

Inter-method agreement between digital occlusograms and mounted-cast articulator analysis in temporomandibular disorders with degenerative joint changes

Gizem Altundal^{ID} · Ece Zorlu^{ID} · Sahra Yıldırım^{ID} · Fatma Yüce^{ID} · İpek Şavkan^{ID} · Sibel Biren^{ID}

Received: 4 June 2026 / Accepted: 29 September 2026
© The Author(s) 2026

Abstract

Background Intraoral scanner occlusograms are increasingly used for non-invasive digital occlusal analysis, yet their agreement with an articulator-mounted reference in temporomandibular disorder (TMD) patients with degenerative joint changes is unknown. This is a methodological question about the fidelity of digital occlusal recording in a population with altered occlusal anatomy, not a claim that occlusal factors cause TMD or that occlusal adjustment treats it. This study aimed to quantify per-cusp inter-method agreement between an iTero Element 5D Occlusogram and a semi-adjustable articulator with 8 µm articulating paper on identical mounted casts, and to compare agreement between TMD patients with cone-beam computed tomography (CBCT)-confirmed degenerative joint changes and Class I controls.

Methods Sixty adults (30 TMD with DC/TMD Axis I and CBCT-confirmed degenerative joint changes; 30 Class I controls selected to have comparable age and sex distributions) had Type IV stone casts mounted on a SAM 3 articulator in centric relation, recorded with the Dawson bimanual manipulation technique; after removal of the interocclusal record the articulator was closed until the mounted casts came into occlusal contact, and the incisal pin was locked to preserve that standardised position; both methods were then read with the casts held in it. Two operators, blinded to study-group allocation during their respective occlusal assessments, classified 12,718 cusp/surface sites on identical casts using 8 µm Arti-Fol (reference) and the iTero Element 5D Occlusogram (index). The Occlusogram threshold was selected by maximising pooled Cohen's κ, Matthews correlation coefficient (MCC), and F1 score. Group comparisons were adjusted for within-patient clustering using a patient-level generalised estimating equation (GEE) and per-patient nonparametric tests.

Results At the optimised ≤ 0.2 mm threshold, inter-method agreement fell from substantial in controls to fair in TMD patients (Cohen's κ 0.636 vs. 0.392; $p = 2.5 \times 10^{-6}$). Sensitivity decreased by 20.1% points in TMD (93.16% vs. 73.08%) and positive predictive value by 21.2 points (49.32% vs. 28.15%; both $p < 0.001$); specificity and negative predictive value remained high ($\geq 96.9\%$ and $\geq 99.5\%$, respectively) and overall accuracy was 96.51% in TMD and 98.11% in controls. Both reductions persisted after adjustment for within-patient clustering (GEE odds

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



ratio 0.17 [95% CI 0.05–0.62]; $p=0.007$ for sensitivity and 0.51 [0.30–0.89]; $p=0.017$ for positive predictive value). False positives clustered predominantly at mandibular incisal edges.

Conclusions Even under controlled mounted-cast conditions, agreement between the iTero Occlusogram and the articulator-mounted reference dropped from substantial in controls to fair in TMD patients with degenerative joint changes, driven by reduced sensitivity relative to the articulating-paper comparator. Digital occlusograms should therefore be interpreted with caution when accurate static contact information is required in this population, and confirmation with a physical occlusal indicator on mounted casts may be advisable before irreversible restorative procedures. These findings concern the fidelity of digital occlusal recording and do not bear on the aetiology or management of TMD itself.

Keywords Temporomandibular Joint Disorders · Dental Occlusion · Dental Articulators · Cone-Beam Computed Tomography · Sensitivity and Specificity · Reproducibility of Results

Introduction

Temporomandibular disorders (TMDs) are the most common non-odontogenic source of orofacial pain, with global prevalence around 31–43% in adults and a consistent female predominance [1, 2], with female-to-male ratios reported between 1.5:1 and 6:1 in clinical and population-based studies [3, 4]. Contemporary understanding holds that occlusal factors are not the primary aetiological determinants of TMD, and systematic reviews have concluded that occlusion alone cannot account for TMD pathogenesis [5, 6]; irreversible occlusal adjustment is accordingly not regarded as an evidence-based treatment for TMD or degenerative joint disease [5, 6]. The present study therefore makes no assumption of a causal occlusion–TMD link. Accurate static occlusal recording nonetheless remains clinically necessary in these patients for reasons unrelated to TMD aetiology—restorative and prosthodontic rehabilitation, implant occlusion, occlusal-splint fit, and orthodontic finishing—and TMD with degenerative joint changes is of interest here specifically as a challenging anatomical condition in which the fidelity of digital occlusal capture has not been established [7, 8].

Articulating paper on mounted casts has constituted the conventional reference for static occlusal analysis for several decades [9, 10]. No occlusal-analysis method meets a strict gold-standard definition [11, 12], but 8 μ m articulating paper on accurately mounted casts is widely accepted as a clinical reference because of its near-natural occlusal tactile threshold (8–10 μ m) [9, 13, 14]. Articulating-paper marks nevertheless remain operator-dependent and subject to false-positive smearing [11, 13–15], and mechanical articulators cannot fully replicate jaw flexure or muscle-guided closure trajectories [16, 17].

Intraoral scanners (IOS) have introduced a non-invasive, three-dimensional alternative for occlusal analysis [18, 19]. The iTero Element 5D Occlusogram (Align Technology, San Jose, CA, USA) partitions interocclusal distance into seven 0.2 mm-wide color-coded bins, from < 0 mm (red, intersecting) to 1.0–1.2 mm (blue) [20]. In vitro validation has shown that virtual occlusal contacts can reproduce articulator contacts with sensitivities of 0.76–0.89 when an appropriate distance threshold is selected [21–23]. Clinical evidence remains heterogeneous: Esposito et al. [19] reported that IOS and 8 μ m articulating paper did not match for any tooth except the upper central incisors and first premolars (Fleiss κ 0.06–0.32); Fraile et al. [24] found moderate IOS-paper agreement (κ = 0.56); and Korkut et al. [20] showed that occlusal contact areas vary systematically with restoration type and clinician experience even in controlled restorative contexts.

Two interrelated gaps persist. First, no published clinical validation has explicitly compared an iTero-based occlusogram against an articulator-mounted reference in patients with TMD diagnosed by DC/TMD Axis I criteria [25, 26]. Static occlusal analysis is clinically consequential in several settings, including implant dentistry, full-mouth rehabilitation, and complex restorative treatment, and no claim of relative clinical priority is intended here; patients with TMD and degenerative joint changes are of specific methodological interest because condylar and articular remodelling may alter the maxillomandibular relationship from which static intercuspation is

recorded, so inter-method agreement established in joints without degenerative change cannot be assumed to transfer to this group. Second, intraoral acquisition introduces uncontrolled variables (head position, salivation, bite-reproduction variability) particularly pronounced in TMD patients [27]; an in-laboratory mounted-cast comparison provides a more controlled comparator design [22, 23, 28]. Pohl et al. [27] compared intraoral scans taken before and after a session of manual therapy in ten patients with TMD against untreated healthy controls followed over the same interval, and reported a mean Euclidean inter-arch displacement of 0.347 mm in the TMD group versus 0.113 mm in controls. Those values describe how much the occlusal relationship shifted over the study interval rather than a baseline difference between groups, but they indicate that the maxillomandibular relationship is less positionally stable in TMD, which is the property an occlusal record has to capture.

This study aimed to (i) determine the per-cusp/per-surface inter-method agreement between the iTero Element 5D Occlusogram and a SAM 3 articulator with 8 µm articulating paper (the clinical comparator) on identical mounted Type IV stone casts; (ii) test whether agreement and classification performance differ between TMD patients with CBCT-confirmed degenerative joint changes and asymptomatic Class I controls; and (iii) localise inter-method disagreement anatomically. We hypothesised that (a) inter-method agreement would be reduced in TMD, on the premise that condylar and articular changes may alter the maxillomandibular relationship in which the casts are mounted and read, and (b) the agreement gap would be larger at posterior than at anterior teeth.

Materials and methods

Study design, ethics, and operator standardisation

This single-centre, prospective, cross-sectional comparative-methods (inter-method agreement) study was conducted at the Department of Orthodontics, Faculty of Dentistry, Istanbul Kent University, Istanbul, Turkey, between December 2025 and April 2026. The study protocol was approved by the Health Sciences Research Ethics Committee of Istanbul Kent University (Protocol No. E-10420511-050-51906; meeting date 19 November 2025; decision number 2025-09) and conducted in accordance with the Declaration of Helsinki (2013 revision). Written informed consent was obtained from all participants prior to data collection. The study was not prospectively registered in a public trial registry because it is a single-centre, investigator-initiated methodological comparison of two occlusal-analysis modalities on identical mounted casts, with no clinical intervention and no patient-level outcome; the full study protocol is described in full in the Methods. Reporting follows the STARD 2015 guidelines [29], adapted for an inter-method agreement / comparative-methods design in which neither method constitutes an absolute reference, with the completed checklist and a participant flow diagram (Supplementary Fig. S2) provided as Supplementary materials.

To prevent carry-over bias between the two methods, all measurements were performed on identical mounted casts by two examiners working independently, as recommended by STARD item 10b [29]. Operator A, an experienced orthodontist (> 5 years' clinical experience in TMD-related orthodontic practice), performed the alginate impressions, SAM 3 articulator mounting, and the articulating-paper occlusal analysis (the reference comparator, Sect. "[Reference comparator: SAM 3 articulator with 8 µm articulating paper on Type IV stone mounted casts](#)"). Operator B, an orthodontist trained in digital workflow, performed the iTero Element 5D scanning of the same mounted casts and the Occlusogram color-map analysis (the index test, Sect. "[Index test: iTero Element 5D Occlusogram of the same mounted casts](#)"). Operator A was aware of the participants' clinical status during the initial examination and impression-taking. The casts were, however, coded before articulator mounting, and Operator A performed the mounting and the articulating-paper assessment without access to patient identities or study-group information. Operator B was also blinded to study-group allocation during the digital assessment. Both operators remained blinded to the results obtained by the other method until data lock. Each operator completed a method-specific calibration session on five training cases not included in the analytic cohort—Operator A on articulating-paper reading and Operator B on Occlusogram colour-bin reading—repeating each case after a

washout interval. Intra-operator (test–retest) Cohen’s κ exceeded 0.80 for both operators (Operator A: 0.84 [95% CI 0.78–0.90]; Operator B: 0.87 [95% CI 0.82–0.92]) before main-cohort acquisition was initiated.

Participants

A consecutive sample of 66 adults (35 TMD candidates; 31 control candidates) was screened. Throughout this report, “screened” refers to these 66 individuals assessed for eligibility and “enrolled” to the 60 who entered the analytic cohort. After application of inclusion/exclusion criteria and removal of records with missing or technically inadequate data, the final analytic cohort comprised 60 participants ($n=30$ TMD; $n=30$ controls).

The TMD group consisted of patients presenting with temporomandibular joint complaints who were diagnosed according to DC/TMD Axis I criteria during routine clinical evaluation [25, 26]. All TMD patients additionally had existing CBCT records obtained as part of their routine diagnostic work-up. CBCT scans were independently interpreted and reported by a single experienced maxillofacial radiologist (more than 10 years’ experience) using the imaging criteria for TMJ degenerative joint disease (DJD) established by Ahmad et al. [30] and incorporated into the DC/TMD imaging protocol [25], together with the broader morphological classification of Vasegh et al. [31]. In accordance with Ahmad et al. [30], the radiographic diagnosis of DJD was made when at least one of the following definite criteria was identified on at least two consecutive slices: surface erosion, osteophyte formation, subchondral cyst, or generalised subcortical sclerosis. Condylar flattening, classified as indeterminate for osteoarthritis within this framework [30], was documented when present but was not used as a stand-alone inclusion criterion. Patients were included in the TMD group only when the radiologist’s report confirmed at least one definite DJD finding (erosion, osteophyte formation, subchondral cyst, and/or generalised subcortical sclerosis). The control group consisted of asymptomatic skeletal and dental Class I individuals with no self-reported TMD symptoms and no clinical signs suggestive of TMD on screening examination. In line with the ALARA principle, no CBCT examinations were acquired specifically for research purposes; the control cohort was therefore assembled from patients who had attended other departments of the Faculty of Dentistry (for example, for implant planning, third-molar assessment, or endodontic evaluation) and whose CBCT scans had already been obtained for those unrelated dental indications. From this pool, individuals fulfilling all inclusion criteria were invited for clinical examination, and only those whose CBCT scans were reported by the same maxillofacial radiologist as showing no definite DJD findings according to the Ahmad et al. [30] criteria were enrolled as controls. Because subclinical temporomandibular joint alterations may occur even in otherwise asymptomatic individuals, the requirement of a radiologist-verified CBCT in the control group was used to reduce potential misclassification bias.

Common inclusion criteria were age 18–55 years, full permanent dentition including fully erupted third molars in occlusion, and Angle Class I dental relationship. Exclusion criteria for both groups were previous orthodontic treatment, prosthetic restorations involving more than two teeth, any missing teeth, anterior or posterior open bite, severe attrition (Smith and Knight tooth wear index ≥ 3 on any tooth), bruxism diagnosed by clinical history, neurological or rheumatologic comorbidities affecting jaw function, and pregnancy. The control group was selected to achieve an age and sex distribution comparable to that of the TMD group, in order to account for the established female predominance of clinical TMD populations [3, 4]. This approach was planned before control recruitment. Individual matching was not performed, and no formal frequency-matching procedure based on predefined age or sex categories was applied; eligible controls were selected with consideration of the overall age and sex distributions of the TMD group. Written informed consent was obtained from all participants prior to inclusion.

Reference comparator: SAM 3 articulator with 8 μ m articulating paper on Type IV stone mounted casts

All procedures in this section were performed by Operator A, blinded to the iTero Occlusogram results. Maxillary and mandibular full-arch alginate impressions (Tropicalgin; Zhermack, Italy) were poured within 30 min of acquisition in Type IV high-strength gypsum (ISO 6873; Fujirock EP, GC Europe, Belgium), chosen for its

dimensional stability, fracture resistance during repeated articulator closures, and established use in prior IOS-versus-articulator validation [22–24, 32, 33].

A SAM 3 face-bow registration (SAM Präzisionstechnik, Munich, Germany) was used to mount the maxillary cast on a SAM 3 semi-adjustable articulator at the participant's arbitrary hinge axis. The mandibular cast was mounted in centric relation rather than in maximum intercuspation. Before registration, participants rested to relax the masticatory muscles and the mandible was guided into centric relation using the Dawson bimanual manipulation technique [34] rather than the habitual intercuspation position. The record was taken in Almore Bite Registration Wax (American Orthodontics, Sheboygan, WI, USA), selected over conventional pink baseplate wax for its higher rigidity and thermal stability (baseplate-wax distortions reported between 25 and 110 μm) [35, 36]. The wax was folded to a thickness that left approximately 3 mm of interocclusal separation at the most posterior teeth, so that the registered position was free of tooth guidance. The mandibular cast was mounted against this record with the incisal guide pin set to the vertical dimension of the record and resting on the incisal guide table, so that the registered position was held while the mounting plaster set; the wafer was removed once the plaster had set. The articulator was programmed with mean condylar inclination (30°) and Bennett angle (15°).

Once the wafer had been removed, the casts were separated by the thickness of the record, and the incisal guide pin was shortened to permit closure of the articulator. The upper articulator member was then closed in a controlled manner by the operator along its arc of closure until the mounted casts reached their initial occlusal contact position. The incisal pin was subsequently adjusted to contact the incisal guide table and locked to preserve this position for all subsequent measurements, referred to hereafter as the standardised occlusal position established after removal of the interocclusal record. No further adjustment of the pin was made during the repeated closures. This position does not represent maximum intercuspation, which is achieved in vivo under muscular guidance after further mandibular movement beyond the position established in the articulator. Locking the pin fixes the occlusal vertical dimension at which every subsequent closure is performed, so that all three markings are made at an identical vertical dimension and the contact pattern cannot become progressively denser across repeated closures. The pin setting was not altered between closures, consistent with the standard remount protocol, in which pin–table separation indicates a change in occlusal vertical dimension [37], and with the framework of Lin et al. [38], who used pin-elevation increments to standardise occlusal vertical dimension changes. Rigid fixation of the incisal pin before contact marking has been used for the same purpose by Morsy and Hammad [39], who secured the pin with autopolymerising resin at each maxillomandibular relation before registering contacts with articulating paper. Both the articulating-paper assessment and the iTero Occlusogram analysis were subsequently performed with the casts maintained in this identical locked position, so that the two methods were evaluated under the same maxillomandibular relationship.

Static occlusal contacts were registered with 8 μm Arti-Fol articulating paper (Bausch GmbH, Cologne, Germany), held with reusable forceps and interposed simultaneously on the right and left sides. The 8 μm thickness was selected for its proximity to the natural occlusal tactile threshold of 8–10 μm and its established use as a clinical reference for static contact identification [9, 13, 14, 24]. The articulator was closed to this locked position three consecutive times by a single trained operator, with the pin setting unchanged throughout; each marked region was recorded as “contact” if the dye mark was reproducibly visible across all three closures, and as “no contact” otherwise. The 212 anatomic sites per patient (140 posterior cusp/marginal-ridge regions: 7 sites/tooth across 20 posterior teeth, namely mesiobuccal cusp, distobuccal cusp, mesiolingual cusp, distolingual cusp, central fossa, mesial marginal ridge, distal marginal ridge; and 72 anterior incisal-edge/cingulum/marginal-ridge regions: 6 sites/tooth across 12 anterior teeth, namely incisal edge, labial surface, mesial marginal ridge, distal marginal ridge, lingual fossa, cingulum) were defined a priori using the regional schema established by DeLong et al. [22, 23] and adapted by Solaberrieta et al. [40]. Both arches were scored independently, so a single occlusal contact between an opposing pair of surfaces was recorded at two sites—the maxillary region and its mandibular counterpart—and the number of positive sites is accordingly about twice the number of physical contact pairs. After the dye marks were photographically documented, the cast surfaces were carefully cleaned with a soft brush

and isopropyl alcohol to prevent visual contamination of the subsequent digital scan; the casts remained in their face-bow-determined articulator mounting throughout this cleaning step.

Index test: iTero Element 5D Occlusogram of the same mounted casts

All procedures in this section were performed by Operator B on the same Type IV stone casts that had been mounted on the SAM 3 articulator and used for the reference-comparator analysis (Sect. "Reference comparator: SAM 3 articulator with 8 μ m articulating paper on Type IV stone mounted casts"), with the dye marks photographically documented and removed prior to scanning. This in-laboratory mounted-cast scanning approach was deliberately preferred over direct intraoral scanning to allow a controlled comparison between the two methods on identical occlusal substrates, removing intraoral confounders such as head position, salivation, soft-tissue obstruction and bite-reproduction variability that are particularly difficult to standardise in TMD patients [27]. This study design has been used by several previous IOS-versus-articulator validation studies, including Solaberrieta et al. [40] (Vel-Mix mounted casts on a Panadent PSH articulator scanned for virtual occlusal record), DeLong et al. [22, 23] (mounted casts as the basis for 3D image-derived contacts), Lee et al. [28], who noted that the limited intraoral space during direct scanning reduces accuracy and argued for cast-based studies that can control these conditions and improve reproducibility, and Sigvardsson et al. [41] (occlusal contacts on stone casts mounted in an articulator versus IOS-derived measurements).

The maxillary and mandibular casts were scanned independently with the iTero Element 5D scanner (Align Technology, San Jose, CA, USA) operated in restorative mode for the highest available occlusal-surface resolution. Scans were acquired using iTero Element 5D software version 2.19.0, with builds 2.19.0.89 and 2.19.0.107 recorded during the study period. Occlusal analyses were performed using OrthoCAD version 5.9.1.50 (Align Technology, San Jose, CA, USA). Bench accuracy of this device under ADA/ANSI 132 testing has been reported as 0.03–0.24% average error for full-jaw long-distance accuracy [42]. The two arches were then aligned to one another through a sequence of four bilateral inter-arch buccal scans (left first molars, left canines, right canines, right first molars), with the casts physically held in the same standardised locked position by the SAM 3 articulator mounting throughout the inter-arch scan acquisition. Both methods were therefore read at one and the same maxillomandibular position; this cast-mediated fixation eliminates the intraoral bite-reproduction step and replaces it with the face-bow- and centric-relation-record-determined position already established during cast mounting (Sect. "Reference comparator: SAM 3 articulator with 8 μ m articulating paper on Type IV stone mounted casts"). Following acquisition, the integrated Occlusogram tool (OrthoCAD; Align Technology) generated a color map of interocclusal distance partitioned into seven bins of 0.2 mm width: red (< 0 mm, intersecting surfaces), orange (0–0.2 mm), yellow (0.2–0.4 mm), green (0.4–0.6 mm), cyan (0.6–0.8 mm), light blue (0.8–1.0 mm), and blue (1.0–1.2 mm) [20]. For each anatomic site defined in Sect. "Reference comparator: SAM 3 articulator with 8 μ m articulating paper on Type IV stone mounted casts", Operator B recorded, using a standardised inspection protocol, the Occlusogram colour representing the smallest interocclusal distance. Where more than one colour was present within the same site, the colour corresponding to the smallest distance was selected (for example, red rather than orange). The classification was therefore based on the minimum interocclusal distance within the site and not on the relative area occupied by each colour.

Threshold selection

Because the Occlusogram bin that best corresponds to a true “contact” has not been clinically established, three candidate cut-offs were evaluated: ≤ 0 mm (red bin only), ≤ 0.2 mm (red + orange bins), and ≤ 0.4 mm (red + orange + yellow bins). For each candidate, four pooled metrics were calculated across the entire cohort ($n = 60$; 12,718 cusp/surface sites): Cohen’s κ , the Matthews correlation coefficient (MCC), the F1 score, and Youden’s J. The threshold maximising the prevalence-sensitive metrics (κ , MCC, F1) was selected as the operating point for all subsequent analyses (Fig. 1; Table 1). This approach was preferred over Youden’s J alone because contact

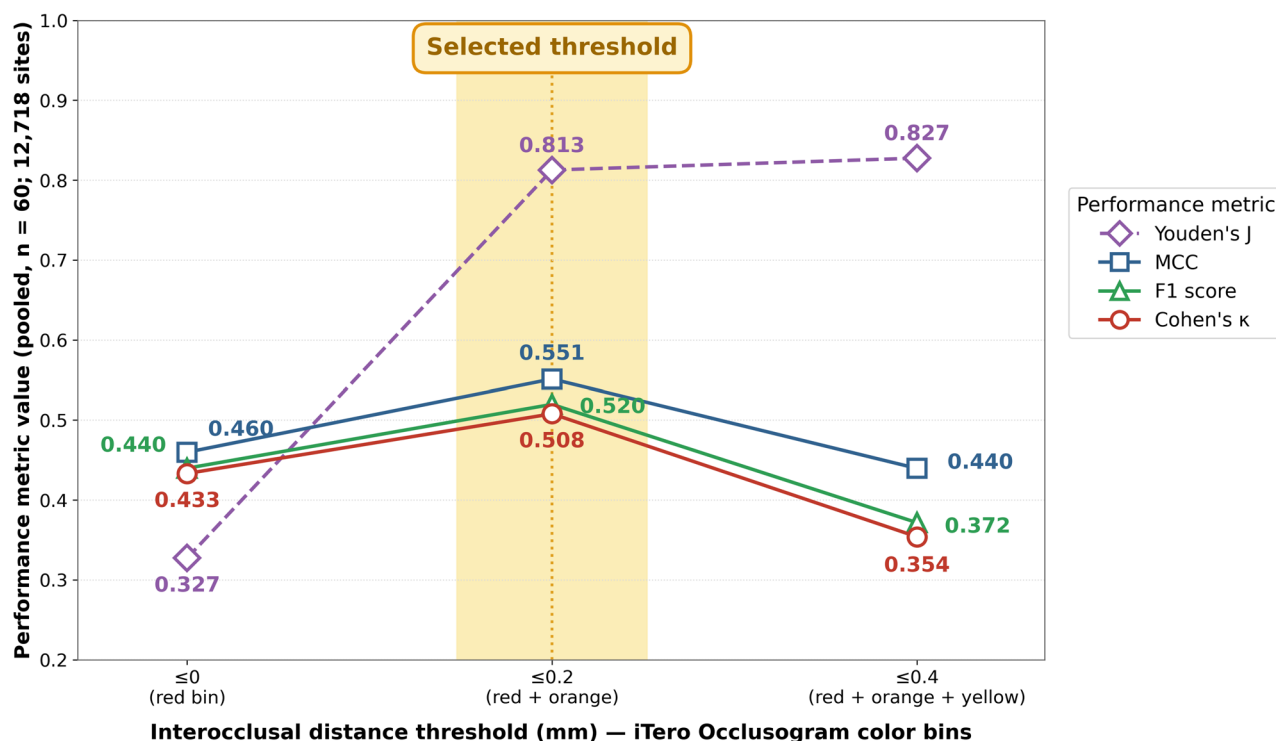


Fig. 1 Threshold selection across the iTero Occlusogram color scale. Pooled agreement metrics ($n=60$; 12,718 cusp/surface sites) plotted against three candidate interocclusal distance thresholds. The three prevalence-sensitive metrics (Cohen's κ , MCC, F1) peaked simultaneously at ≤ 0.2 mm

Table 1 Threshold-selection metrics across the three candidate iTero Occlusogram thresholds (pooled, $n=60$; 12,718 cusp/surface sites)

Metric	≤ 0 mm (red only)	≤ 0.2 mm (red + orange)	≤ 0.4 mm (+yellow)
True positives	73	185	194
False positives	38	306	629
False negatives	148	36	27
True negatives	12,459	12,191	11,868
Sensitivity (%)	33.03	83.71	87.78
Specificity (%)	99.70	97.55	94.97
PPV (%)	65.77	37.68	23.57
NPV (%)	98.83	99.71	99.77
Cohen's κ	0.433	0.508	0.354
MCC	0.460	0.551	0.440
F1 score	0.440	0.520	0.372
Youden's J	0.327	0.813	0.827

The ≤ 0.2 mm threshold (values shown in bold) simultaneously maximised the three prevalence-sensitive agreement metrics (Cohen's κ , MCC, F1) and was adopted as the operating point for all subsequent analyses

The ≤ 0 mm bin had κ , MCC and F1 comparable to ≤ 0.2 mm and a higher PPV (65.77% vs. 37.68%), but substantially lower sensitivity (33.03% vs. 83.71%); the ≤ 0.4 mm bin gave the highest Youden's J, but PPV fell to 23.57% (approximately three in four reported "contacts" would not correspond to articulator-confirmed contacts)

The SAM 3 / 8 μ m articulating-paper reference was held fixed across thresholds (221 reference-positive of 12,718 evaluable cusp/surface sites); thresholds differ only in the iTero contact-detection cut-off. Throughout, the two-by-two labels denote concordance or discordance with the articulating-paper comparator rather than absolute correctness or error (Sect. "Outcome metrics and unit of analysis")

TP True positives, FP False positives, FN False negatives, TN True negatives, PPV Positive predictive value, NPV Negative predictive value, MCC Matthews correlation coefficient

prevalence at the cusp level is low ($\approx 1.7\%$), and prevalence-insensitive metrics like Youden's J can favour over-inclusive thresholds at the expense of clinically meaningful positive predictive value [43]. Because the same cohort was used for both threshold selection and subsequent performance evaluation, this threshold-selection analysis is presented as exploratory and hypothesis-generating rather than confirmatory; external validation in an independent cohort, using the ≤ 0.2 mm cut-off as a pre-specified operating point, is warranted (also discussed in Sect. "Discussion", Limitations).

Outcome metrics and unit of analysis

For the selected threshold, the following metrics were calculated separately for the TMD and control groups, with the SAM-articulator/8- μ m-paper reference treated as the clinical reference comparator: (i) sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy, each with two-sided 95% Wilson confidence intervals; (ii) Cohen's κ with 95% confidence intervals obtained by Fisher's z transformation and Landis–Koch interpretation [44]; (iii) F1 score and Matthews correlation coefficient (MCC), the latter being the recommended single-number summary in imbalanced binary classification [45]; (iv) tooth-level and surface-level counts of the two directions of inter-method disagreement (sites classified as contact by the Occlusogram but not by the comparator, and sites classified as contact by the comparator but not by the Occlusogram) and their anatomic distribution. Throughout the manuscript the conventional two-by-two labels true positive, false positive, true negative and false negative are retained for brevity and for comparability with the diagnostic-accuracy literature, but they denote concordance or discordance with the articulating-paper comparator and not correctness or error in an absolute sense, since neither method constitutes a gold standard for the presence of an occlusal contact.

The unit of analysis was the anatomic site (cusp/surface), pre-defined a priori at 212 sites per patient (140 posterior cusp/marginal-ridge regions and 72 anterior incisal-edge/surface regions; Sect. "Reference comparator: SAM 3 articulator with 8 μ m articulating paper on Type IV stone mounted casts"). This per-cusp/per-surface unit was selected over per-tooth or per-patient summaries to capture clinically relevant intra-tooth heterogeneity in occlusal proximity (e.g., mesiobuccal vs. distobuccal cusps may have different contact status), consistent with the analytic schemas of DeLong et al. [22, 23], Solaberrieta et al. [40], Bostancıoğlu et al. [18] (per-cusp/per-area, 5 contact areas per tooth), and Sigvardsson et al. [41] (per-area). Because sites are nested within patients, the patient was treated as the unit of clustering for all group comparisons (Sect. "Statistical analysis"); per-patient κ , F1, and MCC distributions (Supplementary Table S1) together with a patient-level generalised estimating equation constituted the cluster-aware inference, and the per-site tabulations are presented as accompanying descriptive statistics.

Statistical analysis

Sample size was estimated for a κ -difference test between two independent samples (TMD vs. control) at a two-sided α of 0.05 and 80% power. Assuming $\kappa_{\text{TMD}} \approx 0.40$ and $\kappa_{\text{control}} \approx 0.65$ with an assumed contact prevalence of 3% per cusp and an intracluster correlation of 0.10, the minimum sample size was 28 per group; we enrolled 30 per group to allow for protocol deviations. Comparability of the sex distribution between TMD and control groups was tested with Pearson's χ^2 test, and age was compared between groups using an independent-samples t test. Because occlusal sites are nested within patients, the patient was the unit of clustering for all inferential comparisons. Cluster-adjusted inference for the two headline contrasts—sensitivity (probability that iTero detects a SAM-confirmed contact) and positive predictive value (probability that an iTero-reported contact is SAM-confirmed)—used generalised estimating equation (GEE) logistic regression with an exchangeable working correlation, the patient as the cluster, and cluster-robust (sandwich) standard errors. These two contrasts are the primary cluster-adjusted analyses. The sensitivity model was fitted on the comparator-positive sites and the positive predictive value model on the Occlusogram-positive sites; one TMD participant contributed

no Occlusogram-positive sites, so the latter was fitted on 491 sites from 59 participants. Two further GEE models with the same specification are reported as supportive analyses: overall site-level agreement between the two methods, and the false-positive rate, that is the probability that the Occlusogram classified a comparator-negative site as contact. Cluster-robust confidence intervals for the between-group difference in each metric were additionally obtained by a patient-level bootstrap (5,000 resamples of patients within each group). Per-patient κ , F1, MCC, sensitivity, and specificity distributions were compared with the Mann–Whitney U test (effect size: Cliff's δ). Unadjusted per-site two-proportion z-tests and the Donner–Eliasziw κ -comparison test [46] are reported as descriptive statistics accompanying the cluster-adjusted analyses. Anatomic-subgroup κ values were calculated for posterior (overall, maxilla, mandible, molar, premolar) and anterior (overall, maxilla, mandible, canine, incisor) regions and visualised on a heatmap with Landis–Koch zone shading. False-positive and false-negative event counts per tooth and surface were tabulated and visualised on a dental map of inter-method disagreement. All analyses were performed with Python 3.11 (NumPy, SciPy, scikit-learn, statsmodels, matplotlib) and double-checked in R 4.4 (irr, psych). All confidence intervals for proportions were Wilson score intervals and those for κ were obtained by Fisher's z transformation, reported uniformly as “95% CI lower–upper” in square brackets, with “to” replacing the dash when a confidence limit is negative. Exact p-values are reported for all comparisons. Statistical significance was set at $p < 0.05$. No adjustment for multiple comparisons was applied [47]. The primary cluster-adjusted analyses comprised the two GEE contrasts for sensitivity and positive predictive value; the remaining comparisons were considered supportive or exploratory and interpreted accordingly. As a sensitivity check for multiplicity, both primary contrasts retained statistical significance when evaluated against a Bonferroni-adjusted threshold of $\alpha = 0.025$; this check applies to those two contrasts only and does not constitute a multiplicity adjustment across the supportive and exploratory analyses.

Results

Participants

Thirty TMD patients (24 female, 6 male; mean age 32.4 ± 8.9 years) and 30 asymptomatic controls selected to have comparable age and sex distributions (23 female, 7 male; mean age 31.7 ± 9.4 years) were analysed. All TMD participants met DC/TMD Axis I clinical criteria and showed CBCT-documented degenerative temporomandibular joint changes, each having at least one definite radiographic finding (surface erosion, osteophyte formation, subchondral cyst and/or generalised subcortical sclerosis) according to the criteria of Ahmad et al. [30]. Condylar flattening was recorded where present but, being classified as indeterminate for osteoarthritis within that framework, was never sufficient on its own for classification as degenerative joint disease (Sect. “Participants”). No control participant showed definite degenerative findings on review of existing CBCT records. The female-to-male ratio in the TMD cohort (4:1) is consistent with established TMD epidemiology [3, 4]. Demographic characteristics were comparable between groups: sex distribution did not differ (Pearson $\chi^2 = 0.077$, $p = 0.78$), nor did age (independent-samples t test, $t = 0.30$, $p = 0.77$). Of the 66 individuals screened, 6 were excluded before the analytic cohort was defined and were therefore not enrolled: 3 because of incomplete iTero acquisitions, 2 because of damaged stone casts, and 1 because of intraoperator confirmation of an exclusion criterion (Class II tendency at re-examination). Of the 12,720 planned sites ($212 \text{ sites} \times 60 \text{ participants}$), one cusp/surface record per group failed to export from the OrthoCAD Occlusogram data table—a single-record write error in the software export step, affecting one site in one TMD and one control participant—leaving 6,359 evaluable sites per group and 12,718 in total. This export loss is the sole source of the 12,720→12,718 difference; the evaluable count is 12,718 throughout the text, tables, and figures.

Threshold selection

Pooled across the entire cohort, Cohen's κ peaked at the ≤ 0.2 mm threshold ($\kappa = 0.508$), as did MCC (0.551) and F1 (0.520) (Fig. 1; Table 1). Youden's J was marginally higher at ≤ 0.4 mm (0.827 vs. 0.813), but this gain was offset by a 14.1%-point loss in pooled PPV (37.68% $\rightarrow$ 23.57%) and a 0.154-point loss in pooled κ (0.508 $\rightarrow$ 0.354). The ≤ 0 mm (red-only) bin had substantially lower sensitivity than the ≤ 0.2 mm threshold (33.03% vs. 83.71%) despite a higher positive predictive value (65.77% vs. 37.68%), and κ , MCC and F1 were comparable between the two. Restricting positivity to the red bin therefore classified fewer sites as contacts but failed to register a considerably larger proportion of comparator-positive sites; at the very low cusp-level contact prevalence observed here, high specificity alone is uninformative for clinical decision-making. The ≤ 0.2 mm threshold (red + orange Occlusogram bins) was retained as the operating point for all subsequent analyses, in agreement with prior IOS-based digital occlusion work that defined contacts and "near contacts" within this distance range [22–24]. The cut-off is therefore consistent with the range already used in that literature and is not solely a product of optimisation within the present dataset. Contact force and functional loading were not measured here, and no functional interpretation of the ≤ 0.2 mm distance is intended.

Agreement and classification performance: TMD versus controls

At ≤ 0.2 mm, the SAM-articulator reference identified 104 contacts in the TMD group and 117 contacts in the control group (per-cusp prevalence 1.64% and 1.84%, respectively). The corresponding confusion matrices (Fig. 2A, B) showed that iTero correctly classified 6,061 of 6,255 SAM-negative sites in TMD (TN) but generated 194 false positives (FP), versus 6,130 TN and 112 FP in controls. False negatives (FN) were 28 in TMD and 8 in controls.

All eight diagnostic metrics differed in the predicted direction (Fig. 2C; Table 2, Section A). Sensitivity was substantially higher in controls (93.16% [95% CI 87.09–96.49]) than in TMD patients (73.08% [95% CI 63.84–80.67]), a difference of 20.09% points ($p = 5.4 \times 10^{-5}$). PPV showed a similar 21.17-pp gap (49.32% [95% CI 42.80–55.87] vs. 28.15% [95% CI 23.12–33.79]; $p = 1.5 \times 10^{-6}$). Cohen's κ moved from "substantial" agreement in controls (0.636 [95% CI 0.572–0.701]) to "fair" agreement in TMD (0.392 [95% CI 0.313–0.471]; $\Delta\kappa = -0.244$, $p = 2.5 \times 10^{-6}$); MCC and F1 followed the same pattern (controls 0.670 and 0.645; TMD 0.440 and 0.406). Specificity (96.90% TMD vs. 98.21% controls) and NPV ($\geq 99.5\%$ in both groups) remained high throughout because of the strong cusp-level prevalence imbalance. Overall accuracy, defined as the proportion of all sites classified concordantly by the two methods, was 96.51% [95% CI 96.03–96.93] in TMD and 98.11% [97.75–98.42] in controls ($\Delta = -1.60\%$ points; $p = 2.3 \times 10^{-8}$). The absolute difference is small because the great majority of sites are comparator-negative and are classified as such by both methods; overall accuracy is reported here for completeness and is of limited interpretive value at this prevalence, which is why the prevalence-sensitive metrics were used for the primary comparison. These differences persisted after accounting for within-patient clustering. In the patient-level GEE, the odds of iTero detecting a SAM-confirmed contact were markedly lower in TMD than in controls (sensitivity odds ratio 0.171 [95% CI 0.047–0.621]; $p = 0.007$). The cluster-adjusted positive predictive value was likewise lower in TMD than in controls (odds ratio 0.512 [95% CI 0.296–0.886]; $p = 0.017$); this model was fitted on the 491 Occlusogram-positive sites contributed by 59 participants, one TMD participant having no Occlusogram-positive site. Per-patient analyses agreed: κ , F1, and MCC each showed medium-effect-size reductions in TMD (Cliff's $\delta = -0.387$, -0.387 , and -0.404 ; Mann–Whitney $p = 0.010$, 0.010 , and 0.007), and per-patient sensitivity was lower in TMD ($p = 0.023$; Cliff's $\delta = -0.260$), whereas per-patient specificity did not differ ($p = 0.303$). A patient-level cluster bootstrap (5,000 resamples) gave concordant cluster-robust intervals for the headline differences (Δ sensitivity -20.1 pp [95% CI -34.6 to -5.8]; $\Delta\kappa -0.24$ [95% CI -0.40 to -0.08]; both $p < 0.01$), while the specificity difference remained non-significant [95% CI -3.0 to $+0.2$ pp]. The between-group difference was therefore most evident in the detection of comparator-confirmed contacts and in overall

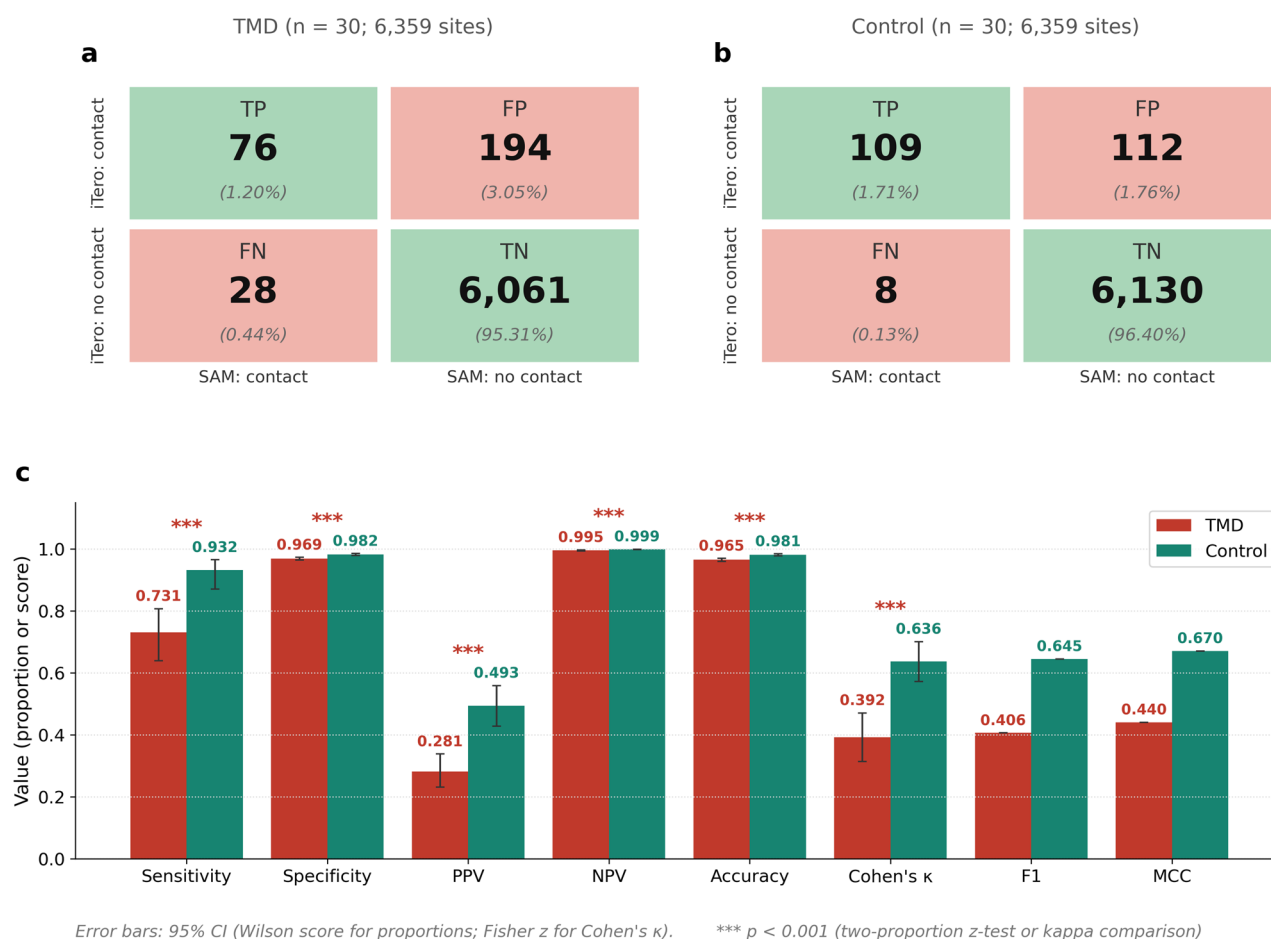


Fig. 2 Classification performance of the iTero Element 5D against the SAM 3 articulator reference comparator at the ≤ 0.2 mm threshold. **a**, **b** Confusion matrices for the TMD and control groups. **c** Classification-performance metrics by group. Error bars show 95% confidence intervals (Wilson score for the proportion-based metrics; Fisher z transformation for Cohen's κ); F1 and MCC are shown as point estimates; *** indicates $p < 0.001$ in the unadjusted per-site comparison (two-proportion z-test; Donner–Eliasiw test for Cohen's κ), which is descriptive only. Between-group inference accounting for within-patient clustering is reported in Table 2 (GEE $p = 0.007$ for sensitivity and $p = 0.017$ for positive predictive value)

agreement, whereas specificity remained similar between the groups in the cluster-aware analyses, and it is robust to the clustered structure of the data.

Anatomic subgroup analysis

The TMD-control κ gap varied by anatomic region (Table 2, Section B; Supplementary Fig. S1). These anatomic subgroup analyses are exploratory and descriptive: they were not part of the primary comparison (Sect. "Statistical analysis"), the per-subgroup site counts are comparatively small, and the values are presented to generate hypotheses rather than to support inference about individual regions. The largest gaps were observed at canines ($\kappa_{\text{TMD}} = 0.354$ vs. $\kappa_{\text{control}} = 0.642$; $\Delta = 0.288$), premolars (0.397 vs. 0.651; $\Delta = 0.254$), and at the mandibular anterior region (0.501 vs. 0.754; $\Delta = 0.253$). Molars showed the lowest agreement in both groups (TMD 0.265; control 0.322) and, correspondingly, the smallest TMD-control gap ($\Delta = 0.057$), which may indicate that posterior cusp anatomy presents a challenge for IOS-based contact identification largely independent of TMD status; this subgroup analysis was exploratory and does not establish such an effect. The maxillary anterior and incisor

Table 2 Per-cusp inter-method agreement between the iTero Occlusogram and the SAM 3 / 8 µm articulating-paper reference at the ≤ 0.2 mm threshold, by group (TMD with CBCT-confirmed degenerative joint changes vs. control)

Metric / Anatomic subgroup	TMD (<i>n</i> =30)	Control (<i>n</i> =30)	Δ (TMD–Control)	<i>p</i> -value
Section A. Overall inter-method agreement and classification performance (all 6,359 sites per group)				
True Positives	76	109	–33	—
False Positives	194	112	+ 82	—
False Negatives	28	8	+ 20	—
True Negatives	6,061	6,130	–69	—
Sensitivity (% , 95% CI)	73.08 (63.84–80.67)	93.16 (87.09–96.49)	–20.09	5.4 × 10^{–5}
Specificity (% , 95% CI)	96.90 (96.44–97.30)	98.21 (97.84–98.51)	–1.31	2.3 × 10 ^{–6}
PPV (% , 95% CI)	28.15 (23.12–33.79)	49.32 (42.80–55.87)	–21.17	1.5 × 10^{–6}
NPV (% , 95% CI)	99.54 (99.34–99.68)	99.87 (99.74–99.94)	–0.33	7.7 × 10 ^{–4}
Overall accuracy (% , 95% CI)	96.51 (96.03–96.93)	98.11 (97.75–98.42)	–1.60	2.3 × 10 ^{–8}
Cohen's κ (95% CI)	0.392 (0.313–0.471)	0.636 (0.572–0.701)	–0.244	2.5 × 10^{–6}
MCC	0.440	0.670	–0.230	0.007 ^a
F1 score	0.406	0.645	–0.239	0.010 ^a
Cluster-adjusted analysis (patient the unit of clustering; Methods 2.7)				
GEE sensitivity, OR (95% CI) ^b	0.171 (0.047–0.621)	Control (reference)	OR=0.171	0.007
GEE PPV, OR (95% CI) ^b	0.512 (0.296–0.886)	Control (reference)	OR=0.512	0.017
GEE overall agreement, OR (95% CI) ^b	0.532 (0.293–0.967)	Control (reference)	OR=0.532	0.038
GEE false-positive rate, OR (95% CI) ^b	1.751 (0.943–3.250)	Control (reference)	OR=1.751	0.076
Per-patient κ, median (MWU) ^c	0.541	0.757	δ = –0.387	0.010
Per-patient sensitivity (MWU) ^c	—	—	δ = –0.260	0.023
Per-patient F1 (MWU) ^c	0.552	0.764	δ = –0.387	0.010
Per-patient MCC (MWU) ^c	0.600	0.772	δ = –0.404	0.007
Per-patient specificity (MWU) ^c	—	—	δ = –0.156	0.303
Section B. Cohen's κ by anatomic subgroup (descriptive/exploratory; not part of the primary comparison)				
<i>Anatomic subgroup</i>	<i>TMD κ (95% CI)</i>	<i>Control κ (95% CI)</i>	<i>Δ</i>	<i>Landis–Koch</i>
Molars	0.265 (0.119–0.411)	0.322 (0.232–0.412)	–0.057	Fair
Premolars	0.397 (0.230–0.563)	0.651 (0.541–0.761)	–0.254	Fair → Substantial
Canines	0.354 (0.100–0.608)	0.642 (0.471–0.813)	–0.288	Fair → Substantial
Incisors	0.535 (0.415–0.656)	0.755 (0.671–0.839)	–0.220	Moderate → Substantial

Table 2 (continued)

Section B. Cohen's κ by anatomic subgroup (descriptive/exploratory; not part of the primary comparison)

Anatomic subgroup	TMD κ (95% CI)	Control κ (95% CI)	Δ	Landis–Koch
Maxillary anterior	0.484 (0.322–0.645)	0.699 (0.601–0.797)	–0.216	Moderate → Substantial
Mandibular anterior	0.501 (0.351–0.651)	0.754 (0.665–0.843)	–0.253	Moderate → Substantial

Key between-group metrics (sensitivity, PPV, Cohen's κ) are shown in bold

Section A reports the overall per-cusp inter-method agreement and classification performance (all 6,359 cusp/surface sites per group)

Section B reports Cohen's κ stratified by anatomic subgroup for hypothesis-generating purposes and was not part of the primary comparison

95% confidence intervals were computed as Wilson score intervals for proportions and via Fisher's z transformation for κ , reported uniformly as "value (95% CI, lower–upper)"; exact p-values are reported throughout. Landis–Koch interpretation: 0.00–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1.00 almost perfect. Subgroup tooth groups follow the FDI numbering system

Prevalence of reference-confirmed contacts: TMD 1.64% (104 sites); control 1.84% (117 sites)

The two-by-two labels (true/false positive and negative) denote concordance or discordance with the articulating-paper comparator, not correctness or error in an absolute sense, since neither method constitutes a gold standard (Sect. "Outcome metrics and unit of analysis")

Between-group inference accounting for within-patient clustering is provided by the GEE and per-patient analyses shown in the cluster-adjusted block; a patient-level cluster-bootstrap sensitivity analysis is additionally reported in the Results; the accompanying per-site two-proportion and κ -comparison p-values are unadjusted and reported for descriptive purposes only

PPV Positive predictive value, NPV Negative predictive value, MCC Matthews correlation coefficient, Δ Group difference, OR Odds ratio

^a MCC and F1 have no closed-form group test; the value shown is the patient-level Mann–Whitney comparison, ^b GEE, generalised estimating equation (exchangeable working correlation, patient cluster, cluster-robust standard errors; control=reference) The sensitivity model was fitted on the 221 comparator-positive sites from all 60 participants and the positive predictive value model on the 491 Occlusogram-positive sites from 59 participants; one TMD participant contributed no Occlusogram-positive site and therefore did not enter the latter model. Sensitivity and positive predictive value were the primary cluster-adjusted contrasts; the overall-agreement and false-positive-rate models were supportive analyses, ^c Per-patient distributions compared by the Mann–Whitney U test; effect size Cliff's δ ; per-patient κ was evaluable in all 30 TMD and 30 control participants

regions showed intermediate gaps ($\Delta=0.216$ and 0.220 , respectively) and remained in the moderate range in TMD patients (0.484 and 0.535), versus substantial agreement in controls (0.699 and 0.755).

Tooth- and surface-level distribution of inter-method disagreement

Pooled across both groups, 306 false-positive events and 36 false-negative events were observed at the tooth-surface level (Fig. 3). False positives clustered heavily at mandibular incisal edges, with the highest hotspots (≥ 6 false-positive events each) being tooth 41-IncisalEdge ($n=11$), 42-IncisalEdge ($n=10$), 32-IncisalEdge ($n=8$), 44-MesioBuccal ($n=8$), 31-IncisalEdge ($n=7$), 13-IncisalEdge ($n=7$), and four further sites tied at six events each (21-IncisalEdge, 35-DistoBuccal, 46-DistoBuccal, and 47-DistoBuccal). False negatives were rarer and more dispersed, with the most frequent occurrences at tooth 23-IncisalEdge, 27-MesioLingual, 21-IncisalEdge, and 31-LabialSurface ($n=2$ each). The clear concentration of false positives at mandibular anterior incisal edges suggests that overbite/overjet relationships in this region are systematically registered by the Occlusogram as contact at sites where the articulating-paper comparator records none, an interpretation consistent with the geometric ambiguity that arises when the scanner projects vertical proximity along the long axis of an incisor crown rather than perpendicular to the occluding surface. This systematic false-positive pattern at the mandibular incisal edges is illustrated for two representative cases in Fig. 4.

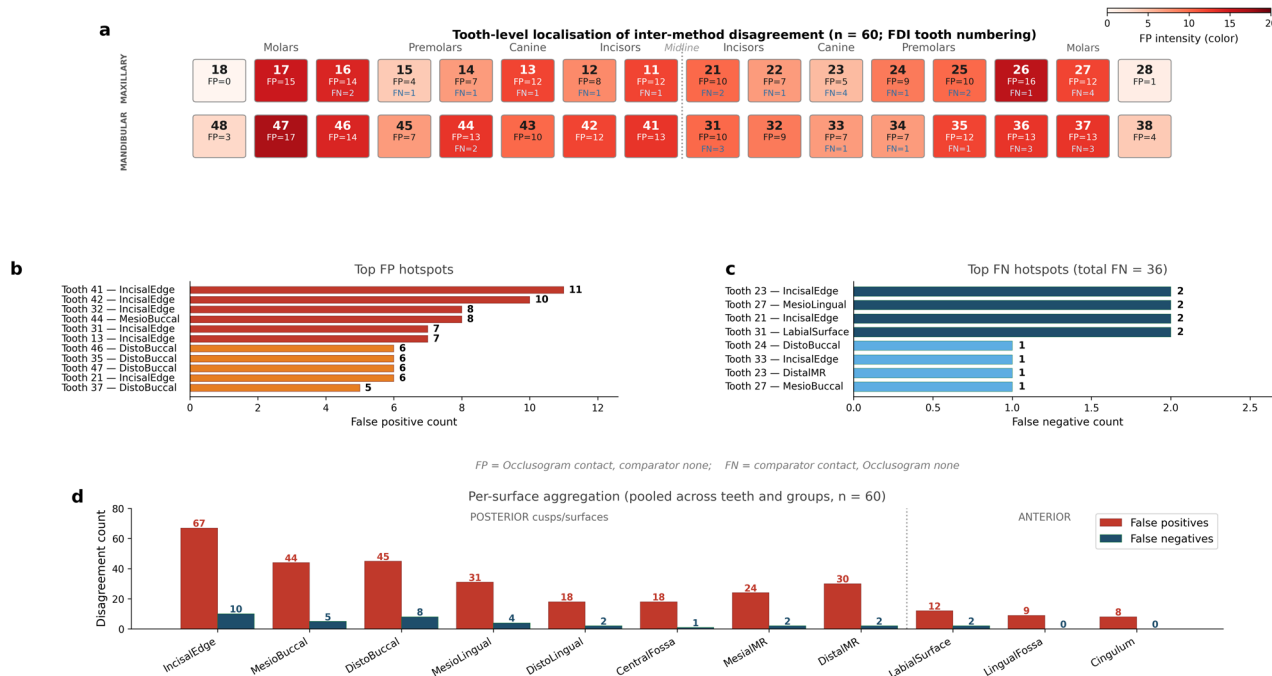


Fig. 3 Tooth- and surface-level localisation of inter-method disagreement between the iTero Occlusogram and the articulating-paper comparator at the ≤ 0.2 mm threshold (pooled across TMD and control groups; $n = 60$). **a** Tooth-level distribution of false positives (FP, red shading intensity) and false negatives (FN, blue) by FDI tooth numbering. **b** Top false-positive hotspots, concentrated at mandibular incisal edges. **c** Top false-negative hotspots, rarer and more dispersed (total FN=36). **d** Per-surface aggregation of inter-method disagreement pooled across teeth and groups. Panel labels within the figure use the conventional two-by-two terminology defined in Sect. "Outcome metrics and unit of analysis"; these denote discordance with the articulating-paper comparator rather than error in an absolute sense

Discussion

The present study found lower inter-method agreement between the digital Occlusogram and the articulator-mounted comparator in patients with TMD and degenerative joint changes than in asymptomatic Class I controls. Under a controlled mounted-cast design in which both methods were read from identical Type IV stone casts mounted in the same SAM 3 articulator, agreement at the optimised ≤ 0.2 mm threshold fell from the substantial range in Class I controls ($\kappa = 0.636$) to only fair in the TMD group ($\kappa = 0.392$). The deficit was concentrated in the detection of comparator-confirmed contacts—sensitivity and overall agreement rather than specificity—and was anatomically structured, with false positives clustering at the mandibular incisal edges and the largest group difference at the canines, premolars, and mandibular anterior region. Because both methods were read from the same casts, this gradient cannot be attributed to any live functional difference between the groups. Why it arises is not established by the present design: an interaction between degeneratively altered occlusal morphology and the scanner's proximity-based contact inference is one possible explanation, but no mechanism was tested in this study.

The SAM 3 articulator with 8 μ m articulating paper is used here as a clinical reference comparator rather than a true gold standard, following the consensus that no occlusal-analysis method has been proven ideal [11, 12, 19, 40]. Articulator condylar settings and instrumented recordings of the same condylar movements do not correspond exactly, with discrepancies of up to 3.4° (mean 1.2°) reported [16], and mechanical articulators cannot fully replicate jaw flexure [17]. Our aim was not to establish absolute occlusal truth (a goal that is challenging to attain in the absence of a true reference [22, 40]), but to quantify agreement between two clinically deployed systems against a well-established comparator.

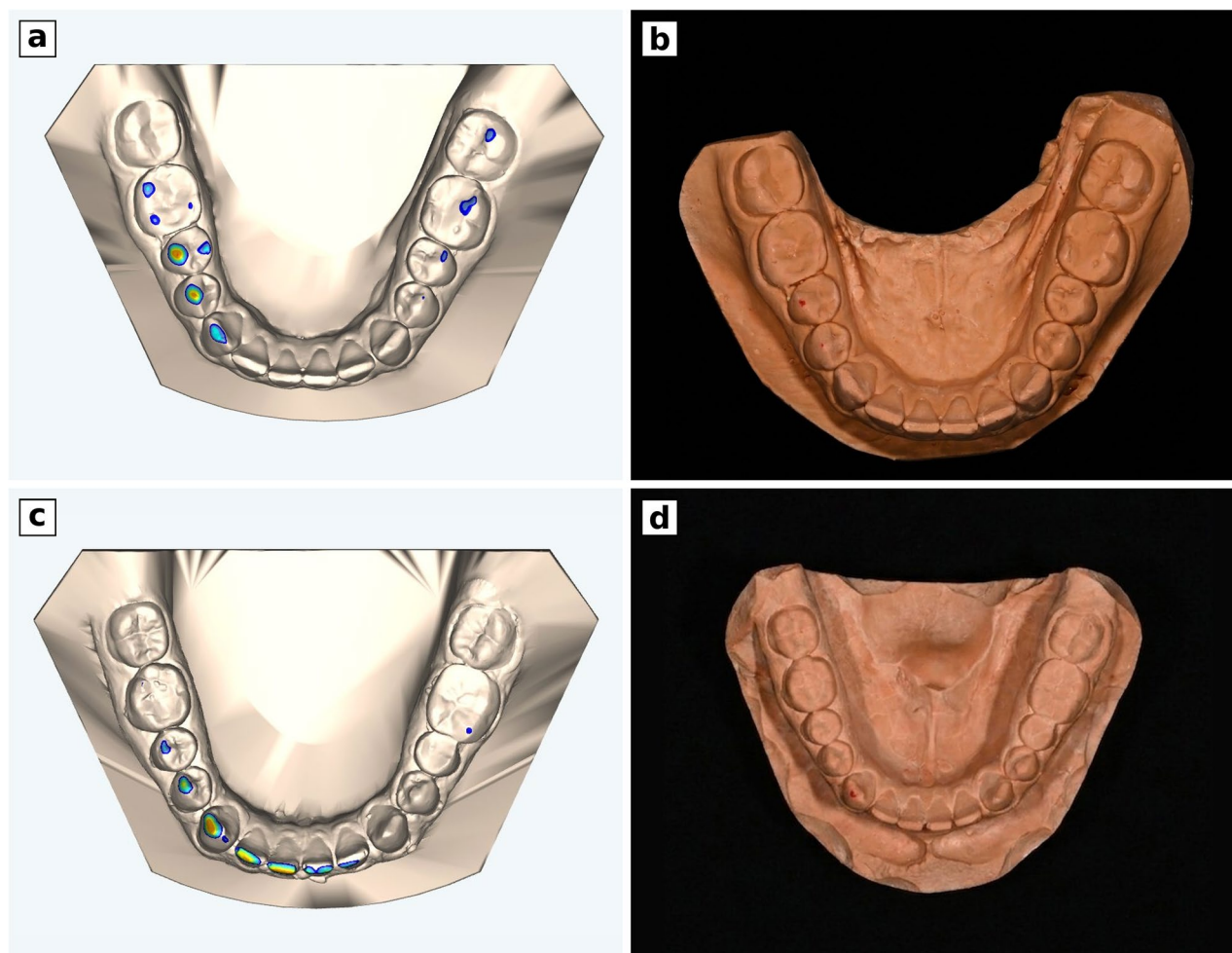


Fig. 4 Illustrative mandibular examples of inter-method concordance and discordance between the iTero Element 5D Occlusogram and the SAM 3 articulator with 8 μ m Arti-Fol at the ≤ 0.2 mm interocclusal distance threshold. **a, b** Control participant: the iTero Occlusogram (a) shows discrete colour contacts on the right mandibular posterior cusps (FDI 44–46 region), and the corresponding mounted Type IV stone cast (b) demonstrates concordant Arti-Fol marks at the same anatomic sites, exemplifying inter-method agreement at these sites. **c, d** TMD participant with CBCT-confirmed degenerative joint changes: the iTero Occlusogram (c) registers an extensive contact band across the mandibular incisal edges (FDI 32–42), whereas the corresponding mounted cast (d) shows no Arti-Fol marks at these sites, illustrating the systematic false-positive pattern at the mandibular incisal edges identified as the principal disagreement hotspot in Fig. 3. Third molars (FDI 18/28/38/48) were present and in occlusion in all participants and were included in the analysis (20 posterior teeth per participant); they lie outside the cropped occlusal field shown in these representative panels

The low absolute number of reference-confirmed contacts (104 of 6,359 sites in TMD and 117 of 6,359 in controls; per-cusp prevalence 1.64% and 1.84%, corresponding to approximately 3.5 and 3.9 positive sites per participant) is lower than the per-subject contact counts usually reported for dentate adults and therefore requires explicit comment. The principal reason is the occlusal position at which both methods were read. The reference comparator did not evaluate contacts after forcing the mounted casts into maximum intercuspation. The casts were mounted in centric relation, the interocclusal record was removed, the articulator was closed until the casts came into occlusal contact, and the incisal pin was locked to preserve that standardised position (Sect. "[Reference comparator: SAM 3 articulator with 8 \$\mu\$ m articulating paper on Type IV stone mounted casts](#)"); the reference therefore records the contacts present in that position rather than the contact distribution of maximum intercuspation, which the mandible reaches in vivo only after translating from that first contact under muscular guidance.

Whenever a discrepancy between centric relation and maximum intercuspation is present, initial tooth contact on the arc of closure involves a more limited contact distribution than maximum intercuspation; such discrepancies have been reported in TMD populations and may be more pronounced where degenerative condylar changes are present. Published per-subject contact counts, by contrast, are obtained in maximum intercuspation and are therefore not the appropriate comparator for these figures. A further consequence of the site schema compounds this: because both arches were scored independently, one contact between an opposing pair of surfaces yields two positive sites, so the 104 and 117 positive sites correspond to roughly 52 and 59 contact pairs, or fewer than two per participant. Three further features of the protocol act in the same direction. First, the positive unit was not a contact mark but one of 212 pre-defined anatomic landmarks—incisal edges, marginal ridges, fossae, and cuspal tips alike—most of which are not in contact in the standardised position at which the casts were read; a mark was scored positive only when it was centred on such a landmark, so dye deposited on a cuspal incline or between two adjacent landmarks contributed no positive site. This rule anchors both methods to an identical spatial frame, which is what an agreement analysis requires, but it necessarily yields fewer positive units than protocols that count every discernible mark. Second, the reference definition was deliberately specificity-weighted: an 8 µm foil—among the thinnest indicators in clinical use—was employed, and a site was scored as contact only when the dye mark was reproducibly visible across all three standardised closures at the locked vertical dimension, a criterion that necessarily under-records light or intermittent contacts. Third, static registration on mounted casts omits the occlusal loading and mandibular flexure present *in vivo*.

Reported numbers of occlusal contacts at maximum intercuspation in dentate adults vary widely between studies, and this variation tracks the indicator used and its physical properties at least as much as the population examined [48]; McDevitt and Warreth, for example, recorded 436 contacts across 38 normal dentitions [14]. The values reported here should accordingly be read as landmark-anchored, high-specificity positive-site counts recorded on mounted casts in the standardised position described above, not as an estimate of the number of functional occlusal contacts at maximum intercuspation, and they are not directly comparable with counts obtained in maximum intercuspation, with thicker indicators, or by intraoral registration. Because both methods were scored on the identical site grid, each under its own positivity rule applied uniformly across both groups, the between-group comparison is unaffected by this conservatism; the low prevalence does, however, inflate specificity and negative predictive value and depress κ and positive predictive value, which is why prevalence-sensitive metrics were triangulated and why the absolute counts are treated below as a limitation rather than as a finding.

Both methods were applied to identical articulator-mounted casts, eliminating intraoral confounders (head position, salivation, bite-reproduction variability) that are particularly difficult to standardise in TMD patients [27]. This design has precedent in prior IOS-articulator validation work [22, 28, 40, 41]. The trade-off is reduced *in-vivo* generalisability. The mounted-cast design removes intraoral sources of variance (saliva, patient movement, bite-reproduction variability) that are added, not subtracted, during direct intraoral scanning; it is therefore plausible that the inter-method difference would be at least as large *in vivo*. This is a hypothesis rather than a measured result: the present findings establish a controlled baseline only and should not be extrapolated to routine intraoral scanner performance in either direction without a direct *in-vivo* study.

A mounted-cast static analysis cannot capture the full repertoire of functional mandibular behaviour. Static contact analysis nevertheless remains the entry point of most digital occlusal workflows, and the centric relation closure position studied here is the one in which occlusal-splint fabrication, occlusal equilibration and the assessment of a centric relation to maximum intercuspation discrepancy are performed, so a digital system that performs poorly at this level is unlikely to be fully compensated by dynamic measurement. The present mounted-cast comparison thus provides a static baseline that subsequent dynamic and force-resolved work (T-Scan [13, 24], Modjaw, jaw-tracking) can extend.

The findings supported the primary hypothesis of lower inter-method agreement in the TMD group. The exploratory anatomic subgroup analysis did not support the anticipated larger between-group difference in posterior regions: although molars showed the lowest agreement in both groups, the between-group difference there was smaller than at the canines, premolars and mandibular anterior region. The fair-versus-substantial gradient

between TMD and control groups is directionally consistent with the positional instability reported by Pohl et al. [27], although the two designs are not equivalent: those authors measured how far the occlusal relationship shifted over a study interval (0.347 mm in TMD versus 0.113 mm in healthy controls) rather than agreement between two recording methods at one time point. Displacements of that order, expressed at the cusp level, are nevertheless of the same magnitude as the ≤ 0.2 mm decision threshold used here, which offers a plausible frame for why iTero misses 26.9% of reference-confirmed contacts in our TMD group while missing only 6.8% in controls. CBCT-confirmed structural condylar and articular changes in our cohort offer a possible structural substrate for this gap that pure pain-related cohorts cannot: condylar flattening, erosion, osteophyte formation and subcortical sclerosis may each subtly alter the maxillomandibular spatial relationship preserved by the SAM 3 articulator and read by the Occlusogram. The study was not designed to test this proposal. No association between individual radiographic features and the magnitude of inter-method disagreement was examined, and the suggestion is therefore a hypothesis for future work rather than a finding of the present analysis.

Because both methods were applied to the same articulator-mounted casts, the observed gap cannot be attributed to live neuromuscular behaviour during measurement. Any patient-level influence can enter only through the single upstream step shared by both methods—the centric relation bite registration used at mounting. Beyond that step, the differences are most parsimoniously explained at the level of cast geometry and image acquisition: the DJD-altered occlusal surfaces and maxillomandibular relationship captured at mounting, together with the projection geometry by which the Occlusogram infers proximity, rather than any dynamic instability that a static mounted cast cannot express. We therefore offer an interaction between altered occlusal morphology and the scanner's proximity-based contact inference as the most parsimonious hypothesis for the sensitivity gap, while noting that the present static design can neither confirm it nor exclude alternative explanations; we deliberately refrain from stronger neuromuscular or mechanistic claims that this design cannot support [5, 27].

Our absolute performance figures, expressed relative to the articulating-paper comparator, are broadly congruent with the IOS validation literature when the comparator and threshold are aligned. DeLong et al. [22, 23] reported sensitivities of 0.76–0.89 and specificities of 0.95–0.98 against shimstock and transillumination references. Solaberrieta et al. [40] reported a mean point-to-point deviation of 0.069 mm. Mitirattanakul et al. [49] reported high specificity, positive predictive value and accuracy for the iTero Element 2, but for the detection of tooth wear against a micro-CT reference: a different classification task, with a different reference and a different positive unit. Those values are therefore not directly comparable with occlusal-contact classification and are cited here only as evidence that the device resolves occlusal-surface detail reliably. Our control-group figures sit comfortably within the range reported by the contact-classification studies cited above. The TMD-group figures, by contrast, fall at or below the lowest published in vitro benchmarks, an observation that reframes prior in vitro validations as best-case estimates that may not transfer to clinically symptomatic populations even under controlled mounted-cast conditions.

The observed asymmetry of disagreement has several plausible mechanistic explanations. The disproportionate clustering of false positives at mandibular incisal edges (11/306 at tooth 41; 10/306 at tooth 42) is compatible with the geometric particularity of the lower-incisor anatomy in Class I relationships, in which overbite may place the lower incisal edge within 0.2 mm of the upper-incisor lingual surface even when no functional occlusal load is borne by the contact. The present data localise where the two methods diverge but do not establish why, so this geometric account is offered as a hypothesis rather than as a tested explanation. An additional source of inter-method disagreement may arise from the different criteria by which each method classified an anatomic site as positive. The minimum-distance rule allowed a small area of close proximity to render an entire site positive on the Occlusogram (Sect. "[Index test: iTero Element 5D Occlusogram of the same mounted casts](#)"), whereas the articulating-paper comparator required a reproducible mark centred on the predefined landmark across three closures. This difference may have contributed to some of the Occlusogram-positive, comparator-negative findings. It cannot, however, account on its own for the lower agreement observed in the TMD group, since the same colour-selection rule and the same articulating-paper criterion were applied in both groups; whether the asymmetry affected the two groups unequally was not assessed. The Occlusogram also measures interocclusal

proximity along the closing arc; it does not distinguish between a static interocclusal gap that would be bridged as the mandible advances into maximum intercuspation and a true loaded contact with measurable bite force. The 8 µm articulating paper, in contrast, only registers a mark when the closure force is sufficient to compress the dye layer, creating a systematic asymmetry consistent with the conceptual analysis of DeLong et al. [22] and Solaberrieta et al. [40], who noted that digital-impression-derived contacts represent a geometric proximity rather than a force-mediated contact, and with the comparative work of Bostancioğlu et al. [18].

The female predominance observed in our TMD cohort (24/30, 80%) and in the control group (23/30, 77%) is consistent with the established epidemiology of TMD, where clinical populations show female-to-male ratios of 4:1 to 6:1 and population-based studies document ratios of 1.5:1 to 20:1 [3, 4]. The comparable sex distribution between groups ($\chi^2 = 0.077$, $p = 0.78$) makes sex an unlikely confounder of the TMD-versus-control comparison. The cohort thus reflects the population most likely to seek occlusal/TMD assessment in routine practice and supports the external validity of our findings within this demographic, while limiting strict generalisation to male-predominant TMD subsets.

Three practical recommendations follow. They concern the safe use of digital occlusal recording and do not imply that occlusal adjustment is a treatment for TMD; contemporary evidence does not support irreversible occlusal therapy for TMD or degenerative joint disease [5, 6]. They are derived from mounted casts read in a single standardised centric relation closure position; direct intraoral scanning, maximum intercuspation and clinical outcomes were not evaluated in this study. They are therefore offered as provisional, scope-limited suggestions rather than as requirements for practice.

First, the iTero Element 5D Occlusogram appears better suited to screening and visualisation in TMD with degenerative joint changes than to serving as a stand-alone basis for irreversible occlusal intervention. With a TMD-group sensitivity of 73.08% and PPV of 28.15%, roughly one in four comparator-confirmed contacts were not displayed by the scanner in this cohort, and approximately seven of every ten sites the scanner displayed as contacts did not correspond to comparator-confirmed contacts. In practice these two figures act in opposite directions and carry different consequences. Contacts that the Occlusogram does not display are the more consequential for planning, because a contact that is not displayed is one the clinician does not know is present: an occlusal scheme designed on the digital map alone may leave a comparator-confirmed contact unaccounted for on a new restoration, on a splint intended to distribute contacts evenly, or during orthodontic finishing. The Occlusogram gives no indication of which sites are liable to be missed, so reading the map more attentively would not be expected to recover them; an independent physical check of the contacts would be required. The surplus of displayed contacts has the opposite character. It is conservative diagnostically but costly procedurally, since acting on it means adjusting surfaces that carry no comparator-confirmed contact, and such adjustment is irreversible. A workable compromise in this population is to use the Occlusogram to decide where to look, a task for which it is quick, non-invasive and reproducible, and an 8 µm foil on the mounted cast to verify a site before any irreversible adjustment is considered, reading the two side by side at the same articulator position as in the present protocol. Where a mounted cast is not available, the contacts displayed digitally are at minimum an approximation and should be checked with a thin foil before any irreversible step. Where selective grinding, restorative occlusal redesign or splint contact removal is contemplated in this population, the present data suggest that the Occlusogram alone provides a weak evidential basis for the decision.

Second, posterior contact information warrants particular caution: agreement at molars was the lowest of any subgroup in both populations ($\kappa = 0.265$ in TMD; 0.322 in controls). Because this was equally true of controls, it is better read as an intrinsic limitation of Occlusogram-based contact identification at posterior cusps than as a TMD-specific finding. Where posterior contact information derived from a digital occlusogram is to be acted on in TMD patients, we suggest that it be cross-checked by mounted-cast analysis with 8 µm articulating paper. Because the present data were acquired at the centric relation closure position, this recommendation applies directly to workflows performed in that position—splint fabrication, occlusal equilibration and orthodontic planning—and should be extended to intercuspal-position tasks such as prosthodontic or implant occlusion verification only with corresponding caution.

Third, mandibular incisal edges constitute a structured false-positive zone: 36 of 306 false-positive events ($\approx 12\%$) clustered at four mandibular incisal edges (FDI 41, 42, 32, 31). iTero “anterior contact” reports in this region therefore warrant cross-confirmation on a mounted cast before they are used to inform decisions about anterior intrusion, overbite reduction or anterior occlusal scheme redesign. Workflows combining static iTero analysis with dynamic measurement (T-Scan [13, 24], Modjaw, jaw-tracking) should treat the static iTero step as a screening layer in TMD patients with degenerative joint changes; confirmation on an articulator-mounted cast would be a reasonable precaution before a T-Scan-driven occlusal-equilibration protocol is undertaken.

Strengths include: DC/TMD Axis I clinical examination [25, 26]; per-cusp unit of analysis (12,718 sites) with cluster-adjusted inference (patient-level GEE and per-patient tests; Supplementary Table S1); two-operator stratified blinding [29]; controlled SAM 3 closure with the incisal pin locked to preserve a single standardised occlusal position for both methods [37–39]; Type IV gypsum [32] and a thermally stable bite-registration wax [35, 36]; the mounted-cast design [22, 28, 40, 41]; and threshold selection by triangulation of three prevalence-sensitive metrics.

Several limitations should be acknowledged. This was a single-centre study with a Class I–restricted cohort, which limits strict generalisation to other malocclusion classes, male-predominant TMD subsets, or TMD subtypes without degenerative changes (isolated myalgia, isolated disc displacement, pain-only TMD). The sample was also modest in absolute terms. Thirty participants per group satisfied the a priori power calculation for the planned κ contrast, but a sample of this size supports precise estimation only of a group difference of the magnitude anticipated; the anatomic-subgroup estimates in particular (Table 2, Section B) rest on small numbers of comparator-positive sites and have correspondingly wide confidence intervals. The findings should therefore be regarded as requiring confirmation in larger and more heterogeneous cohorts before they are generalised to routine practice. The composition of the control group is a further consideration. Because the ALARA principle precluded acquiring CBCT for research purposes alone, controls were assembled from individuals whose scans had already been obtained for unrelated dental indications such as implant planning, third-molar assessment or endodontic evaluation. Patients who undergo CBCT for these reasons are not a random sample of the asymptomatic population, and a degree of selection bias cannot be excluded; its direction with respect to inter-method agreement is not predictable from the present data. All controls were nonetheless required to be clinically asymptomatic, dentally and skeletally Class I, and free of definite radiographic degenerative findings on review by the same maxillofacial radiologist. The mounted-cast design enhances internal validity but limits in-vivo extrapolation, and functional and dynamic dimensions of TMD were not assessed; future work integrating jaw-tracking, T-Scan, or virtual-articulator instruments is warranted. No occlusal-analysis method has been proven ideal [11, 12]. The occlusal position studied also bounds what these findings cover. Because both methods were read in the standardised occlusal position established after removal of the interocclusal record, the results describe agreement for initial centric relation contacts—the position that is clinically relevant to splint fabrication, occlusal equilibration and orthodontic planning—and cannot be extrapolated to the full contact distribution of maximum intercuspation, which was not registered. The initial occlusal contact position was established through controlled closure of the articulator by a single operator. Locking the incisal pin preserved that position for all subsequent measurements, but identification of the initial contact position involved a degree of operator judgement, and the closure force applied at each marking was not instrumented, whereas comparable protocols have applied a defined static load [39]. Both factors may have influenced the recorded contact distribution and should be considered when interpreting the findings. Blinding was likewise not absolute: the casts were coded before mounting and neither operator had access to group allocation during assessment, but Operator A had examined the participants beforehand, so recognition of individual casts cannot be entirely excluded. The reference definition imposes a further limitation: requiring a mark reproducible across three standardised closures on articulator-mounted casts yields a conservative, high-specificity contact set, so the absolute number of positive sites under-represents the contacts present under function in vivo. These constraints bound generalisation of the absolute figures, although not the between-group comparison, in which each method’s positivity rule was applied uniformly to both groups. The same cohort was used both to select the ≤ 0.2 mm threshold and to evaluate performance at that threshold.

A cut-off chosen to maximise agreement within a given dataset will, on average, perform less well in a new one, so the agreement and classification figures reported here are likely to be optimistic to an unknown degree. The between-group contrast may be less directly affected than the absolute performance estimates, because the same threshold was applied to both groups; it is not, however, immune to bias introduced by data-driven threshold selection, and we cannot establish that any optimism operated to the same degree in each group. The absolute estimates should therefore be regarded as potentially optimistic, and the between-group findings also warrant confirmation in an independent cohort. The ≤ 0.2 mm cut-off is accordingly presented as exploratory and hypothesis-generating, and external validation in an independent cohort, with the threshold pre-specified rather than fitted, is required before it is adopted as a general operating point. Within-patient clustering was addressed directly rather than left as a residual concern: the primary group comparisons were adjusted for clustering using a patient-level GEE with cluster-robust standard errors and corroborated by per-patient nonparametric tests, so no inference treats sites within a patient as independent. Future work should pursue multicentre validation with the ≤ 0.2 mm cut-off pre-specified, evaluation using direct intraoral scanning to quantify the added in-vivo variance, extension to other malocclusion classes and to TMD subtypes without degenerative change (isolated myalgia, disc displacement, pain-only TMD), and integration of dynamic, force-resolved occlusal analysis (T-Scan, Mod-jaw, jaw-tracking) alongside the static digital occlusogram.

Conclusions

On identical Type IV stone casts mounted in centric relation and read in a single standardised occlusal position preserved by a locked incisal pin, an interocclusal distance threshold of ≤ 0.2 mm (red+orange iTero Occlusogram bins) maximised pooled agreement metrics against a SAM 3 articulator with 8 μ m articulating paper used as a clinical reference comparator. At this threshold, agreement was substantial in asymptomatic Class I controls ($\kappa=0.636$) but only fair in TMD patients with CBCT-confirmed degenerative joint changes ($\kappa=0.392$), with a 20.09-percentage-point drop in sensitivity and a 21.17-percentage-point drop in positive predictive value. Disagreement was anatomically structured, with false positives concentrated at mandibular incisal edges and the largest TMD-control gaps at canines, premolars, and the mandibular anterior region. Because this agreement gap was observed even under controlled mounted-cast conditions, the iTero Element 5D Occlusogram should be interpreted with caution when accurate static contact information is required in TMD patients with degenerative joint changes; where digital occlusal data will inform irreversible restorative, prosthodontic, or implant procedures, confirmation with a physical occlusal indicator on a mounted cast is advisable. These findings concern the fidelity of digital occlusal recording in a challenging anatomical population and do not bear on the aetiology or clinical management of TMD, for which occlusal adjustment is not an evidence-based therapy.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/s12903-026-10094-y>.

Acknowledgements The authors gratefully acknowledge the institutional support of the Department of Orthodontics, Faculty of Dentistry, Istanbul Kent University.

Authors' contributions G.A.: Conceptualization, Methodology, Investigation (Operator A — cast mounting and reference-standard articulating-paper analysis), Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing, Project administration. E.Z.: Investigation (per-cusp data acquisition and verification), Data curation, Writing – review & editing. S.Y.: Investigation (Operator B — iTero Element 5D scanning and index-test analysis), Writing – review & editing. F.Y.: Investigation (radiologic interpretation of CBCT scans according to the Ahmad et al. and DC/TMD imaging criteria), Validation, Writing – review & editing. İ.Ş.: Methodology, Writing – review & editing. S.B.: Conceptualization, Supervision, Resources, Writing – review & editing. All authors read and approved the final version of the manuscript.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability The datasets generated and/or analysed during the current study are not publicly available due to institutional data-protection policy but are available from the corresponding author on reasonable request and after appropriate ethical clearance.

Declarations

Ethics approval and consent to participate This study was performed in line with the principles of the Declaration of Helsinki (2013 revision). Approval was granted by the Health Sciences Research Ethics Committee of Istanbul Kent University (meeting date 19 November 2025; Protocol No. E-10420511-050-51906; decision number 2025-09). Written informed consent was obtained from all individual participants included in the study.

Consent for publication Not applicable. The figures contain dental cast and occlusogram images only and do not include any information by which an individual participant can be identified.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Alqutaibi AY, Alhammadi MS, Hamadallah HH, et al. Global prevalence of temporomandibular disorders: a systematic review and meta-analysis. *J Oral Facial Pain Headache*. 2025;39:48. <https://doi.org/10.22514/jofph.2025.025>.
2. Huang Z, Liu Y, Su B, et al. Temporomandibular disorder prevalence in malocclusion patients: a systematic review and meta-analysis. *Head Face Med*. 2025;21:13. <https://doi.org/10.1186/s13005-025-00490-0>.
3. Pascu L, Haiduc RS, Almășan O, Leucuța DC. Occlusion and temporomandibular disorders: a scoping review. *Med (Kaunas)*. 2025;61:791. <https://doi.org/10.3390/medicina61050791>.
4. Bueno CH, Pereira DD, Pattussi MP, Grossi PK, Grossi ML. Gender differences in temporomandibular disorders in adult populational studies: a systematic review and meta-analysis. *J Oral Rehabil*. 2018;45:720–9. <https://doi.org/10.1111/joor.12661>.
5. Manfredini D, Lombardo L, Siciliani G. Temporomandibular disorders and dental occlusion. A systematic review of association studies: end of an era? *J Oral Rehabil*. 2017;44:908–23. <https://doi.org/10.1111/joor.12531>.
6. Lekaviciute R, Kriauciunas A. Relationship between occlusal factors and temporomandibular disorders: a systematic literature review. *Cureus*. 2024;16:e54130. <https://doi.org/10.7759/cureus.54130>.
7. Mélou C, Leroux L, Bonnesoeur M, Le Padellec C, Bertaud V, Chauvel-Lebret D. Relationship between natural or iatrogenic malocclusions and temporomandibular disorders: a case control study. *Cranio*. 2024;42:206–14. <https://doi.org/10.1080/08869634.2021.1933307>.
8. Pullinger AG, Seligman DA. Quantification and validation of predictive values of occlusal variables in temporomandibular disorders using a multifactorial analysis. *J Prosthet Dent*. 2000;83:66–75. [https://doi.org/10.1016/s0022-3913\(00\)70090-4](https://doi.org/10.1016/s0022-3913(00)70090-4).
9. Davies S. Good occlusal practice in advanced restorative dentistry. A guide to good occlusal practice. Cham: Springer; 2022. pp. 79–118.
10. Kerstein RB, Radke J. Clinician accuracy when subjectively interpreting articulating paper markings. *Cranio*. 2014;32:13–23. <https://doi.org/10.1179/0886963413Z.0000000001>.
11. Carey JP, Craig M, Kerstein RB, Radke J. Determining a relationship between applied occlusal load and articulating paper mark area. *Open Dent J*. 2007;1:1–7. <https://doi.org/10.2174/1874210600701010001>.
12. Manziuc MM, Savu MM, Almășan O, et al. Insights into occlusal analysis: articulating paper versus digital devices. *J Clin Med*. 2024;13:4506. <https://doi.org/10.3390/jcm13154506>.

13. Qadeer S, Kerstein R, Kim RJY, Huh JB, Shin SW. Relationship between articulation paper mark size and percentage of force measured with computerized occlusal analysis. *J Adv Prosthodont*. 2012;4:7–12. <https://doi.org/10.4047/jap.2012.4.1.7>.
14. McDevitt WE, Warreth AA. Occlusal contacts in maximum intercuspation in normal dentitions. *J Oral Rehabil*. 1997;24:725–34. <https://doi.org/10.1111/j.1365-2842.1997.tb00268.x>.
15. Suese K. Progress in digital dentistry: the practical use of intraoral scanners. *Dent Mater J*. 2020;39:52–6. <https://doi.org/10.4012/dmj.2019-224>.
16. Celar AG, Tamaki K. Accuracy of recording horizontal condylar inclination and Bennett angle with the Cadiax compact. *J Oral Rehabil*. 2002;29:1076–81. <https://doi.org/10.1046/j.1365-2842.2002.00951.x>.
17. Solaberrieta E, Otegi JR, Mínguez R, Etxaniz O. Improving the digital workflow: direct transfer from patient to virtual articulator. *Int J Comput Dent*. 2013;16:285–92.
18. Bostancıoğlu SE, Toğay A, Tamam E. Comparison of two different digital occlusal analysis methods. *Clin Oral Investig*. 2022;26:2095–109. <https://doi.org/10.1007/s00784-021-04191-1>.
19. Esposito R, Masedu F, Cicciù M, Tepedino M, Denaro M, Ciavarella D. Reliability of recording occlusal contacts by using intraoral scanner and articulating paper. *J Dent*. 2024;142:104872. <https://doi.org/10.1016/j.jdent.2024.104872>.
20. Korkut B, Tüter Bayraktar E, Şenol AA, Uyar E, Yeler C, Yazıcı G. Digital comparison of occlusal vertical contacts between direct composite and indirect CAD/CAM restorations: an in vivo quantitative assessment. *Clin Exp Health Sci*. 2024;14:501–9. <https://doi.org/10.33808/clinexphealthsci.1400916>.
21. DeLong R, Heinzen M, Hodges JS, Ko CC, Douglas WH. Accuracy of a system for creating 3D computer models of dental arches. *J Dent Res*. 2003;82:438–42. <https://doi.org/10.1177/154405910308200607>.
22. DeLong R, Ko CC, Anderson GC, Hodges JS, Douglas WH. Comparing maximum intercuspal contacts of virtual dental patients and mounted dental casts. *J Prosthet Dent*. 2002;88:622–30. <https://doi.org/10.1067/mpd.2002.129379>.
23. DeLong R, Knorr S, Anderson GC, Hodges J, Pintado MR. Accuracy of contacts calculated from 3D images of occlusal surfaces. *J Dent*. 2007;35:528–34. <https://doi.org/10.1016/j.jdent.2007.02.004>.
24. Fraile C, Ferreiroa A, Romeo M, Alonso R, Pradies G. Clinical study comparing the accuracy of interocclusal records, digitally obtained by three different devices. *Clin Oral Investig*. 2022;26:1957–62. <https://doi.org/10.1007/s00784-021-04174-2>.
25. Schiffman E, Ohrbach R, Truelove E, et al. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for clinical and research applications. *J Oral Facial Pain Headache*. 2014;28:6–27. <https://doi.org/10.11607/jop.1151>.
26. Ohrbach R, Gonzalez Y, List T, Michelotti A, Schiffman E. (2014) Diagnostic criteria for temporomandibular disorders (DC/TMD) clinical examination protocol. www.rdc-tmdinternational.org
27. Pohl TDP, Harhoff AC, Ries J, Wichmann M, Matta RE. Changes of the occlusal relationship in patients with temporomandibular disorders (TMD) after manual therapy: a pilot study. *Clin Exp Dent Res*. 2025;11:e70147. <https://doi.org/10.1002/cre2.70147>.
28. Lee H, Cha J, Chun YS, Kim M. Comparison of the occlusal contact area of virtual models and actual models: a comparative in vitro study on Class I and Class II malocclusion models. *BMC Oral Health*. 2018;18:109. <https://doi.org/10.1186/s12903-018-0566-7>.
29. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies. *BMJ Open*. 2016;6:e012799. <https://doi.org/10.1136/bmjopen-2016-012799>.
30. Ahmad M, Hollender L, Anderson Q, Kartha K, Ohrbach R, Truelove EL, John MT, Schiffman EL. Research diagnostic criteria for temporomandibular disorders (RDC/TMD): development of image analysis criteria and examiner reliability for image analysis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2009;107:844–60. <https://doi.org/10.1016/j.tripleo.2009.02.023>.
31. Vasegh Z, Safi Y, Sanaei Azar M, Ghazizadeh Ahsaie M, Arianezhad SM. Assessment of bony changes in temporomandibular joint in patients using cone beam computed tomography – a cross sectional study. *Head Face Med*. 2023;19:47. <https://doi.org/10.1186/s13005-023-00392-z>.
32. Pudi S, Raj ARD, Balaji H, Rengasamy V, Matsa P, Anand VV. Dimensional accuracy in four types of type IV die stone after delayed pouring: a comparative study. *J Pharm Bioallied Sci*. 2024;16:S3164–6. https://doi.org/10.4103/jpbs.jpbs_654_24.
33. Aljohani MS, Bukhari HA, Alshehri M, Alamoudi A. Accuracy of the different materials used to fabricate a verification jig of implant-supported fixed complete dental prostheses: an in vitro study. *Cureus*. 2022;14:e29794. <https://doi.org/10.7759/cureus.29794>.
34. Dawson PE. Evaluation, diagnosis, and treatment of occlusal problems. 2nd ed. St. Louis: Mosby; 1989.
35. Dwivedi A, Maru K, Sharma A. A comparative evaluation of three dimensional accuracy of different types of interocclusal recording materials-an in vitro study. *Med Pharm Rep*. 2020;93:280. <https://doi.org/10.15386/MPR-1453>.
36. Rani P, Ranjan M, Sinha T, Singh A, Keshri M, Prakash J. Assessment of the capability to accurately capture the vertical relationship by the various bite registration materials. *J Pharm Bioallied Sci*. 2024;16:S745–7. https://doi.org/10.4103/jpbs.jpbs_990_23.

37. Zarb GA, Bolender CL, Eckert SE, et al. Prosthodontic treatment for edentulous patients: complete dentures and implant-supported prostheses. 13th ed. St. Louis: Mosby; 2013.
38. Lin YC, Scialabba R, Lee JD, Sun J, Lee SJ. Assessment of occlusal vertical dimension change in mechanical and virtual articulation: a pilot study. *Dent J (Basel)*. 2022;10:212. <https://doi.org/10.3390/dj10110212>.
39. Morsy N, Hammad I. Recording maximal intercuspation and border positions of the mandible with intraoral scanner using the acquisition software's multi-occlusion function. *J Adv Prosthodont*. 2024;16:221–30. <https://doi.org/10.4047/jap.2024.16.4.221>.
40. Solaberrieta E, Otegi JR, Goicoechea N, Brizuela A, Pradies G. Comparison of a conventional and virtual occlusal record. *J Prosthet Dent*. 2015;114:92–7. <https://doi.org/10.1016/j.prosdent.2015.01.009>.
41. Sigvardsson J, Nilsson S, Ransjö M, Westerlund A. Digital quantification of occlusal contacts: a methodological study. *Int J Environ Res Public Health*. 2021;18:5297. <https://doi.org/10.3390/ijerph18105297>.
42. Align Technology. iTero Element 5D imaging system: ADA/ANSI 132 bench testing data on file. San Jose: Align Technology; 2024.
43. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21:6. <https://doi.org/10.1186/s12864-019-6413-7>.
44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74. <https://doi.org/10.2307/2529310>.
45. Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy. *BioData Min*. 2021;14:13. <https://doi.org/10.1186/s13040-021-00244-z>.
46. Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic. *Stat Med*. 1992;11:1511–9. <https://doi.org/10.1002/sim.4780111109>.
47. Bender R, Lange S. Adjusting for multiple testing—when and how? *J Clin Epidemiol*. 2001;54:343–9. [https://doi.org/10.1016/s0895-4356\(00\)00314-0](https://doi.org/10.1016/s0895-4356(00)00314-0).
48. Qadeer S, Türp JC. Number of occlusal contacts during maximum intercuspation in dentate adults: a systematised review. *J Oral Rehabil*. 2023;50:617–26. <https://doi.org/10.1111/joor.13451>.
49. Mitirattanakul S, Neoh SP, Chalarmchaichaloenkit J, et al. Accuracy of the intraoral scanner for detection of tooth wear. *Int Dent J*. 2023;73:56–62. <https://doi.org/10.1016/j.identj.2022.06.004>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Gizem Altundal¹  · Ece Zorlu¹  · Sahra Yıldırım¹  · Fatma Yüce²  · İpek Şavkan¹  · Sibel Biren¹ 

✉ Gizem Altundal
turkmenngizem@gmail.com; gizem.altundal@kent.edu.tr

¹ Department of Orthodontics, Faculty of Dentistry, Istanbul Kent University, Istanbul, Turkey

² Department of Oral Diagnosis and Radiology, Faculty of Dentistry, Istanbul Kent University, Istanbul, Turkey