Investigating three methods of assessing the clinically relevant trueness of two intraoral scanners

› **C. OSNES[1], J. CAPLAN[2], M. FERRARI[3], A. KEELING[4]**

[1]Research Assistant, School of Dentistry, University of Leeds, Leeds, UK - PhD student in the Department of Medical Biotechnologies, University of Siena, Siena, Italy
[2]Private dental practitioner, Private practice, St Albans, UK
[3]Professor, Department of Medical Biotechnologies, University of Siena, Siena, Italy
[4]Associate Professor, School of Dentistry, University of Leeds, Leeds, UK

ABSTRACT

Aims Intraoral scanners (IOS) are used for a wide range of treatments. Most IOSs produce data appropriate for local work, such as crowns, but evidence suggests that full-arch scans result in more erroneous scans, which may affect the fit of clinical appliances. There are no standardized methods for assessing the quality of IOSs. Though many studies have investigated the accuracy of scanners, one may find the reported values are difficult to interpret in a clinical context.
Materials and methods This study investigated the trueness of two IOSs, using three metrics. The clinical value of each metric is discussed. A dentate model was scanned 10 times using two intraoral scanners. Three methods were used to assess the trueness of the scans against a scan produced in a laboratory scanner.
Results The mean unsigned distance deviation between a laboratory scan and the Primescan scans was 0.016(±0.006)mm. The mean unsigned distance deviation between the laboratory scan and the Omnicam scans was 0.116(±0.01)mm. The arch width between molars was 55.44mm for the Solutionix scan. The arch width of the Primescan was 55.439(±0.075)mm, while the Omnicam reported 54.672(±0.065)mm. The mean proportion of the Primescan scans deviating beyond 0.1mm when compared against the Solutionix was 0.7(±2.0)%. The equivalent for the Omnicam was 42.1(±2.5)%.
Conclusions All methods indicated significantly different results between the scanners. The Primescan produced truer scans than the Omnicam, regardless of measurement method. The intermolar-width and proportion beyond 0.1mm methods may give more clinically relevant insight into the trueness of scan data than current gold-standard methods.

INTRODUCTION

Intraoral scanners (IOS) can be of great convenience to the dental practitioner and their patients. Interest in capturing full arch scans has been growing over the past decade in an effort to increase the range of treatment modalities offered by the digital workflow. Investigations report a tendency for intraoral scanners to produce clinically acceptable digital impressions over short distances, while complete arch scans may suffer from distortions at a scale which may have potential clinical implications (1–4). Despite this, dental design software offers the ability to provide full arch prostheses based on intraoral scans. The onus is on the end-user to decide on appropriate use, based on the published evidence. Unfortunately, there is a lack of consensus on both the degree of trueness required for a particular procedure, and the metric by which to measure the said trueness.

For example, the level of trueness required for full-arch implant work is unresolved, with studies citing a range of 10 to 150 microns as minimal tolerance for the passive fit of an implant framework (5). Wismejer et al. report that "CAD/CAM technology has not eliminated the risks for hardware-related complications" in implant-supported reconstructions, implying that though single crowns and abutments can be reliably produced using CAD/CAM solutions, the current performance of the complete CAD/CAM workflow does not fall within the tolerance required for optimal full-arch implant work (6). As such, an awareness of the full-arch accuracy of a scanner could play a deciding part in whether or a not a clinician chooses to rely on a virtual impression

as means of data acquisition in the digital treatment workflow.

An important question arises in how best to measure the quality of full arch dental scans. Any measurements should give a good clinical indication of the potential quality of fit of a prosthesis. The precision (repeatability) of a scanner is a common metric, whilst trueness is also useful when correct values are known *a priori* (though this is rarely the case *in vivo*). However, how best to measure and report accuracy (meaning both trueness and precision) is much disputed; and robustly assessing the quality of 3D (three-dimensional) scan data in a clinically relevant manner is an unsolved problem.

The authors have previously presented evidence to highlight the inherent flaws in using the commonly used metric of global mean deviation between repeated scans as a measure of clinical appropriateness (7). Mean deviation is likely to report smaller scan errors than might be present in a scan, leading to an overly optimistic appraisal. Likewise, as all scan data produced by dental scanners have already undergone filtering and noise removal prior to output mesh generation (generally as an STL file), removing a portion of the most extremely deviating values of a data-set, as is often reported in the field, may result in artificially precise data prior to analysis. Further, any measurement relying on aligning multiple scans will be subject to a margin of error; as there is rarely a single, true solution to 3D alignment problems. Investigating errors such as those accumulated over a full-arch scan can be a challenge greatly affected by the artificial minimization of error, as investigated by O'Toole et al. (8), and may result in analyses underreporting global errors. This may further bring the relevance of reporting the mean deviation of meshes into question, as this value may reflect more on the success of the alignment algorithm to minimize deviation than the quality of the scan data.

Ender et al. (2019) investigated full arch versus segment scans (9). The authors report a generally higher precision in posterior segment scans. However, it might be suggested that the mean distance deviation reported for scan segments is likely to provide favorable results for the posterior segments, as straighter sections, as opposed to the curved anterior sections, are more likely to align in such a manner as to minimize any distance deviation between repeated scans. Because of this, aligning and measuring isolated segments discards most, if not all, cross arch error through optimal (though potentially incorrect) mesh alignment.

One possible solution to overcome the minimization of error caused by relying on scan alignments may be to forego measurements relying on global alignment, and instead measure distances between robustly identified key points within a single scan (10–12). This could provide an insight into any arch distortion introduced by the scanner during the scanning process, including location specific errors, without suffering from alignment

minimization artifacts. When considering the clinical fit of a full arch prosthesis, a metric such as cross-arch distance error might be considered more clinically relevant. The 2019 study (11) used specific mesh vertices as virtual key points, as opposed to introducing physical features of interest onto the scan object to investigate virtual occlusion (10,13).

This study compared the trueness values reported using three different analysis methods on the same two sets of intraoral scan data captured from two contemporary IOSs, with the aim of gaining an insight in the clinical applicability of the various methods. The IOSs used were Primescan and Omnicam (Dentsply Sirona) and the methods compared were A) the unsigned mean distance deviation, B) the linear distance between (virtual) key features on the dental arch, as described above (and referred to as inter-molar width), and C) the percentage surface area of a scan deviating beyond 0.1mm (a simplified version of the method reported by Wismeijer et al.) (6).

Null hypotheses:

That there is no significant difference between trueness, as measured using mean surface deviation compared to a reference scan, between the two test IOSs. That there is no significant difference between trueness, as measured using linear cross-arch distance deviation compared to a reference scan, between the two test IOSs. That there is no significant difference between trueness, as measured using the percentage of the surface deviating beyond 0.1mm compared to a reference scan, between the two test IOSs

MATERIALS AND METHODS

A dentate type IV stone maxillary model was scanned ten times with a Primescan, CEREC 5.0.0 (Dentsply Sirona) (P1 -P10), and ten times with an Omnicam, CEREC 4.6 (Dentsply Sirona) intraoral scanner (O1-O10), using the manufacturer's recommended scanning strategies. The scanned model had been poured more than 30 days prior to scanning.

All scans were recorded in one session by an experienced operator. Both scanners had been calibrated prior to scanning. All scans were exported as high-resolution STL (stereolithography) files.

To produce an indication of a trueness metric, the model was scanned once using a verified (VDI 2634/2) lab scanner (Rexcan DS2, Solutionix) which has a quoted resolution of <10μm when measured against the industry standard. All Primescan (P1–P10) and Omnicam (O1-O10) scans were aligned to the Solutionix scan. The alignment algorithm used was iterative closest point implemented using the freely available Open3D software (14), following a subsampling of all scans to produce pointclouds with a point distance no greater
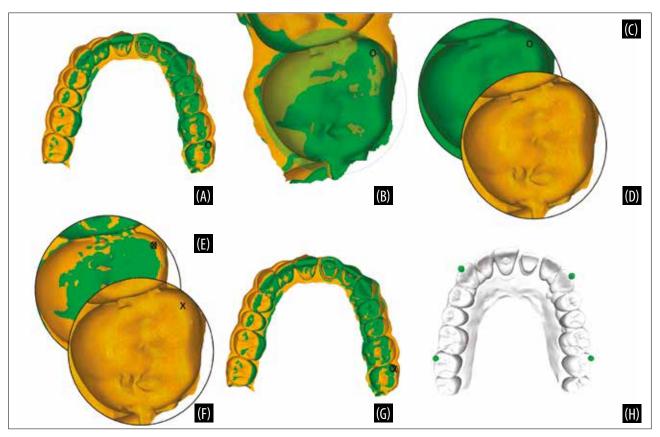
FIG. 1 Identifying anatomically identical keypoints on different scans to enable direct inter-arch distance comparisons. **A)** Following global alignment of a pair of scans, a keypoint, UL7, has been identified on one scan (labelled "o") and we wish to identify this virtual keypoint on all subsequent upper arch scans from both IOSs. The arch width discrepancy in full arch IOS scans leads to poor alignments if performed over the full arch. In step **B)** we take a small section (10mm radius) of the source arch, and a similar selection from the test arch **(C, D)** and finely align these two **(E)**. The precise location of the UL7 keypoint is then mapped onto the test arch **(E, F)**, (labelled "x"), and an inverse transformation is then applied to carry the keypoint back to the untransformed test arch **(G)**. This process was repeated for all four keypoints across the arch **(H)** on all scans.

than 25 microns. Once aligned, the scans were reverted to their original point spacing. All meshes were cropped identically using a cropping lasso (Contour Select, LDD) defined on P1, and applied to all 20 meshes. This ensured that all future measurements would be taken from identical regions across all scans (Fig. 1).

Topologically identical key points were identified on the upper right second molar (UR7) on the Solutionix scan and all 20 IOS scans using the method outlined by Gintaute et al. (11). Three more key points were similarly identified on UR3, UL3 and UL7.

Statistical analysis was done using SPSS Statistics 26 (IBM). Independent two-sample t-tests were used to assess the difference in trueness between the two intraoral scanners, across the three metrics investigated; mean deviation, inter-molar width and proportion beyond 0.1mm.

RESULTS

Surface comparison against Solutionix
The mean unsigned distance deviation between the verified Solutionix scan and the ten Primescan scans was 0.016($\pm$0.006)mm. The mean signed standard deviation for the Primescan was 0.021($\pm$0.009)mm .

The mean unsigned distance deviation between the verified Solutionix scan and the ten Omnicam scans was 0.116($\pm$0.01)mm. The mean signed standard deviation for the Omnicam was 0.158($\pm$0.025)mm. There was a significant difference between the unsigned distance deviations produced by the two intraoral scanners $p < 0.001$).

Inter-molar width
The arch width between the left and right molars was 55.44mm for the Solutionix scan. The arch width between the left and right molars was 55.439($\pm$0.075)mm for Primescan. This same distance was 54.672($\pm$0.065)mm for Omnicam. The perimeter distance of the single Solutionix scan was 152.40mm. The mean perimeter distance of the Primescan was 152.38($\pm$0.076)mm. The mean perimeter distance for the Omnicam was 151.29($\pm$0.06)mm (Table 1). There was a significant difference between the arch widths produced by the two intraoral scanners ($p < 0.001$).

	UR7 to UR3	UR3 to UL3	UL3 to UL7	UL7 to UR7 (Inter-molar width)
Primescan	-0.012 (±0.012)	-0.006 (±0.010)	-0.006 (±0.014)	-0.001 (±0.075)
Omnicam	-0.065 (± 0.018)	-0.199 (±0.017)	-0.085 (±0.012)	-0.768 (±0.065)

TABLE 1 Mean difference in key point distance from Solutionix scan (mm).

Upper bound deviation against Solutionix

The mean proportion of the Primescan scans deviating beyond 0.1mm when compared against the Solutionix was 0.7(±2.0)%.

The mean proportion of the Omnicam scans deviating beyond 0.1mm when compared against the Solutionix was 42.1(±2.5)%. The difference between the two intraoral scanners in proportion of scan deviating beyond 0.1mm from the Solutionix was significant (p<0.001).

DISCUSSION

This study investigated the full arch trueness of two intraoral scanners using three different methods of assessment. There was a significant difference in global mean unsigned deviation between the two scanners (p<0.001). There was a significant difference in inter-molar width recorded by the two scanners (p<0.001). There was a significant difference in proportion of scan deviating beyond 0.1mm from the Solutionix scan p<0.001). Thus, the null hypotheses must all be rejected. The Primescan produced data significantly closer to the verified lab scanner (an indication of trueness) compared

to Omnicam, both when assessing using mean unsigned deviation and linear cross arch distance, Omnicam consistently under-reported the linear intermolar width, in effect, narrowing the arch form. The Primescan reported only fractional amounts of scan data deviating beyond 0.1mm (0.7(±2.0)%) from the Solutionix scan, whereas an average of 42.1(±2.5)% of each Omnicam scan deviated beyond this distance.

All three metrics indicated that the Primescan produced truer data than the Omnicam. However, unlike the mean distance metric, the key point method gave a better intuition as to the potential quality of fit of a cross-arch prosthesis. For example, the casual reader might consider the mean unsigned deviation error of the Omnicam (0.116±0.01mm) to be clinically tolerable, envisaging that the fit of a cross-arch framework would require only a small adjustment. Conversely, the cross-arch linear error metric revealed that Omnicam consistently under-estimated the intermolar width by a much larger value of 0.768 (± 0.065)mm. This degree of framework inaccuracy would require significant chairside adjustment, or more likely, remaking. By contrast, the Primescan linear cross-arch error averaged -0.001 (±0.075) mim, which could more confidently be assumed to produce a well-fitting full arch prosthesis.
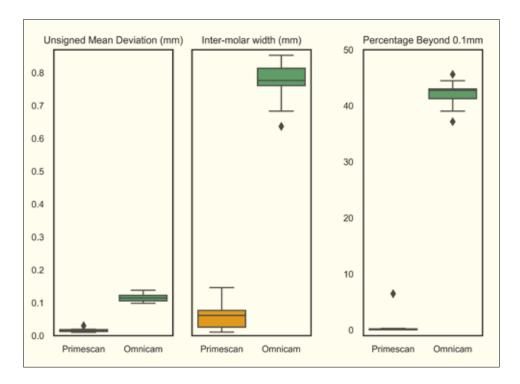

FIG. 2 Results from each of the three methods plotted. The line indicates median value, the box upper and lower quartile, while the whiskers show overall distribution. Outliers are indicated with a diamond.
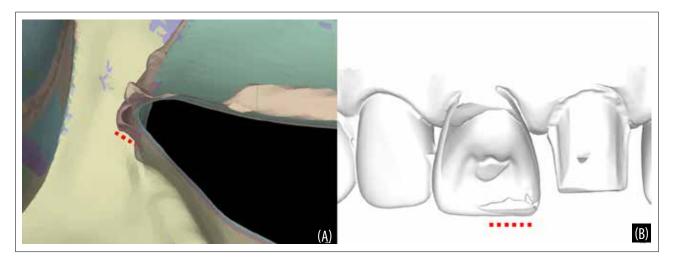
FIG. 3. Screencaptures displaying edge sharpening artefacts occurring on the Primescan scans. **A)** shows an occlusal view looking down on the prep and edge sharpening occurring on the distal aspect of the tooth preparation on two of the five scans displayed. The red dotted line is ≈ 0.2mm. This error would lead to a much larger cement thickness than desired in any crown produced, but the error would be difficult to detect. **B)** shows a cross-section of a Primescan scan, with a tunnel burrowing ≈ 4.4mm (red dotted line) across the mesial aspect of an anterior tooth (UR1). This would potentially have interfered with any CAD design, had the artefact occurred on a prepared tooth.

Hence the key point method appears to discriminate better between IOSs and their likely clinical potential.

Our key point method requires no physical placement of landmarks, making it simpler to implement than previous studies, which required fixed reference objects such as metal bars or spheres (10,15). Interestingly in the latter study, using spheres *in vivo* on 50 test subjects, an intermolar error of 0.828($\pm$0.265)mm for the Omnicam (measurement D1_4 in their paper) was reported. This agrees well with our value of 0.768 ($\pm$0.065)mm, with the slightly poorer trueness in the Kuhr et al. study perhaps being due to a combination of older Omnicam software and the fact their study was performed *in vivo*. This might also explain the lower precision in their study (as given by the standard deviation).

Our key point method could easily be employed to measure precision *in vivo*, which may be a more clinically informative metric than the commonly used mean surface deviation. However, the problem remains in employing our metric to measure trueness *in vivo*, in that we have no reference values for the key point separation distances. Given the evidence of numerous papers regarding good conventional impressions outperforming IOSs over full arches, it would seem appropriate to use physical silicone or polyether impressions as a 'gold standard' in future work, when attempting to assess IOS trueness over a full arch (4,9,15,16).

Our third test metric, percentage of surface lying beyond 0.1mm, may also hold value as a broad comparison of IOS accuracy and the user may select a threshold value appropriate to their needs. Here, we report that 42.1 $\pm$ 2.5% of the Omnicam scan surface lay beyond 0.1mm of the true value. That almost half the entire scan is poorer than 0.1mm might allow a clinician to make an informed choice on appropriate use. Conversely, Primescan (0.7 $\pm$ 2.0%) revealed a strong improvement in trueness as judged by this metric.

It is interesting to note that the noise in both scanners, as measured via the standard deviations across scans, did not differ significantly.

An inherent challenge in accuracy validation of intraoral scanners *in vivo* is the lack of a measurable reference. Hence, while *in vivo* scans can be used to measure precision and repeatability, trueness validation of *in vivo* scans can be challenging. As a result, a large number of intraoral scanner accuracy studies rely on *in vitro* studies. A number of previous studies demonstrate that intraoral scanning reduces scan accuracy, due to movement restrictions and the optically challenging environment within the oral cavity (17–19). Results obtained *in vitro* can therefore be assumed to be an optimal scenario and likely to produce artificially favorable conclusions.

Visual comparison between the Omnicam and Primescan scans made it evident that the Primescan data had undergone significant edge sharpening (20), resulting in artificially sharp margins and severe mesh artifacts. One such artifact, a tunnel burrowing half-way across the distal aspect of an anterior tooth would potentially have interfered with any CAD design, had the artifact occurred on a prepared tooth. There seems to be a commercial drive to make IOS scans appear better using digital enhancements (21).These algorithms are potentially risky, because clinicians will see a sharp looking scan, but it may no longer actually represent the patient. Our metrics will not inform with regards to these, and so further work is required to assess the local effects (for example on crown margin trueness) of edge enhancement and interproximal sharpening.

CONCLUSION

We present a comparison of three methods for assessing the quality of 3D data produced by two IOSs. The virtual keypoint, and percentage of scan deviating beyond 0.1mm methods may both give a clearer insight into clinical scanner trueness than the commonly reported unsigned mean surface deviation. Due to the virtual method of keypoint creation, the method can be used on scan data obtained both *in vitro* and *in vivo*.

Primescan produced significantly truer results than Omnicam, under all three metrics. Its clinical use over full arches would appear to be more appropriate than Omnicam. However, the Primescan was found to perform notable edge-sharpening, to the point of data deterioration; the clinical effect of this aspect of data manipulation should be investigated further.

DECLARATIONS

Availability of data and material

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

MF is editor for the Journal of Osseointegration. There are no other conflicts of interest.

Funding

Nothing to declare

Authors' contributions

JC collected all scan data. CO and AK analysed and interpreted the data and contributed equally in writing the manuscript. MF contributed to the writing of the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Chochlidakis KM, Papaspyridakos P, Geminiani A, Chen CJ, Feng IJ, Ercoli C. Digital versus conventional impressions for fixed prosthodontics: A systematic review and meta-analysis. J Prosthet Dent 2016;116:184–90. doi:10.1016/j.prosdent.2015.12.017.
2. Tsirogiannis P, Reissmann DR, Heydecke G. Evaluation of the marginal fit of single-unit , complete-coverage ceramic restorations fabricated after digital and conventional impressions : A systematic review and meta-analysis. J Prosthet Dent 2016;116:328-335.e2. doi:10.1016/j.prosdent.2016.01.028.
3. Ahlholm P, Sipilä K, Vallittu P, Jakonen M, Kotiranta U. Digital Versus Conventional Impressions in Fixed Prosthodontics: A Review. J Prosthodont 2018;27:35–41. doi:10.1111/jopr.12527.
4. Ender A, Attin T, Mehl A. In vivo precision of conventional and digital methods of obtaining complete-arch dental impressions. J Prosthet Dent 2016;115:313–20. doi:10.1016/j.prosdent.2015.09.011.
5. Kan JY, Rungcharassaeng K, Bohsali K, Goodacre CJ, Lang BR. Clinical methods for evaluating implant framework fit. J Prosthet Dent 1999;81:7–13. doi:10.1016/S0022-3913(99)70229-5.
6. Wismeijer, Daniel; Buser, Daniel; Chen S. Biological and Hardware Complications in Implant Dentistry. Quintessence Publishing; 2019.
7. Osnes CA, Wu JH, Venezia P, Ferrari M, Keeling AJ. Full arch precision of six intraoral scanners in vitro 2019. doi:10.1016/j.jpor.2019.05.005.
8. Toole SO, Osnes C, Bartlett D, Keeling A. Investigation into the accuracy and measurement methods of sequential 3D dental scan alignment. Dent Mater 2019;35:495–500. doi:10.1016/j.dental.2019.01.012.
9. Ender A, Zimmermann M, Mehl A. Accuracy of complete- and partial-arch impressions of actual intraoral scanning systems in vitro. Int J Comput Dent 2019;22:11–9.
10. Güth JF, Edelhoff D, Schweiger J, Keul C. A new method for the evaluation of the accuracy of full-arch digital impressions in vitro. Clin Oral Investig 2016;20:1487–94. doi:10.1007/s00784-015-1626-x.
11. Gintaute A, Keeling AJ, Osnes CA, Zitzmann NU, Ferrari M, Joda T. Precision of maxillo-mandibular registration with intraoral scanners in vitro. J Prosthodont Res 2019. doi:10.1016/j.jpor.2019.05.006.
12. Kühle A. Calibration Procedures for Atomic Force Microscopes. Adv Tech Assess Surf Topogr Dev a Basis 3D Surf Texture Stand "Surfstand" 2003:175–94. doi:10.1016/B978-190399611-9/50007-4.
13. Kuhr F, Schmidt A, Rehmann P, Wöstmann B. A new method for assessing the accuracy of full arch impressions in patients. J Dent 2016;55:68–74. doi:10.1016/j.jdent.2016.10.002.
14. Zhou Q-Y, Park J, Koltun V. Open3D: A Modern Library for 3D Data Processing 2018.
15. Kuhr F, Schmidt A, Rehmann P, Wöstmann B. A new method for assessing the accuracy of full arch impressions in patients. J Dent 2016;55:68–74. doi:10.1016/j.jdent.2016.10.002.
16. Ender A, Mehl A. Full arch scans : Conventional versus digital impressions - An in-vitro study. Int J Comput Dent 2011.
17. Flügge T V., Schlager S, Nelson K, Nahles S, Metzger MC. Precision of intraoral digital dental impressions with iTero and extraoral digitization with the iTero and a model scanner. Am J Orthod Dentofac Orthop 2013;144:471–8. doi:10.1016/j.ajodo.2013.04.017.
18. Kim J-H, Kim K-B, Kim S-H, Kim W-C, Kim H-Y, Kim J-H. Quantitative evaluation of common errors in digital impression obtained by using an LED blue light in-office CAD/CAM system. Quientessenz Int 2015;46.
19. Keeling A, Wu J, Ferrari M. Confounding Factors Affecting the Marginal Quality of an Intra-Oral Scan. J Dent 2017.
20. Fan R, Jin X. Controllable Edge Feature Sharpening for Dental applications. Comput Math Methods Med 2014. doi:http://dx.doi.org/10.1155/2014/873635.
21. Kim T, Cho Y, Kim D, Chang M, Kim Y. Tooth Segmentation of 3D Scan Data Using Generative Adversarial Networks 2020:1–17. doi:10.3390/app10020490.