Influence of Scanner Precision and Analysis Software in Quantifying Three-Dimensional Intraoral Changes: Two-Factor Factorial Experimental Design

Saoirse O'Toole¹, BA, BDentSc, PhD, MFD RCSI; David Bartlett¹, BDS (Hons), PhD, FDSRCS; Andrew Keeling², BSc (Hons), BDS, PhD; John McBride³, BSc (Hons), PhD; Eduardo Bernabe¹, BDS, PhD; Luuk Crins⁴, MSc, DDS; Bas Loomans⁴, PhD, DDS

¹Centre for Clinical, Oral and Translational Sciences, King’s College London, London, United Kingdom
²University of Leeds, Leeds, United Kingdom
³University of Southampton, Southampton, United Kingdom
⁴Radboud University Medical Centre, Nijmegen, Netherlands

Corresponding Author:
Saoirse O'Toole, BA, BDentSc, PhD, MFD RCSI
Centre for Clinical, Oral and Translational Sciences
King’s College London
London
United Kingdom
Phone: 44 2071887462
Email: saoirse.otoole@kcl.ac.uk

Abstract

Background: Three-dimensional scans are increasingly used to quantify biological topographical changes and clinical health outcomes. Traditionally, the use of 3D scans has been limited to specialized centers owing to the high cost of the scanning equipment and the necessity for complex analysis software. Technological advances have made cheaper, more accessible methods of data capture and analysis available in the field of dentistry, potentially facilitating a primary care system to quantify disease progression. However, this system has yet to be compared with previous high-precision methods in university hospital settings.

Objective: The aim of this study was to compare a dental primary care method of data capture (intraoral scanner) with a precision hospital-based method (laser profilometer) in addition to comparing open source and commercial software available for data analysis.

Methods: Longitudinal dental wear data from 30 patients were analyzed using a two-factor factorial experimental design. Bimaxillary intraoral digital scans (TrueDefinition, 3M, UK) and conventional silicone impressions, poured in type-4 dental stone, were made at both baseline and follow-up appointments (mean 36 months, SD 10.9). Stone models were scanned using precision laser profilometry (Taicaan, Southampton, UK). Three-dimensional changes in both forms of digital scans of the first molars (n=76) were quantitatively analyzed using the engineering software Geomagic Control (3D Systems, Germany) and freeware WearCompare (Leeds Digital Dentistry, UK). Volume change (mm³) was the primary measurement outcome. The maximum point loss (μm) and the average profile loss (μm) were also recorded. Data were paired and skewed, and were therefore compared using Wilcoxon signed-rank tests with Bonferroni correction.

Results: The median (IQR) volume change for Geomagic using profilometry and using the intraoral scan was −0.37 mm³ (−3.75-2.30) and +0.51 mm³ (−2.17-4.26), respectively (P<.001). Using WearCompare, the median (IQR) volume change for profilometry and intraoral scanning was −1.21 mm³ (−3.48-0.56) and −0.39 mm³ (−3.96-2.76), respectively (P=.04). WearCompare detected significantly greater volume loss than Geomagic regardless of scanner type. No differences were observed between groups with respect to the maximum point loss or average profile loss.

Conclusions: As expected, the method of data capture, software used, and measurement metric all significantly influenced the measurement outcome. However, when appropriate analysis was used, the primary care system was able to quantify the degree of change and can be recommended depending on the accuracy needed to diagnose a condition. Lower-resolution scanners may underestimate complex changes when measuring at the micron level.
**Introduction**

All clinicians should be able to quantify and assess whether a degenerative health condition is stable or progressing. This is possible in some diseases that have accurate biomarkers but is not always possible for diseases of the soft and hard tissues. Measurement has typically taken the form of recording subjective visual changes; therefore, physical measurements of change are needed.

In dentistry, quantitative measurement of differences between sequential 3D scans of the teeth is typically used to diagnose erosive tooth wear. This is a condition in which excesses acids from the diet and stomach can dissolve the teeth. Due to changes in diet and health, the prevalence of erosive tooth wear has been increasing, now affecting 1 in every 3 adults globally [1]. Quantitative validation of tooth wear has only been possible in university hospitals thus far [2-5]. This validation is achieved by scanning accurate molds of the teeth with laser profilometers to create a precise digital map of the surface with repeatable, calibrated point coordinates. As directly scanning teeth with lab-based profilometers has not been possible, scans of molds of the teeth have been aligned and compared using custom-built or commercial engineering software to quantify changes. Data capture is typically very accurate with this approach, and small process errors have been calculated to be in the range of 15 microns [6,7]. However, the reliance on research laboratory–based scanners and complex engineering analysis software is expensive and unfeasible for use in primary care settings [6-8].

Digital handheld scanners, known as intraoral scanners, take digital maps of the teeth and are increasingly being used in primary care. Intraoral scanners do not generate aerosols and are more amenable to effective cross-infection control compared to conventional impressions that generate aerosols and can harbor pathogenic microorganisms [9]. These advances are particularly relevant for ensuring complete disinfection during the current COVID-19 pandemic. Intraoral scanners capture data via different methods, ranging from video capture to the use of confocal, triangulation, or active waveform principles. Rather than relying on accurate calibrated data point collection on an unmoving subject, multiple data points are captured and stitched together with company-specific algorithms. Errors are generated when the scanner fails to collect sufficient data to stitch a digital map of the surface (undersampling) [10,11] or when the process fails, particularly with more than one tooth [12,13]. Furthermore, data stitching algorithms often interpolate or smooth missing or erroneous data; therefore, the data points are estimated, nonuniform, and lack adequate surface detail for changes to be measured at the micron level.

The software currently used by commercial companies to analyze digital maps rely on an iterative closest point (ICP) algorithm to merge the maps to the closest possible alignment, without considering if the proposed alignment solution makes biological sense [14]. We previously demonstrated that this estimation leads to distortions and can result in physiologically impossible outcomes [15,16]. We recently incorporated feature-recognizing elements [16,17] into an ICP algorithm to minimize these errors and created an open-source freeware to be used alongside any 3D scan. Although this method has been validated against previous gold-standard software [18], it has not yet been tested on longitudinal clinical data.

The combination of data collection from primary care and free, user-friendly software for analysis may create new opportunities for monitoring disease. However, the accuracy of measuring change in scans will be influenced by the scanner, software, and their interaction. In this study, we used a factorial design to compare data obtained from profilometric scans of casts and those obtained from direct intraoral scans using two types of registered software: commercial software (Geomagic Control, 3D Systems, Germany) and freeware (WearCompare, Leeds Digital Dentistry, UK). We expected to see differences in the measurements obtained between the scanners but we did not know whether this difference would be clinically significant. The primary null hypothesis was that the dental wear data, specifically the volume change, average profile loss, and maximum point loss, detected by the profilometer will not be different to those obtained with the intraoral scanner. The secondary null hypothesis was that the software used to analyze the data will not influence the volume change, average profile loss, and maximum point loss observed for either scanner.

**Methods**

**Participants**

Data were collected from a larger clinical longitudinal erosive tooth wear study (Radboud Tooth Wear Project ABR code: NL31371.091.10) [19,20]. Study participants had been referred by general dental practitioners to the Department of Dentistry of Radboud University Medical Center (Nijmegen, the Netherlands) for management of erosive tooth wear. Patients in the monitoring arm who provided additional written consent for their data to be transported to the United Kingdom (ABR codes NL31401.091.10) and to perform additional analyses were included in the study (n=25; mean age 35.8, SD 6.8 years; 20 men and 5 women). A power calculation was performed in GPower vs 3.1 [21] using a two tailed test, demonstrating that to obtain a correlation of 0.4 between the scanners at 95% power with P<.05, a sample size of 75 was required.

**Digital and Dental Impressions**

Data transported to the United Kingdom included digital intraoral scans obtained using Lava Chairside Oral Scanner (3M, USA) at baseline and 3M TrueDefinition Intraoral Scanner (3M, UK) at follow up, and analog dental impressions taken with silicone (Ivoclar Virtual 380, Ivoclar Vivodent, Liechtenstein, Europe). Impressions were poured in type-3
dental stone (SLR Dental GmbH, Germany) within 24 hours according to the manufacturer’s instructions.

Both the digital and dental impressions were captured by the same trained operator. The point clouds of recognized index teeth (ie, the occlusal surface of the first molars [22,23]) were isolated by the operator (ST) and set aside for evaluation. Each analog study model (n=100) was scanned using a noncontact triangulation laser profilometer (XYRIS 2000TL, Taicaan Technologies, Southampton, UK) in a raster pattern using a step-over of 50 µm with a repeatability error of 2.6 µm [24]. This generated a 3D point cloud dataset for comparison.

Measurements and Software
Quantitative analysis of the change between sequential scans from the profilometer and intraoral scanner was performed using both the commercial software Geomagic Control 2011 (Geomagic, Morrisville, NC, USA) and the freeware WearCompare (Leeds Digital Dentistry, Leeds, UK). Data points selected by the operator on the buccal and lingual surfaces were chosen as reference areas and used for analysis according to previously published protocols [16]. For Geomagic, a best-fit alignment of 1000 data points on reference surfaces, followed by a refined alignment using 5000 data points, was performed. For the reference alignment, the occlusal surface was deleted from the dataset, leaving the buccal and lingual reference surfaces. The transformation matrix was then applied to the complete displaced dataset to realign it with the same orientation. For WearCompare, an initial global alignment utilizing a feature-based recognition system was performed. The same buccal and lingual reference surfaces were selected for refined ICP alignments, which highlights corresponding reference areas within 25 microns of each other. The occlusal surface was selected to be measured and all measurements were taken perpendicular to the occlusal surface.

Volume change (mm$^3$), maximum point loss (µm), and the mean loss over the surface (µm) were analyzed for each surface for both scanners and software types. As a secondary volumetric analysis, any positive values, indicating either gain or error, were set to zero.

Statistical Analysis
This study utilized a two-factor factorial experimental design comparing two different methods of data capture (profilometer and intraoral scanner) and two different analysis software types with different alignment principles (Geomagic Control and WearCompare). Descriptive statistics of all measurement metrics were calculated, and normality was assessed using the Shapiro-Wilks test and histogram assessment. Since the data were calculated, and normality was assessed using the

### Results

From the original data collected in the Netherlands, 76 surfaces were analyzed representing an average follow-up time of 36 months (SD 10.9). The data were initially analyzed using the previous gold-standard commercial software Geomagic. Laboratory profilometry data analyzed in Geomagic showed a median volume loss of –0.37 mm$^3$ (IQR –3.75-2.30), whereas a median volume gain was observed for the intraoral scan data of +0.51 mm$^3$ (IQR –2.17-4.26), representing a significant difference (P<.001). The median profile loss was 55.8 µm (IQR 24.43-77.60) and 43.65 µm (IQR 29.93-77.95) for the laboratory profilometer and intraoral scanner, respectively (P<.001). The maximum point loss on the occlusal surface was 398.4 µm (IQR 238.7-533.7) and 303.9 µm (IQR 217.6-483.0) for the profilometer and intraoral scan, respectively (P=0.1).

Data from the freeware WearCompare showed a median volume change for the profilometer scan of –1.21 mm$^3$ (IQR –3.48-0.56) and –0.39 mm$^3$ (IQR –3.96-2.76) for the intraoral scan (P=0.04). The median profile loss was 44.80 µm (IQR 29.48-91.63) and 43.10 µm (IQR 24.43-77.60) for the profilometer scan and intraoral scan, respectively (P=0.18). The median maximum point loss on the occlusal surface for the profilometer scan was 317.1 µm (IQR 198.0-466.4) and was 278.3 µm (IQR 170.8-494.0) for the intraoral scan (P=0.77). Therefore, no statistically significant differences were observed between the profilometer scans and intraoral scans when measurements were analyzed in WearCompare.

However, WearCompare detected significantly greater volume loss than analysis in Geomagic (P<.001), regardless of the scanner type. There were no differences between software in terms of average profile loss (P=.28) or maximum point loss (P=.26). When positive values were set to zero, the median volume change for the profilometer was unchanged, whereas intraoral scan volume loss evident. A significant difference was observed between the profilometry and intraoral scan data for Geomagic analysis (P=.02) but not for WearCompare analysis (P=.36) (Table 1).

### Table 1. Volume changes (mm$^3$) observed over 3 years when positive data were set to zero, which is a commonly used method by many commercial companies.

| Scan type                | Geomagic       | WearCompare   | P value |
|--------------------------|----------------|---------------|---------|
| Laser profilometer, median (IQR) | –0.37 (–3.75-0.00) | –1.21 (–3.48-0.00) | .30     |
| Intraoral scanner, median (IQR)    | 0.00 (–2.17-0.00) | –0.39 (–3.95-0.00) | .09     |
Moderate ICCs were observed in analyzing volume change data between the scanners and software types: 0.476 (95% CI 0.281-0.632) for Geomagic and 0.457 (95% CI 0.259-0.618) for WearCompare (both P<.001). WearCompare and Geomagic data showed slightly stronger ICC using the intraoral scanner (0.673, 95% CI 0.529-0.780; P<.001) than with the profilometer (0.525, 95% CI 0.341-0.671, P<.001).

Discussion

Principal Findings

This study demonstrates the differences in outcomes that can be observed when using low-resolution primary care digital scanners and precision measurements from hospital laboratory profilometers for measuring biological changes at the micron level. As expected, increased volume change values were observed using the higher-resolution and calibrated profilometer scans compared to the intraoral scans. Surprisingly, this difference was only statistically significant when using commercial software, previously considered to be the gold standard, for the analysis. The custom-built freeware outperformed the commercial software. The null hypothesis was therefore partially rejected. This finding suggests that if the analysis is conducted accurately, it may compensate for the decreased resolution of the scanner. This is a promising finding and has implications for the development of primary care systems.

There are several possible reasons for the reduced volume changes observed with intraoral scanners. Data interpolation or the mathematical averaging of data points across a surface can smoothen the topography of the surface and may overlook small discrepancies/areas of change in the surface. Smooth surface lesions, potentially on the buccal and lingual reference areas, will be subjected to heavy data undersampling internally in the scanner as the topography is not deemed to be as important. Smoothened surfaces are more susceptible to inaccuracies in data registration and alignment [15] as they will increase the mathematical tendency to minimize differences toward any sloped surfaces (in this case the occlusal surface). This can result in inaccuracies in alignment and biologically implausible outcomes. Analysis in a software that ignores features or the holistic geometric shape, such as Geomagic used in this study, will be particularly susceptible to this effect. Combining Geomagic analysis with the intraoral scan data resulted in an overall volumetric tooth tissue gain, which is physiologically impossible, indicating large errors within the analysis process.

For WearCompare, errors did not occur to the same extent, resulting in overall negative values for wear progression in both the profilometry and intraoral scans. Recent techniques developed at Radboud University involve using reference areas for alignment on the occlusal surfaces in addition to the buccal and lingual surfaces, providing additional control of the alignment along the Z axis. This may facilitate less translation and angular errors, and consequently less positive values. However, this comes at a tradeoff of increased analysis time and may also underestimate wear if an ICP algorithm is used for the Z axis. Further research will focus on validating this technique.

The correlation between wear measurements taken with the scanners was moderate as there are inherent but different errors for each form of data capture. Undetected, subvisual errors on casts or scans may have been present and subsequently analyzed as wear data. The profilometer is unable to scan undercuts, which means that less of the surface area can be used for selective surface alignment. In contrast, the intraoral scanner was successful at scanning undercuts. However, missing data or incomplete intraoral scans can also create errors whereby the triangle size is distorted and measurements can be skewed [25]. Recognizing where such errors may lie in each scan type will facilitate more accurate analysis.

Differences were observed between the profilometer and intraoral scan data when positive values were omitted from the analysis. Discounting positive values, which is commonly done with many types of commercial software and when profile loss and maximum point loss measurement metrics are reported, do not show error within the system. We observed that this can cause clinically significant changes to the outcome. Color maps of aligned scans can visually indicate areas of change, but the quantification does not always reflect the severity of progression. Discounting positive data increases the likelihood that a poor alignment will not be detected and wear underestimation or overestimation can occur. Reporting the negative changes may only be useful when trying to communicate wear to patients, but these metrics have limited diagnostic potential when measuring successive rates of wear.

Limitations

This study has several limitations. The analysis was performed blinded to the sequence of scanning to limit bias. However, there were often indications of sequence, such as surface restorations or clear visual wear progression in the interim period. Although only one model of intraoral scanner was used, the hardware and software changed over the 3-year period of the study, emphasizing that research in this fast-moving field becomes rapidly outdated. Other intraoral scanners will have slightly different methods of processing missing data and interpolating irregularities, and it is possible that slightly different results may be achieved with different intraoral scanners. Single-tooth analysis was performed to maximize accuracy, which also limits generalizability to full-arch analysis. A large limitation in longitudinal wear analysis is that the true wear progression is unknown. One has to assume that wear has occurred and positive values represent errors in alignment or in the data capture process. This makes it difficult to identify any form of measurement as a gold standard.

Conclusion

This study shows that low-resolution scanners can be used for measurements at the micron level provided appropriate analysis techniques and software are used. This could represent a step change in the way that erosive tooth wear is diagnosed and treated. From a dental point of view, the ability to view digital scans with increased magnification on a monitor also offers an increased diagnostic advantage. However, there is a duty of care on the profession and research community to not overestimate the quantitative capabilities of digital scanners to inform treatment or care outcomes until we are certain that they are
adequately sensitive and specific to do so. This will depend on the level of accuracy required from the analysis process to diagnose disease progression within a feasible diagnostic window. The resolution and accuracy of primary care scanning tools is likely to increase rapidly and further work should concentrate on reducing the process errors inherent within each measurement system.

Acknowledgments
This study was funded by the Academy of Medical Sciences (REF: SGL020/1082).

Conflicts of Interest
None declared.

References
1. Schlueter N, Luka B. Erosive tooth wear - a review on global prevalence and on its prevalence in risk groups. Br Dent J 2018 Mar 09;224(5):364-370. [doi: 10.1038/sj.bdj.2018.167] [Medline: 29495027]
2. Rodriguez J, Austin R, Bartlett D. In vivo measurements of tooth wear over 12 months. Caries Res 2012;46(1):9-15. [doi: 10.1159/000334786] [Medline: 22156738]
3. Mitchell HL, Chadwick RG, Ward S, Manton SL. Assessment of a procedure for detecting minute levels of tooth erosion. Med Biol Eng Comput 2003 Jul;41(4):464-469. [doi: 10.1007/bf02348909]
4. Pintado MR, Anderson GC, DeLong R, Douglas WH. Variation in tooth wear in young adults over a two-year period. J Prosthet Dent 1997 Mar;77(3):313-320. [doi: 10.1016/0022-3913(97)70189-6]
5. O’Toole S, Newton T, Moazzez R, Hasan A, Bartlett D. Randomised Controlled Clinical Trial Investigating The Impact of Implementation Planning on Behaviour Related to The Diet. Sci Rep 2018 May 23;8(1):8024-8026. [doi: 10.1038/s41598-018-26418-0] [Medline: 29795123]
6. Rodriguez JM, Bartlett DW. The dimensional stability of impression materials and its effect on in vitro tooth wear studies. Dent Mater 2011 Mar;27(3):253-258. [doi: 10.1016/j.dental.2010.10.010] [Medline: 21075442]
7. Chadwick RG, Mitchell HL, Ward S. Evaluation of the accuracy and reproducibility of a replication technique for the manufacture of electroconductive replicas for use in quantitative clinical dental wear studies. J Oral Rehabil 2002 Jun;29(6):540-545. [doi: 10.1046/j.1365-2842.2002.00894.x] [Medline: 12071922]
8. Rodriguez J, Austin R, Bartlett D. A method to evaluate profilometric tooth wear measurements. Dent Mater 2012 Mar;28(3):245-251. [doi: 10.1016/j.dental.2011.10.002] [Medline: 22094292]
9. Barenghi L, Barenghi A, Cadeo C, Di Blasio A. Innovation by Computer-Aided Design/Computer-Aided Manufacturing Technology: A Look at Infection Prevention in Dental Settings. Biomed Res Int 2019;2019:6092018. [doi: 10.1155/2019/6092018] [Medline: 31467901]
10. Shembesh M, Ali A, Finkelman M, Weber H, Zandparsa R. An In Vitro Comparison of the Marginal Adaptation Accuracy of CAD/CAM Restorations Using Different Impression Systems. J Prosthodont 2017 Oct 08;26(7):581-586. [doi: 10.1111/jopr.12446] [Medline: 26855686]
11. Keeling A, Wu J, Ferrari M. Confounding factors affecting the marginal quality of an intra-oral scan. J Dent 2017 Apr;59:33-40. [doi: 10.1016/j.jdent.2017.02.003] [Medline: 28189718]
12. 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 Sep;20(7):1487-1494. [doi: 10.1007/s00784-015-1626-x] [Medline: 26454734]
13. Giménez B, Özcan M, Martínez-Rus F, Pradíes G. Accuracy of a digital impression system based on parallel confocal laser technology for implants with consideration of operator experience and implant angulation and depth. Int J Oral Maxillofac Implants 2014 Jul;29(4):853-862. [doi: 10.11607/jomi.3343] [Medline: 25032765]
14. Mitchell HL, Chadwick RG. Mathematical shape matching as a tool in tooth wear assessment--development and conduct. J Oral Rehabil 1998 Dec 22;25(12):921-928. [doi: 10.1046/j.1365-2842.1998.00324.x] [Medline: 9888227]
15. Mitchell HL, Koch I, Chadwick RG. Linear interpolation error in measured surfaces in a dental erosion study. Med Biol Eng Comput 2004 Jan;42(1):100-105. [doi: 10.1007/bf02351017]
16. O’Toole S, Osnes C, Bartlett D, Keeling A. Keeling D. Investigation into the accuracy and measurement methods of sequential 3D dental scan alignment. Dent Mater 2019 Mar;35(3):495-500. [doi: 10.1016/j.dental.2019.01.012] [Medline: 30683418]
17. Zhou QY, Park J, Koltun V. Fast global registration. Cham: Springer; 2016 Presented at: 14th European Conference on Computer Vision; October 2016; Amsterdam, Netherlands p. 766-782. [doi: 10.1007/978-3-319-46475-6_47]
18. O’Toole S, Osnes C, Bartlett D, Keeling A. Investigation into the validity of WearCompare, a purpose-built software to quantify erosive tooth wear progression. Dent Mater 2019 Oct;35(10):1408-1414. [doi: 10.1016/j.dental.2019.07.023] [Medline: 31402133]
19. Sterenberg Bamm, Bronkhorst EM, Wetselaar P, Lobbezoo F, Loomans BAC, Huysmans MDNJ. The influence of management of tooth wear on oral health-related quality of life. Clin Oral Investig 2018 Sep 3;22(7):2567-2573 [FREE Full text] [doi: 10.1007/s00784-018-2355-8] [Medline: 29397468]
20. Loomans B, Opdam N. A guide to managing tooth wear: the Radboud philosophy. Br Dent J 2018 Mar 09;224(5):348-356. [doi: 10.1038/sj.bdj.2018.164] [Medline: 29495023]

21. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]

22. Martignon S, López-Macías AM, Bartlett D, Pitts N, Usuga-Vacca M, Gamboa LF, et al. The use of index teeth vs. full mouth in erosive tooth wear to assess risk factors in the diet: A cross-sectional epidemiological study. J Dent 2019 Sep;88:103164. [doi: 10.1016/j.jdent.2019.07.002] [Medline: 31276748]

23. O’Toole S, Lau JS, Rees M, Warburton F, Loomans B, Bartlett D. Quantitative tooth wear analysis of index teeth compared to complete dentition. J Dent 2020 Jun;97:103342. [doi: 10.1016/j.jdent.2020.103342] [Medline: 32360512]

24. Rodriguez JM, Bartlett DW. A comparison of two-dimensional and three-dimensional measurements of wear in a laboratory investigation. Dent Mater 2010 Oct;26(10):e221-e225. [doi: 10.1016/j.dental.2010.07.001] [Medline: 20701961]

25. Nedelcu R, Olsson P, Nyström I, Thor A. Finish line distinctness and accuracy in 7 intraoral scanners versus conventional impression: an in vitro descriptive comparison. BMC Oral Health 2018 Feb 23;18(1):27-11 [FREE Full text] [doi: 10.1186/s12903-018-0489-3] [Medline: 29471825]

Abbreviations

ICC: intraclass correlation
ICP: iterative closest point