was recorded on 6-sites (non-molar teeth) and 8-sites (molar teeth) by the principal examiner. Anonymised scans were assessed by the principal examiner twice (Ex1a, 1b) and two additional examiners (Ex2 and 3). **Results:** 109 participants were assessed; all had at least one site of buccal GR. Ex1a on-scan measurements compared to clinical measurements showed 66.9% sensitivity and 92.5% specificity. There were considerable differences between central and proximal sites, central sites showed greater sensitivity 77.0% and specificity 81.4% compared to proximal sites 38.5% and 97.2% respectively. Diagnostic precision was robust, agreement for intra-examiner variability (Examiner-1a vs 1b) was 98.5% (95% CI 98.1%-98.8%). Inter-examiner variability was 78.9% (95% CI, 77.0%-80.6%) for Ex-1a-Ex-2 and 86.2% (95% CI, 84.8%-87.5%) for Ex-1a-Ex-3. **Conclusion:** Intraoral scan measurements are reproducible and can be utilised with accuracy. Measuring GR clinically is the current gold standard despite the limitations, intraoral scans can be utilised to overcome these limitations.

Clinical Significance:

GR impacts aesthetics and precludes conditions such as dentine hypersensitivity. Completing a full mouth record for gingival recession is time-consuming and accuracy is hindered by soft tissue and angulation constraints. Recording GR on an intraoral scan is highly reproducible and although not identical to clinical measurement, it is arguably quicker for the patient, visualises the cervical area and can be utilised with accuracy and reproducibility for monitoring purposes.

List of tables and figures:

Introduction

Gingival recession (GR) is a highly prevalent condition affecting teeth and surrounding soft tissues [1, 2]. Characterised by apical shift of the gingival margin in relation to the cemento-enamel junction (CEJ) resulting in mucogingival deformity, exposing the tooth root surface to

the oral environment [1]. The aetiology is unknown [1, 3] however several aetiological factors have been identified; anatomical variation - gingival phenotype and prior lack of alveolar bone, bone remodelling after orthodontic treatment, trauma, smoking, gingival inflammation, periodontitis and successful periodontal treatment. [4-10]. All can occur alone or in combination.

GR frequently occurs in adults. Recent epidemiological research [11] confirmed a high prevalence. 87.9% of participants had at least one site with GR 1 mm or greater. A systematic review [12] using the same definition reported a pooled global prevalence estimate of 85%. The clinical significance of GR relates to aesthetic and function, including dentine hypersensitivity [1]. Exposed root increase susceptibility to carious and non-carious cervical lesions [1, 13, 14].

Primary assessment of GR relies on identifying the sites and quantifying extent, by measuring the distance between the CEJ and gingival margin in the coronal-apical direction (recession depth). Clinically, GR depth is measured with a periodontal probe demarcated in one millimetre increments, an essential instrument for visual-tactile dental examination [15], enabling assessment at a clinically acceptable level [16] [17]. Identified limitations include variations in probe position and angulation, to CEJ identification and rounding errors to whole mm [18-21]. These constraints encouraged researchers to develop novel approaches to assess and measure GR. Analogue methods include stents, digital callipers, endodontic spreaders [22, 23]. Image-based methods focused on intraoral photographs [24-26]. Latterly scanning and 3D models [16, 27, 28] have improved measurement accuracy, eliminating variations in position and angulation of the digital ruler and removed rounding of measurements up [29, 30].

Early, accurate diagnosis is essential for implementing preventive and restorative measures. Novel digital methods primarily focused on treatment outcomes, including evaluating root coverage surgery or monitoring GR [29]. Routine acquisition of intraoral scans for screening and diagnostic purposes has integrated these workflows into everyday clinical practice. Larger cross-sectional studies are necessary [31] to enhance our understanding of the benefits and limitations of intraoral scans compared to the current clinical standard.

This study aimed to compare on-scan GR assessment (index test) with clinical examination using a periodontal probe (reference standard). The objective was to evaluate diagnostic accuracy, precision, and measurement agreement of the index test used by single and multiple examiners relative to the reference standard.

Methods

Study Design

A prospective, observational, blinded, single-arm, cross-sectional, single-site study. Designed to investigate how assessing GR on 3D intraoral scans of teeth compares to dentists' standard clinical assessment (**Figure 1**). Ethical approval was obtained from the London - Riverside Research Ethics Committee (24/PR/0088). All participants provided written informed consent. The study was registered in the ISRCTN registry (ISRCTN62574906). The study was conducted in compliance with ISO 14155:2020 – Clinical investigation of medical devices for human subjects, in accordance with the principles of Good Clinical Practice, and the STARD 2015 guidelines for reporting diagnostic accuracy studies [32].

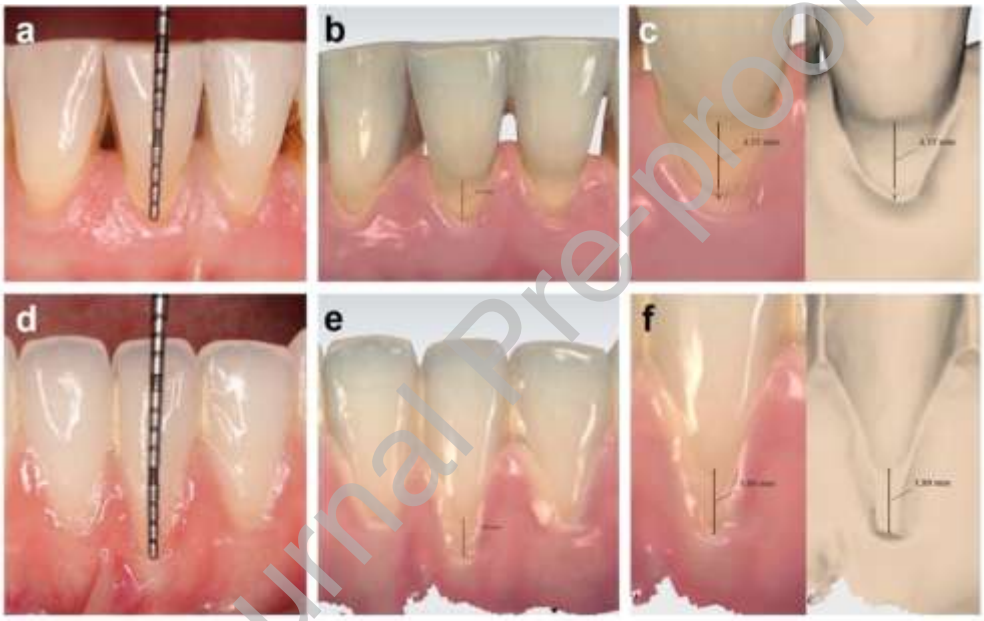

Figure 1. Clinical photo and intraoral scan of gingival recession measurements.

Participants & Eligibility

A consecutive/convenience sample of healthy volunteers with varying degrees of GR ranging from no recession to severe recession were recruited from the University of Bristol Dental Clinical Trials Unit database and individuals responding to advertisements. Following written informed consent, the principle dentist evaluated and screened all participants for study eligibility before enrolment.

Participants underwent medical history and demographic documentation, eligibility screening and oral assessment of hard and soft tissues. Eligible participants were adults aged 18+ years, able and willing to consent, in good general health without conditions posing risk or

affecting data quality, having at least 20 natural teeth excluding third molars. Exclusion criteria included significant diseases or disorders that could risk participant safety, influence study results or impair ability to participate.

Clinical workflow

Following enrolment, the primary examiner (Ex1) scanned the participant's mouth with an intraoral 3D scanner (3Shape TRIOS 5®, 3Shape TRIOS A/S, Copenhagen, Denmark) following the manufacturer's instructions. Ex1 then assessed GR clinically (visual and tactile) using a UNC-15 periodontal probe® (HuFriedy, Germany). This was the end of participant involvement.

Reference standard

The reference standard for measuring recession depth and, therefore, detecting the presence or absence of GR was clinical assessment using a periodontal probe (UNC-15) by the trained and calibrated Ex1. Prior to participant recruitment, Ex1 conducted simulated clinical exercises with Examiner 2 (Ex2). The two examiners differed by one scale point (1mm) for 90/630 (14.3%) of the sites examined.

GR was assessed and recorded in millimetres at six sites, mesio-buccal (MB), central-buccal (CB), disto-buccal (DB), mesio-lingual (ML), central-lingual (CL), and disto-lingual (DL), for anterior and premolar teeth. For molar teeth, eight sites, MB, central-mesio-buccal (CMB), central-disto-buccal (CDB), DB, ML, central-mesio-lingual (CML), central-disto-lingual (CDL), and DL, were recorded (**Figure 2**). The number of recession sites scored was in accordance with standard clinical procedure for specialist or research purposes. The molars had two additional sites recorded due to their increased complexity in root morphology. If root dentine was visualised, the recording was rounded up to the nearest mm, including when <1mm.

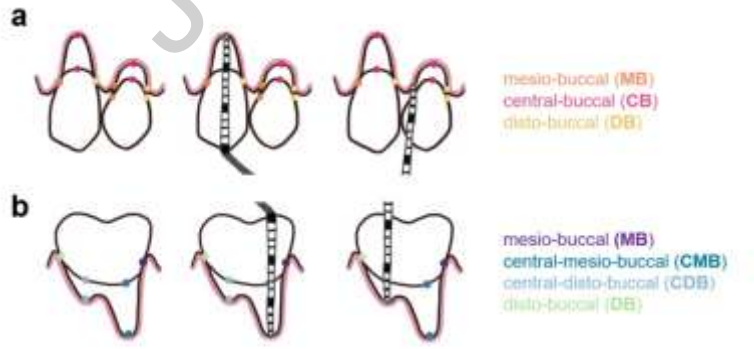

Figure 2. Gingival recession measurement sites – for anterior and premolar teeth, six sites, and for molar teeth, eight sites, were assessed. Only the buccal surface of a non-molar (a) and a molar tooth (b) is visualised, where coloured dots illustrate calibrated sites where measurements were recorded. The same sites apply to palatal/lingual surfaces.

Figure 2. Gingival recession measurement sites – for anterior and premolar teeth, six sites, and for molar teeth, eight sites, were assessed. Only the buccal surface of a non-molar (a) and a molar tooth (b) is visualised, where coloured dots illustrate calibrated sites where measurements were recorded. The same sites apply to palatal/lingual surfaces.

Index test

Ex1 performed visual on-scan assessment of GR (1a), at least two weeks after clinical assessment, without access to the clinical recordings. At least two weeks later, the same Ex1 repeated the assessments (1b). Additionally, two independent examiners blinded to clinical findings, (Ex2 and Ex3) assessed the same images. Intraoral scans were anonymised randomly for all assessments.

On-scan assessments were performed using a measurement tool in custom-designed software (not yet commercially available, 3Shape TRIOS A/S, Copenhagen, Denmark) on the same sites as the reference standard. The software enabled visualisation (including with/without colours and turning on/off specularity) and interaction (including rotation, translation, and zooming in/out) of the scans in the same way as commercially available software in the 3Shape ecosystem, with storage of the measurements enabled. All examiners were trained to use the software. Measurements were recorded in millimetres with two decimals.

Statistical Analysis

The primary outcome was the diagnostic sensitivity and specificity at site level of the on-scan assessment (Ex1a) to predict Ex1's clinical score, both measures dichotomised as zero or positive. The pre-defined minimally acceptable performance goal for both measures is 75%. For comparisons with clinical scoring, on-scan measurements were rounded up to the nearest millimetre. Confidence intervals for sensitivity and specificity were calculated by bootstrapping.

For diagnostic precision, repeated assessments by the same examiner, 1a and b, were used to evaluate intra-examiner variability. Assessments by different examiners, Ex1, Ex2, Ex3, were used to assess inter-examiner variability. Chance-corrected agreement was characterised by Scott's pi [33] with 95% confidence intervals [34].

Measurement agreement between clinical and on-scan assessments was evaluated using Bland-Altman analysis [35] showing bias and 95% limits of agreement. On-scan measurements were not rounded to the nearest millimetre for this analysis. A predefined accepted limit of agreement is ±2 mm [16, 17, 28]

Sample Size Calculation

Sample size was estimated using the prevalence estimate, expected diagnostic sensitivity and specificity, and the acceptable half-width of the 95% confidence interval [36]. Each

measurement site represents an individual statistical unit, assuming within-participant variation reflects between-participant variation, allowing multiple sites per participant. Due to clustering (site, tooth, and participant), all levels were considered, with the participant level being the limiting factor, resulting in a planned 109 participants contributing with ≥20 teeth each (total at least 2180 teeth and 13080 sites). We anticipated sensitivity and specificity of 85%, and a prevalence of 55%. With a sample size of 109 subjects, 60 are then expected to be positive, 49 negative, and the anticipated 95% confidence interval for specificity is 75% to 95%, the interval for sensitivity being narrower. To mitigate age-related bias in GR [11], participants were evenly distributed across five age groups: 18-27, 28-37, 38-47, 48-57, and 58+.

Results

Demographics

116 healthy adult participants were recruited between 6 March and 18 June 2024. Results from the first 7 participants were excluded due to change of study clinician from illness and inability to continue study conduct.. Data from 109 participants was analysed. No adverse events were recorded. **Table 1** summarises demographic characteristics.

***Table 2**. Demographic Data*

	Frequency	Percentage
Gender		
Female	79	72.5
Male	30	27.5
Smoking status		
Smoker	5	4.6
Former smoker	25	22.9
Non-smokers	79	72.5
Ethnic group		
White	77	70.6
Mixed	3	2.8
Asian	14	12.8
Black	13	11.9
Other	2	1.8
	Range	Mean (SD)
Age (years)	19 – 74	41.3 (13.7)

Clinical characteristics

With site level being of interest, only sites measured by all examiners were used for further analysis. Analyses were based on scores for 15574 (77.7%) of a possible 20056 sites, after excluding sites with clinical or any scan reading unavailable (including missing teeth, saliva, staining and calculus hindering visualisation)

All participants had at least one site of buccal GR, and only 3 participants did not exhibit GR at any lingual/palatal site. Maximum buccal and lingual GR scores of 1mm were

seen in 20.2% and 56.0% participants, respectively. Maximum buccal and lingual GR scores of 2mm or greater were seen in 79.8% and 41.3% participants, respectively. This cohort exhibited few participants with GR greater than 5mm. Most of the cohort were periodontally healthy and were not susceptible to periodontitis.

GR was more frequent on buccal surfaces than on the corresponding lingual surfaces, except in the lower anterior region (**Figure 3**). GR was most prevalent at the CB site in non-molar teeth, and CMB site in molar teeth (**Table 2**).

Figure 3. Grouped GR frequency graphs.

Table 2. *Site-level mean & standard deviation GR.*

Teeth	**Tooth sites**							
Non-molar	**MB**		**CB**	**DB**	**ML**		**CL**	**DL**
Measured sites	1981		1992	1988	1766		1815	1779
Mean (mm)	**0.19**		**0.80**	**0.17**	**0.15**		**0.42**	**0.17**
(SD)	(0.51)		(0.93)	(0.48)	(0.43)		(0.62)	(0.49)
Molar	**MB**	**CMB**	**CDB**	**DB**	**ML**	**CML**	**CDL**	**DL**
Measured sites	721	730	712	703	650	653	655	650
Mean (mm)	**0.16**	**0.90**	**0.62**	**0.20**	**0.17**	**0.59**	**0.33**	**0.12**
(SD)	(0.43)	(0.86)	(0.87)	(0.53)	(0.45)	(0.78)	(0.63)	(0.40)

On-scan assessments compared to clinical assessments

The frequency of GR measurements, as measured by each examiner, is shown in **Figure 4a.** On-scan data were rounded up to the nearest mm allowing direct comparison with clinical scoring. In this cohort, most sites showed no GR, 1mm was the commonest among recession sites. Generally, differences between clinical and on-scan GR scores were fairly balanced, with on-scan scores only slightly lower than the corresponding clinical scores - except for Ex2 who over-estimated considerably (**Figure 4b)**.

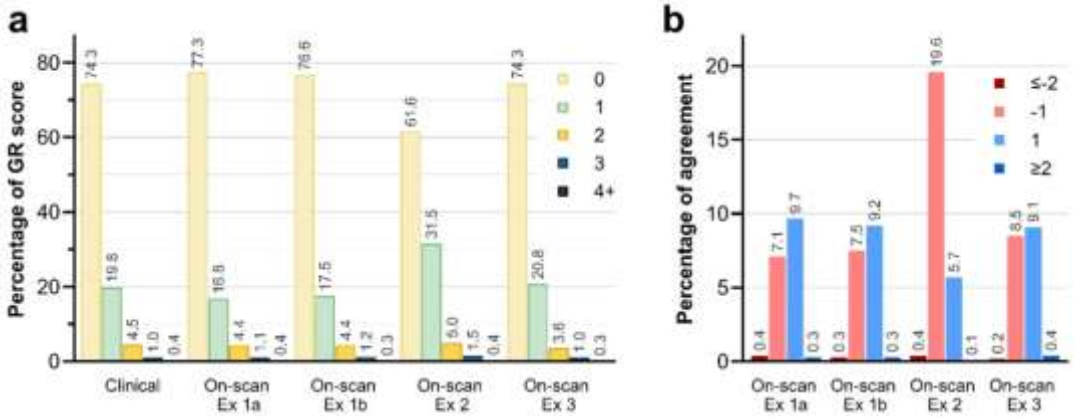

Figure 4. Comparison of on-scan assessments with clinical assessment of GR measurements for all examiners, showing the distribution of GR scores (a) and percentages of over- and under-assessment (b). On-scan overestimates are shown in red and underestimates in blue. n=15574 sites.

Figure 4. Comparison of on-scan assessments with clinical assessment of GR measurements for all examiners, showing the distribution of GR scores (a) and percentages of over- and under-assessment (b). On-scan overestimates are shown in red and underestimates in blue. n=15574 sites.

Diagnostic accuracy and precision of on-scan assessments

Sensitivity and specificity with CI for all examiners are shown in **Table 3**. The sensitivity and specificity for on-scan assessments performed by Ex1a were 66.9% & 92.5%. There are considerable differences between central and proximal areas, with sensitivity and specificity 77.0% & 81.4% for central sites and 38.5% & 97.2% for proximal sites.

Regarding diagnostic precision, agreement for intra-examiner variability (Ex1a vs 1b) is 98.5% (95% CI 98.1%-98.8%), and Scott's pi is 0.96 (95% CI 0.95-0.97). While the crude agreement for inter-examiner variability is 78.9% (95% CI, 77.0%-80.6%) and 86.2% (95% CI, 84.8%–87.5%) for Ex1a-Ex2 and Ex1a-Ex3, respectively, adjustment for chance-corrected agreement reduces these figures to a Scott's pi of 0.50 (95% CI 0.46-0.55) and 0.62 (95% CI 0.59–0.66) for these pairs.

Measurement agreement of on-scan and clinical assessments

Bland-Altman plots showing measurement agreement between clinical and on-scan measurements made by Ex1a are shown in **Figure 5a and b,** and for on-scan measurements by Ex1a and Ex2 in **Figure 5c and d.**

Table 3. *Diagnostic accuracy measures of on-scan assessments as the index test and clinical assessments as the reference standard with bootstrap confidence intervals.*

Series & comparison	Sensitivity	95% CI	Specificity	95% CI
All tooth sites (n=15574)				
Ex1a → clinical	66.9%	63.4% - 70.4%	92.5%	91.1% - 93.9%
Ex1b → clinical	68.4%	65.1% - 71.7%	92.1%	90.6% - 93.6%
Ex3 → clinical	70.7%	66.7% - 74.6%	89.9%	88.4% - 91.3%
Ex2 → clinical	82.8%	80.7% - 84.9%	76.9%	74.6% - 79.2%
Central tooth sites (n=6938)				
Ex1a → clinical	77.0%	73.6% - 80.3%	81.4%	78.7% - 84.1%
Ex1b → clinical	79.0%	76.0% - 82.0%	79.8%	76.9% - 82.7%
Ex3 → clinical	76.5%	72.7% - 80.3%	79.5%	76.9% - 82.2%
Ex2 → clinical	93.7%	92.6% - 94.9%	47.5%	43.0% - 52.0%
Proximal tooth sites (n=9176)				
Ex1a → clinical	38.5%	29.0% - 48.1%	97.2%	95.8% - 98.6%
Ex1b → clinical	38.6%	28.9% - 48.4%	97.3%	95.8% - 98.7%
Ex3 → clinical	54.4%	46.3% - 62.5%	94.3%	92.9% - 95.6%
Ex2 → clinical	52.1%	43.6% - 60.6%	89.4%	87.4% - 91.3%

* **Green** - sensitivity or specificity is statistically significantly higher than the target 0.75; **yellow** - point estimate above but lower limit below 0.75, does not reach statistical significance; **no colour** - sensitivity or specificity is statistically significantly lower than the target 0.75

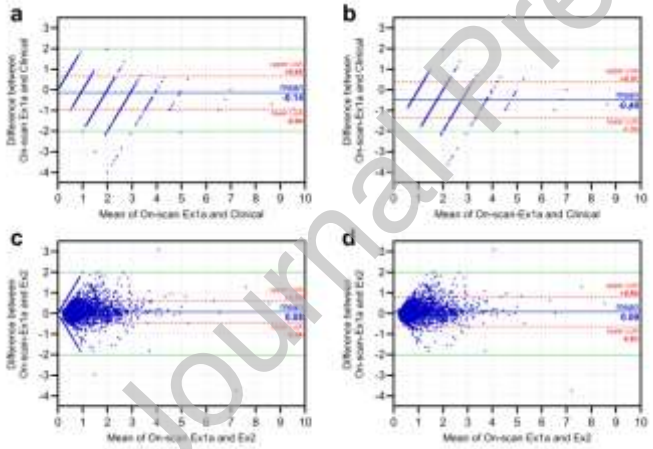

Figure 5. Bland-Altman plots for measurement agreement between on-scan assessment by Ex1a with clinical assessment (plots a and b) and on-scan assessments by Ex2 (plots c and d). Plots b and d present the same comparison as plots a and c except that only sites with evidence of GR are considered from both sets, sites with either score zero were excluded.

Discussion

This study compared on-scan GR assessment with clinical examination using a periodontal probe to evaluate diagnostic performance across single and multiple examiners. On-scan assessment showed acceptable diagnostic accuracy, especially at central sites. Precision

was high for intra-examiner repeatability, reflecting consistent on-scan measurements by the primary examiner, while inter-examiner precision ranged from moderate to substantial, indicating variability among examiners. Measurement agreement met expectations, with narrower-than-anticipated limits of agreement, overall supporting the consistency of on-scan assessment with clinical despite methodological differences between them.

The present study is the first diagnostic accuracy study comparing on-scan and clinical assessments of GR. There is limited literature on the diagnostic accuracy of current methods, as previous studies have mainly focused on sites with GR and have primarily addressed measurement agreement [16, 17, 28, 37]. It has been suggested that for a diagnostic test to be useful, the sum of sensitivity and specificity should be at least 150%, halfway between 100%, no diagnostic information, and 200%, completely accurate [38]. In the present study, when sensitivity and specificity values were combined, the target of 150% was exceeded. To balance the two, individual targets of 75% were set for both sensitivity and specificity, as this was deemed the minimum acceptable for false positives and negatives. Across all sites in this sample, a general trend was observed, of underdiagnosis of GR on scans, characterised by reasonable specificity but sensitivity not reaching the 75% threshold. The overall underdiagnosis stems predominantly from proximal sites, which have very high specificity but low sensitivity. Central sites, which are the most commonly assessed for GR, display more balanced sensitivity and specificity, both achieving the predefined target of 75%.

Using novel methods for established clinical situations requires careful adaptation, with an understanding of the methodological differences. With on-scan assessment, examiners rely solely on the visual aspect of identifying the CEJ, thereby missing the probe's tactile element. We facilitated this by an enhanced visual on-scan examination, allowing examiners to view all captured angles of the tooth without soft tissue impairing vision or access This was particularly useful for second molars and lingual views, in addition the ability to zoom into areas of interest and view with greater detail. Provision of filters on the software allowed examiners to turn colour on/off, aiding CEJ detection and gingival margin position, enhancing their ability to determine landmarks. However, there are also differences in the perceived colour changes on the scan from the enamel to the dentine surfaces., An extremely subtle change in colour and texture compared to what can be seen clinically may make the CEJ difficult to identify sometimes. Even clinically, it is challenging to detect GR <0.5mm accurately. If the examiner zooms in more than it is possible to visualise clinically with magnification, this can lead to false positives resulting in overestimation of GR on-scan compared to clinical assessment, as observed with Examiner 2. Individual examiners can be trusted to produce consistent results, demonstrated by the high intra-examiner reliability of on-scan assessments. Overall, the agreement was reasonable, albeit

imperfect, with Examiner 2 emerging as a modest outlier demonstrated by moderate-substantial inter-examiner variability. Since this is a novel approach, reduced inter-examiner variability may be addressed through further calibration for the on-scan assessments to align examiners' interpretations, which was not performed in the present study.

Limited measurement resolution and the presence of rounding with clinical GR measurements are significant differences compared to on-scan measurements. On-scan digital measurements can be recorded to a tenth of a millimetre or less, whereas clinical measurements cannot be recorded below 1mm with the global standard UNC-15 probe [39]. Limited measurement resolution to whole numbers leads to rounding. The effects of simple mathematical rounding have been well demonstrated [40] rendering different conclusions regarding root coverage outcomes. However, the rounding done unconsciously during the examination doesn't necessarily follow mathematical rules. E.g. this is highly relevant to the 0-1mm range, where if the examiner detects GR, even if it is smaller than 0.5mm, they will prefer to round up to 1 indicating there is GR present, instead of assigning zero, which would indicate no GR present. This corresponds to our observed tendency for clinical measurements to be higher than the on-scan ones, as observed in previous measurement agreement studies using Bland-Altman analysis [28]. The present study results show similar bias of around 0.5mm as in the study by Kuralt et al. [16] and shifted 95% limits of agreement towards overestimation of clinical measurements compared to on-scan ones, but with a narrower range (+0.37mm to -1.33mm vs. +0.84mm to -1.88mm). Pre-defined limits of agreement of 2mm, while consistent with prior literature ([16, 17, 28]), may exceed what would be considered clinically optimal. This threshold reflects current clinical measurement practices and the state of existing methodologies rather than ideal clinical target. Similar oblique patterns, due to a mismatch in resolution, were observed on the Bland-Altman plots as before [16, 17, 28, 37] which, together with the high prevalence of zeros and simply the enormous sample size, raise questions about the appropriateness of use of this methodology for displaying measurement agreement for such datasets. The effect of the presence of zeroes in the measurement agreement results is evident when zeroes are removed (**Figure 5**), but due to the overlap of all the zero points on the plot, the plot doesn't accurately represent how many values are clustered there. The methodology is more suitable for comparing continuous on-scan measurements, which again yield similar results to those observed in previous studies [16]. However, the sheer number of overlapping points on the plot, especially those with one or both readings zero, renders visual interpretation of the pattern practically impossible. This calls for future research to develop appropriate methods for assessing measurement agreement in scenarios like this, especially since digitalisation has been very prominent recently.

The study cohort is the largest collected to date, aiming to assess the diagnostic accuracy of on-scan GR assessment, which also enabled some epidemiological insights. Previous agreement studies in the area focused on recession-affected teeth (usually 2mm+) in smaller sample sizes measuring central sites. Compared to the present study, which included 109 participants with 15574 sites of varying degrees of recession. The prevalence in this study was higher than in previous studies [11] 87.9%, [12] 85%. GR is highly prevalent both in individuals with excellent oral health [41] and those susceptible to periodontitis [10]. Despite this cohort being divided equally across the age categories the population generally consisted of participants with excellent oral hygiene and good oral health, potentially explaining the high prevalence. At the site level, a high percentage of sites showed no GR or GR of 1mm, likely determined by the presence or absence of alveolar bone at these sites in this orally healthy population.

GR was more prevalent on buccal surfaces than the corresponding lingual surfaces. The high level of GR in upper premolars corresponds to findings from a previous study where GR was particularly prevalent at premolars and molars and the majority of recession defects appeared to be in healthy mouths [42], most likely due to presence or absence of underlying bone [5]. GR has been reported more frequently in the upper arch, where the highest prevalence of GR occurred in the canine and first premolar region almost approaching 100% [43], similar to the present study. There was a reported a regional shift to increased GR to the left side of the mouth [43] which corresponds to tooth 25, the most common site with GR in this study. Potentially due to increased predisposition, where the buccal bone plate has premolar fenestration resulting in a reduced buccal bone crest thickness [44].

The present study employed a robust design with multiple examiners to conduct a comprehensive assessment of accuracy and precision. A periodontal probe was used as an established reference standard to test on-scan assessments with three independent, blinded examiners, enabling evaluation of intra- and inter-examiner variability. However, the study also has some limitations. A relatively large proportion of sites were not measurable, with some inconsistency between clinical and on-scan measurements and between examiners. To minimise this impact, analysis used only sites with a full set of measurements for all examiners. The scan quality was excellent, but sites may not have been measured on-scan due to visible measurement landmarks being obscured by staining, plaque or calculus, which were not always as apparent as clinically. Scans were acquired before clinical examination, and removal of these contaminants is not feasible on the scan, as it presents a screenshot at a specific time point, whereas clinically, these can be removed and assessed. Similar to periodontal probe position challenges, on-scan measurements can also be affected by the viewing angle or skewed by a tilt on the z-axis within the 3D model,

increasing the likelihood of over/underestimation, especially for lingual/palatal GR compared to clinical measurements. Clinically recording measurements in the presence of saliva, and soft tissue retraction had its own challenges and although these methods are being compared the differences in data collection are very different and have their own limitations.

Conclusions

When clinicians measure GR in general dental practice, typically only central sites are measured. It is time-consuming to complete a full mouth record, and accuracy is hindered by soft tissue and angulation constraints. Most importantly, GR is measured to the nearest millimetre, and unless a stent is used to guide the probe to the exact location, reproducibility can be erratic. Whilst clinical GR measurement is regarded as the gold standard, it has many limitations which can be overcome with scans. The scan is highly reproducible and can be measured off-site later. Whilst not identical to clinical measurement, it is arguably quicker for the patient, captures a 3D image of the lesion, can be overlaid with templates, and can be utilised with accuracy and reproducibility for monitoring changes in the cervical area.

References

[1] P. Cortellini, N.F. Bissada, Mucogingival conditions in the natural dentition: Narrative review, case definitions, and diagnostic considerations, Journal of Periodontology 89 Suppl 1 (2018) S204-S213.

[2] G. Pini Prato, Mucogingival deformities, Ann Periodontol 4(1) (1999) 98-101.

[3] S. Mythri, S.M. Arunkumar, S. Hegde, S.K. Rajesh, M. Munaz, D. Ashwin, Etiology and occurrence of gingival recession - An epidemiological study, J Indian Soc Periodontol 19(6) (2015) 671-5.

[4] D.M. Kim, S.H. Bassir, T.T. Nguyen, Effect of gingival phenotype on the maintenance of periodontal health: An American Academy of Periodontology best evidence review, Journal of Periodontology 91(3) (2020) 311-338.

[5] P.J. Watson, Gingival recession, J Dent 12(1) (1984) 29-35.

[6] A. Khocht, G. Simon, P. Person, J.L. Denepitiya, Gingival recession in relation to history of hard toothbrush use, Journal of periodontology 64(9) (1993) 900-905.

[7] S. Jepsen, J.G. Caton, J.M. Albandar, N.F. Bissada, P. Bouchard, P. Cortellini, K. Demirel, M. de Sanctis, C. Ercoli, J. Fan, N.C. Geurs, F.J. Hughes, L. Jin, A. Kantarci, E. Lalla, P.N. Madianos, D. Matthews, M.K. McGuire, M.P. Mills, P.M. Preshaw, M.A. Reynolds, A. Sculean, C. Susin, N.X. West, K. Yamazaki, Periodontal manifestations of systemic diseases and developmental and acquired conditions: Consensus report of workgroup 3 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions, J Periodontol 89 Suppl 1 (2018) S237-s248.

[8] L. Levin, Y. Zadik, T. Becker, Oral and dental complications of intra-oral piercing, Dental Traumatology 21(6) (2005) 341-343.

[9] J.M. Albandar, C.F. Streckfus, M.R. Adesanya, D.M. Winn, Cigar, pipe, and cigarette smoking as risk factors for periodontal disease and tooth loss, Journal of periodontology 71(12) (2000) 1874-1881.

[10] H. Wolf, T. Hassell, Color Atlas of Dental Hygiene: Periodontology: Periodontology, Thieme2006.

[11] N. West, M. Davies, A. Sculean, S. Jepsen, R. Faria-Almeida, M. Harding, F. Graziani, R.G. Newcombe, J.E. Creeth, D. Herrera, Prevalence of dentine hypersensitivity, erosive tooth wear, gingival recession and periodontal health in seven European countries, Journal of Dentistry 150 (2024) 105364.

[12] V.S. Yadav, B. Gumber, K. Makker, V. Gupta, N. Tewari, P. Khanduja, R. Yadav, Global prevalence of gingival recession: A systematic review and meta-analysis, Oral Dis 29(8) (2023) 2993-3002.

[13] P.A. Heasman, R. Holliday, A. Bryant, P.M. Preshaw, Evidence for the occurrence of gingival recession and non-carious cervical lesions as a consequence of traumatic toothbrushing, J Clin Periodontol 42 Suppl 16 (2015) S237-55.

[14] P.A. Heasman, M. Ritchie, A. Asuni, E. Gavillet, J.L. Simonsen, B. Nyvad, Gingival recession and root caries in the ageing population: a critical evaluation of treatments, J Clin Periodontol 44 Suppl 18 (2017) S178-S193.

[15] P.M. Preshaw, Detection and diagnosis of periodontal conditions amenable to prevention, BMC Oral Health 15 Suppl 1(Suppl 1) (2015) S5.

[16] M. Kuralt, R. Gašperšič, A. Fidler, The precision of gingival recession measurements is increased by an automated curvature analysis method, BMC Oral Health 21(1) (2021) 505.

[17] D. Schneider, A. Ender, T. Truninger, C. Leutert, P. Sahrmann, M. Roos, P. Schmidlin, Comparison between clinical and digital soft tissue measurements, J Esthet Restor Dent 26(3) (2014) 191-9.
[18] A.F. Hefti, Periodontal probing, Crit Rev Oral Biol Med 8(3) (1997) 336-56.
[19] S.G. Grossi, R.G. Dunford, A. Ho, G. Koch, E.E. Machtei, R.J. Genco, Sources of error for periodontal probing measurements, J Periodontal Res 31(5) (1996) 330-6.
[20] T.L. Watts, Probing site configuration in patients with untreated periodontitis. A study of horizontal positional error, J Clin Periodontol 16(8) (1989) 529-33.
[21] G. Pini-Prato, D. Franceschi, F. Cairo, M. Nieri, R. Rotundo, Classification of dental surface defects in areas of gingival recession, J Periodontol 81(6) (2010) 885-90.
[22] M.P. Santamaria, F.F. Suaid, M.Z. Casati, F.H. Nociti, A.W. Sallum, E.A. Sallum, Coronally positioned flap plus resin-modified glass ionomer restoration for the treatment of gingival recession associated with non-carious cervical lesions: a randomized controlled clinical trial, Journal of periodontology 79(4) (2008) 621-628.
[23] S. Bittencourt, E. Del Peloso Ribeiro, E.A. Sallum, A.W. Sallum, F.H. Nociti, M.Z. Casati, Comparative 6-month clinical study of a semilunar coronally positioned flap and subepithelial connective tissue graft for the treatment of gingival recession, Journal of periodontology 77(2) (2006) 174-181.
[24] S. Kerner, D. Etienne, J. Malet, F. Mora, V. Monnet-Corti, P. Bouchard, Root coverage assessment: validity and reproducibility of an image analysis system, J Clin Periodontol 34(11) (2007) 969-76.
[25] S. Kerner, A. Borghetti, S. Katsahian, D. Etienne, J. Malet, F. Mora, V. Monnet-Corti, J.M. Glise, P. Bouchard, A retrospective study of root coverage procedures using an image analysis system, J Clin Periodontol 35(4) (2008) 346-55.
[26] S. Kerner, J. Garabetyan, S. Le Roch, D. Damman, P. Bouchard, Image distortion of intra-oral photographs: The root coverage model, J Clin Periodontol 47(7) (2020) 875-882.
[27] O. Zuhr, S.F. Rebele, D. Schneider, R.E. Jung, M.B. Hurzeler, Tunnel technique with connective tissue graft versus coronally advanced flap with enamel matrix derivative for root coverage: a RCT using 3D digital measuring methods. Part I. Clinical and patient-centred outcomes, J Clin Periodontol 41(6) (2014) 582-92.
[28] H.N. Fageeh, A.A. Meshni, H.A. Jamal, R.S. Preethanath, E. Halboub, The accuracy and reliability of digital measurements of gingival recession versus conventional methods, BMC Oral Health 19(1) (2019) 154.
[29] N. Gkantidis, K. Dritsas, M. Ghamri, D. Halazonetis, A. Sculean, Methods for 3D evaluation and quantification of gingival recessions and gingival margin changes: Advancements from conventional techniques, Periodontol 2000 (2024).
[30] M. Kuralt, R. Gašperšič, A. Fidler, Methods and parameters for digital evaluation of gingival recession: A critical review, J Dent 118 (2022) 103793.
[31] J.A. Knottnerus, J.W. Muris, Assessment of the accuracy of diagnostic tests: the cross-sectional study, J Clin Epidemiol 56(11) (2003) 1118-28.
[32] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L. Irwig, J.G. Lijmer, D. Moher, D. Rennie, H.C. de Vet, H.Y. Kressel, N. Rifai, R.M. Golub, D.G. Altman, L. Hooft, D.A. Korevaar, J.F. Cohen, S. Group, STARD

2015: an updated list of essential items for reporting diagnostic accuracy studies, BMJ 351 (2015) h5527.
[33] W.A. Scott, Reliability of content analysis: The case of nominal scale coding, Public opinion quarterly (1955) 321-325.
[34] A. Donner, M. Eliasziw, A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation, Stat Med 11(11) (1992) 1511-9.
[35] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 1(8476) (1986) 307-10.
[36] N.M. Buderer, Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity, Acad Emerg Med 3(9) (1996) 895-900.
[37] M. Kuralt, A. Cmok Kučič, R. Gašperšič, A. Fidler, Evaluation of gingival recessions with conventional versus digital methods, Journal of Dentistry 120 (2022) 104093.
[38] M. Power, G. Fell, M. Wright, Principles for high-quality, high-value testing, Evid Based Med 18(1) (2013) 5-10.
[39] J.G. Caton, G. Armitage, T. Berglundh, I.L.C. Chapple, S. Jepsen, K.S. Kornman, B.L. Mealey, P.N. Papapanou, M. Sanz, M.S. Tonetti, A new classification scheme for periodontal and peri-implant diseases and conditions - Introduction and key changes from the 1999 classification, J Clin Periodontol 45 Suppl 20 (2018) S1-s8.
[40] O. Zuhr, S.F. Rebele, D. Schneider, R.E. Jung, M.B. Hürzeler, Tunnel technique with connective tissue graft versus coronally advanced flap with enamel matrix derivative for root coverage: a RCT using 3D digital measuring methods. Part I. Clinical and patient-centred outcomes, J Clin Periodontol 41(6) (2014) 582-92.
[41] G. Serino, J.L. Wennström, J. Lindhe, L. Eneroth, The prevalence and distribution of gingival recession in subjects with a high standard of oral hygiene, Journal of Clinical Periodontology 21(1) (1994) 57-63.
[42] J. Seong, D. Bartlett, R.G. Newcombe, N.C.A. Claydon, N. Hellin, N.X. West, Prevalence of gingival recession and study of associated related factors in young UK adults, Journal of Dentistry 76 (2018) 58-67.
[43] M. Addy, P. Mostafa, R.G. Newcombe, Dentine hypersensitivity: the distribution of recession, sensitivity and plaque, Journal of Dentistry 15(6) (1987) 242-248.
[44] A.S. Jati, L.Z. Furquim, A. Consolaro, Gingival recession: its causes and types, and the importance of orthodontic treatment, Dental Press J Orthod 21(3) (2016) 18-29.

Declaration of interests

☐The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Nicola West reports financial support was provided by University of Bristol. Nicola West reports a relationship with University of Bristol that includes: funding grants. Marko Kuralt reports a relationship with 3Shape that includes: employment. Marko Kuralt has patent #WO2025067852A1 pending to Automatic assesment of soft tissue recession in an oral cavity. Marko Kuralt has patent #WO2025093441A1 pending to Method and system for aligning digital 3d models of teeth. The university of Bristol received a grant from 3Shape If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.