Quantitative assessment of interproximal tooth reduction performed as part of Invisalign<sup>®</sup> treatment in 10 orthodontic practices

Tony Weir, Arun Shailendran, Brett Kerr and Elissa Freer
School of Dentistry, The University of Queensland, Brisbane, Australia

Background: Interproximal reduction (IPR) is a treatment option for orthodontic space gain. The attainment of prescribed objectives in aligner treatment may require IPR that is accurately performed both qualitatively and quantitatively.

Objective: This study assesses the in vivo accuracy of IPR carried out in 10 orthodontic practices as a method of orthodontic space creation.

Methods: A comparison of proposed and achieved amounts of IPR completed (accuracy), the accuracy of IPR within and between upper and lower dental arches, and the accuracy of IPR within and between posterior and anterior arch segments were performed using 3-dimensional digital study models gained via Align's<sup>®</sup> ClinCheck.

Results: The findings indicated that IPR was routinely underperformed by all practices studied. On average, the amount of IPR achieved represented only 44.0% of the total prescribed per tooth in the sample assessed, with a mean discrepancy of 0.16 mm per tooth. There were statistically significant differences only between the overall anterior and posterior groups (p < 0.01) and between maxillary anterior and maxillary posterior groups (p < 0.01); however, these were not clinically significant. Significant differences in IPR performance were noted between different orthodontic practices.

Conclusions: This study demonstrates that the clinical performance of IPR in 10 orthodontic practices consistently fails to achieve the prescribed amount often by large variations. The effect of this under-performance on clinical outcomes remains to be quantified.


Received for publication: May, 2021
Accepted: July, 2021

Tony Weir: anthony.weir@uq.edu.au; Arun Shailendran: a.shailendran@uq.edu.au; Brett Kerr: uqbkerr@ud.edu.au; Elissa Freer: e.freer@uq.edu.au

Introduction

The creation of space in the dental arches is often necessary to achieve orthodontic tooth alignment. The interproximal reduction of teeth (IPR), also known as enamel stripping, reproximation, or slenderisation, is one proposed method to gain space, which involves the removal of enamel from the contact areas of teeth. While gaining space for tooth alignment is a primary indication for IPR, it may also be employed to correct tooth size discrepancies, improve stability, and for aesthetic considerations such as predictable reduction of open gingival embrasures (OGE).

When introduced in 1944, IPR was suggested for use mainly in the anterior segment when a lack of balance presented. Begg’s Stone Age man’s dentition study in 1954 provided a theoretical justification for enamel removal as it simulated the shortening of the dental arch that occurred through natural interproximal attrition. This was later supported by Sheridan, who proposed that IPR, using an air-rotor technique, simulated the natural processes of interproximal wear over time. In 1956, Hudson introduced the first clinical sequence of IPR, employing medium and fine metallic strips for mesiodistal reduction, followed by final polishing and topical fluoride application.
IPR produces no harmful side effects as long as adequate cooling is used during the procedure and the tooth surface is left smooth and self-cleansing.\textsuperscript{9} Adequate polishing after IPR has the potential to leave treated enamel smoother than untreated enamel,\textsuperscript{6–10} thereby reducing the risk of plaque accumulation and subsequent pathology.\textsuperscript{5,11–14}

IPR has been indicated in the following circumstances:

1. Provision of space for alignment when extraction would result in excessive space and non-extraction treatment would result in undue arch expansion.
2. When individual tooth sizes prevent Class I molar and canine relationships from being established.\textsuperscript{15}
3. To improve the shape or contact areas of teeth.
4. To reduce unsightly OGE or ‘black triangles’. OGE are reported as present in 41.9\% of adolescent patients treated for upper central incisor crowding.\textsuperscript{16}
5. To increase post-orthodontic alignment stability. The flattening of contact points in the lower anterior region has been proposed as a method to reduce or prevent relapse because of the proximation of the flat contacts of the orthodontically aligned teeth.\textsuperscript{17–20}

Considerations for IPR include:

- Poor oral hygiene: IPR may roughen the remaining enamel surface, potentially allowing biofilms to accumulate more readily.\textsuperscript{21} However, IPR completed with adequate polishing has been shown to leave enamel smoother than untreated enamel.\textsuperscript{8,10}
- Dental factors: teeth with reduced interproximal enamel thickness or that experience hypersensitivity to cold should not undergo IPR. Furthermore, IPR performed when contacts are poorly aligned may lead to the removal of enamel from undesired locations, giving poor proximal contacts and embrasures, and potentially resulting in aesthetic problems. An investigation of the risk of developing dental caries after IPR has shown that there is no increased risk.\textsuperscript{5,11–14} Periodontal health is similarly reported as unaffected.\textsuperscript{14,17,22,23}

The performance of IPR demands precise clinical skills to achieve the desired reduction of tooth increments which may be as small as 0.1 mm. If enamel is over-reduced, residual space may appear, which seldom spontaneously closes. If under-reduced, the appropriate amount of space desired for tooth movement is not available and alignment may not be achieved.

Clinicians may plan to achieve their intended treatment goals by performing IPR, in which case it may be important to reduce the enamel by the exact amount required. Although several studies have focused on the surface irregularities that could remain after IPR, few studies provide a quantitative evaluation of the removed enamel.\textsuperscript{6,24–27} An in vitro study by Johner et al.\textsuperscript{25} investigated the predictability of the anticipated amount of IPR using three commonly employed IPR techniques on premolars. It was concluded that the actual IPR was typically less than the intended amount, and that the IPR technique did not appear to be a significant predictor of the actual amount of enamel reduced. Nevertheless, there are several limitations in extrapolating such in-vitro outcomes to clinical practice. Replicating biological systems surrounding the teeth in an in vitro study can be challenging, since it is extremely difficult to reproduce the dynamically active periodontal apparatus, tongue, cheeks, and other patient related factors.

An unpublished in vivo study by Blandy et al.\textsuperscript{26} assessed IPR accuracy and found that overall enamel reduction was under-achieved by 55.9\% per tooth surface. The authors concluded that IPR might not be performed with a high degree of clinical accuracy and this could contribute to compromised orthodontic treatment outcomes. A recent study by De Felice et al.\textsuperscript{27} also showed similar results, with accuracy of IPR in the lower arch found to be 37.02\%, and in the upper arch to be 44.95\%. However, both studies had relatively small sample sizes and perhaps were underpowered to detect the small differences in tooth widths before and after IPR. In addition, the operators in the De Felice study, were not blinded, while in Blandy et al.’s study, the IPR data was collected from a single operator.

Numerous studies have shown shortfalls in tooth alignment achieved with aligner treatment, either at the end of treatment or when compared to the digitally prescribed outcome.\textsuperscript{28–32} One potential reason for the inability to achieve tooth alignment is failure to provide sufficient space within the aligners for the teeth to fit. Any shortfall in interproximal reduction could clearly compromise the ability of teeth to fit within an aligner. Therefore, the present study aimed to evaluate the accuracy of IPR using a large sample size with patient data obtained from multiple practitioners and with the operators blinded during measurement to minimise bias. Proposed and achieved amounts of IPR completed (accuracy)
will be assessed from 10 orthodontic practices, and further divided into maxillary, mandibular, anterior, and posterior sub-groups.

Methods
The Australasian Aligner Research Database (AARD) of Invisalign cases has been collected for research purposes. As of July 2020, comprising approximately 10,000 cases, AARD held patient databases for 12 orthodontic practices, whose practitioners had significant experience with orthodontics in general and Invisalign treatment in particular. At the time of this study, sufficient data were available in AARD from 10 practices. Cases were accessed at random until a maximum of 25 per orthodontist were identified and selected based on the below criteria to ensure sufficient sample sizes for all sub-groups. For orthodontic practice A only 22 cases were available. The selection criteria were:

- ClinCheck files available at pre-treatment (T₁) and at the completion of initial Aligners (T₂).
- All IPR carried out as programmed between T₁ and T₂.
- Minimum of 1 mm IPR in total per case.
- Minimum of two sites for IPR.

ClinCheck files were de-identified by the Gatekeeper (TW), and participating orthodontic practice databases were randomly assigned alphabetic labels.

Pilot studies
An initial pilot study was conducted on 40 typodont teeth that were involved in previous research into the accuracy of ClinCheck Pro (CCPro) measurements. This study found that CCPro measurements differed significantly from either calliper or 3D Optical Profilometry (3D OP) values, and concluded that CCPro numbers should not be relied upon. Geomagic Control X software (3D systems, North Carolina, USA; Version 2017.0.3) was used by two of the investigators on this pilot sample to evaluate its accuracy when compared to the gold standards of callipers or 3D OP (see Table I). Although the results showed that Geomagic was statistically comparable to other methods, inter-examiner reliability was weaker.

While scrutinising the Geomagic results, an analysis of measurements on the 40 teeth showed that, contrary to the earlier finding by Shailendran et al., CCPro values now closely approximated those of the gold standard protocols. Reloading the pilot study .adf files into an older version of CCPro (ClinCheck Pro v5.1.1.21 NA) disclosed that the algorithms used to measure tooth widths had been changed in the newer software (ClinCheck Pro v5.7.3.93 INT), as illustrated in Figure 1.

The pilot study was therefore expanded to include all 120 typodont teeth used in the previous study by Shailendran et al. A reliability analysis of these data using ICC and Bland-Altman assessments is shown in Table II and Figure 2. Since the accuracy of the current CCPro was now confirmed, it was determined for the present study that those values would be used instead of Geomagic determinations.

To estimate the sample size needed, a further pilot study was conducted on five cases randomly selected from each of eight orthodontic practice databases assigned to the current study. From the 40 cases, 422 teeth had IPR, from a total of 931 teeth present. Using new CC Pro, the mean planned and achieved

<table>
<thead>
<tr>
<th>Method</th>
<th>Geomagic 1</th>
<th>Geomagic 2</th>
<th>3D OP measure</th>
<th>Callipers</th>
<th>Old CC Pro</th>
<th>New CC Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geomagic 1</td>
<td>1.000</td>
<td>0.986</td>
<td>0.989</td>
<td>0.986</td>
<td>0.971</td>
<td>0.988</td>
</tr>
<tr>
<td>Geomagic 2</td>
<td>0.986</td>
<td>1.000</td>
<td>0.994</td>
<td>0.995</td>
<td>0.989</td>
<td>0.997</td>
</tr>
<tr>
<td>3D OP measure</td>
<td>0.989</td>
<td>0.994</td>
<td>1.000</td>
<td>0.997</td>
<td>0.988</td>
<td>0.996</td>
</tr>
<tr>
<td>Callipers</td>
<td>0.986</td>
<td>0.995</td>
<td>0.997</td>
<td>1.000</td>
<td>0.986</td>
<td>0.997</td>
</tr>
<tr>
<td>Old CC Pro</td>
<td>0.971</td>
<td>0.989</td>
<td>0.988</td>
<td>0.986</td>
<td>1.000</td>
<td>0.991</td>
</tr>
<tr>
<td>New CC Pro</td>
<td>0.988</td>
<td>0.997</td>
<td>0.996</td>
<td>0.997</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Geomagic 1 and 2 denote the two investigator’s independent data. ‘Old CC Pro’ is ClinCheck version 5.1.1.21 NA. ‘New CC Pro’ is ClinCheck version 5.7.3.93 INT.*
Figure 1. Tooth measurements supplied for the same case showing variations between different versions of ClinCheck. Left is using Old CC Pro, right is New CC Pro.

Table II. Correlation matrix \(n = 120\) comparing both ClinCheck versions with ‘gold standard’ measurements.

<table>
<thead>
<tr>
<th></th>
<th>Old CC Pro</th>
<th>New CC Pro</th>
<th>3D OP</th>
<th>Callipers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old CC Pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>0.991**</td>
<td>0.988**</td>
<td>0.988**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>New CC Pro</td>
<td>0.991**</td>
<td>1</td>
<td>0.997**</td>
<td>0.998**</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>3d OP</td>
<td>0.988**</td>
<td>0.997**</td>
<td>1</td>
<td>0.999**</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>Callipers</td>
<td>0.988**</td>
<td>0.998**</td>
<td>0.999**</td>
<td>1</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).**

tooth width differences after IPR were measured, and a summary of these data is shown in Table III.

When the results of the pilot study were examined, it was noted that teeth not undergoing any IPR returned different measurements at \(T_1\) and \(T_2\). The descriptive statistics for this finding are illustrated in Table IV. Using G*Power (version 3.1.9.2) to assess these findings showed that a minimum sample size of 17 in each sub-group would be required.

For each orthodontic practice and for all 10 practices collectively, the tooth-group variables were:

- Anterior (incisors and canines).
- Posterior (premolars and first molars).
The main study comprised a total of 247 cases, with a total of 5,757 teeth, of which IPR was planned on 2,474 teeth. Missing teeth were omitted from analysis, even though the ClinCheck software assigned values to them. Second and third molars, when present, were excluded from the study. For all statistical analyses, teeth without planned IPR were eliminated from measurement.

For each case, two of the investigators (AS and BK) independently opened the $T_1$ and $T_2$ ClinCheck files with ClinCheckPro and entered the values from the Bolton analyses (Figure 3) into an Excel spreadsheet. In ClinCheck, IPR is assigned per contact and therefore, for this study, was assumed to be equally applied to adjacent teeth.

The data were then cross-checked, any errors corrected, and the Excel files exported for statistical analysis. Data were evaluated using SPSS Statistics for Windows, version 22 (IBM Corp., Armonk, N.Y., USA), with significance set to $p = 0.05$.

### Results

The data were checked for outliers attributable to the innate method error of the ClinCheck Pro software. Those found were not discarded as they were inferred to have negligible effect on the results, since the variations were normally distributed, and the sample size was very large.

### Overall

The principal finding of this study was that considerably less IPR was performed than planned, by every orthodontic practice. For all teeth assessed ($n = 2,474$) the mean planned IPR was $0.29 \text{ mm (SD = 0.13)}$, while mean achieved was only $0.13 \text{ mm (SD = 0.16)}$, with narrow 95% CIs (see Figure 4).

As shown in Table V, the mean shortfall was $0.16 \text{ mm per tooth (SD = 0.16)}$ with a $t$-test showing high statistical significance ($p < 0.001$).
Figure 5 presents the overall results for individual orthodontic practices and for all combined, showing a pattern of underperformed IPR by all participants. On average, the amount of IPR achieved represented only 44.0% of the total prescribed. The best result (Orthodontist I) was 75.4%, while the worst

Table IV. Teeth in pilot study of 40 cases for which no IPR was done, to show variation of New CC Pro measurements from T1 to T2.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planned difference</td>
<td>509</td>
<td>0.05</td>
<td>1.00</td>
<td>0.2900</td>
<td>0.14318</td>
</tr>
<tr>
<td>Actual difference</td>
<td>509</td>
<td>-0.78</td>
<td>1.18</td>
<td>0.1303</td>
<td>0.17169</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>509</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Showing pre- and post-IPR measurements from ClinCheck.

Figure 4. Summary chart showing mean planned and achieved tooth width changes for all teeth (n = 2,474) from all 10 orthodontists combined.
(Orthodontist E) managed only 9.9% of the planned total IPR. Using all combined data, as shown in Table VI, linear regression indicated that the average expected IPR could be calculated using a constant of 0.019, and an unstandardised coefficient of 0.493 ($p < 0.01$).

For example, if 0.5 mm of IPR was planned, the expected achieved value would, on average, be 0.27 mm, calculated as:

$$\text{Expected achieved value} = \text{Planned difference} \times \text{Unstandardised coefficient} + \text{Constant}$$

$$0.5 \times 0.493 + 0.19 = 0.266$$

**Sub-groups**

The amount and incidence of IPR varied significantly between the studied orthodontic practices, with some sub-groups having insufficient data for individual orthodontists. The statistical results are therefore presented in Table VII for all orthodontists combined.
Although the mean achieved IPR ranged in the groups from 37.1 to 58.8%, there was a surprising consistency, with most groups close to the overall mean of 44.0%. There were statistically significant differences between the Anterior and Posterior groups (p < 0.01) and between Mx Anterior and Mx Posterior groups (p < 0.01), however, these could not be considered clinically significant with mean shortfall differences of 0.05 and 0.07 mm, respectively. Data for each group, organised by orthodontic practice, are graphically presented in Figure 6. Although there were some differences between orthodontists and tooth groups the overall trend was a consistent finding of underperformance of the planned amount of IPR.

**Discussion**

The accuracy of in vivo IPR using digital study models was assessed for 10 orthodontic practitioners, all of whom had significant experience using the Invisalign appliance. The results indicated that IPR was not completed to a high degree of accuracy, and that the amount completed was generally much smaller than that intended. This was a consistent finding when completing IPR in the upper or lower arch, and in either the posterior or anterior arch segments.

The present study also determined the algorithms used to measure tooth widths had been changed in the newer software version (ClinCheck Pro v5.7.3.93 INT). While the improved accuracy of the data supplied by Align to doctors is to be lauded, it is of concern that the original data had been retrospectively changed and overwritten by Align without informing doctors that such alteration to diagnostic information, upon which they may have relied in the treatment of their patients, had occurred.

Overall IPR was deficient by 56.0% per tooth in the sample assessed. This matches almost exactly the findings of Blandy et al., in which a 55.9% deficiency in the performance of IPR by a single orthodontist was found. The recent study by De Felice et al. reported similar shortfalls, and also showed similar results with lower arch IPR performed at 37.02% and the upper arch at 44.95%. The overall findings of the current study are also consistent with those presented by Johner et al. in their in vitro IPR quantification study, which found...
Figure 6. Subgroups by orthodontist and all combined.
routine under-performance of IPR regardless of the actual method employed.

In the current study an average 0.29 mm of IPR was prescribed per tooth surface. However, a mean of only 0.13 mm was removed, which represented a statistically significant under-achievement, on average, of 0.16 mm per tooth surface. Not only was this statistically significant, but a deficiency of more than 50% between the proposed and achieved amounts of IPR could be deemed clinically significant, particularly if multiple IPR locations were prescribed to gain substantial amounts of space. When IPR is prescribed as a specific treatment modality with a specific amount required, it is important that it is completed as close as possible to the value set. As the amount of IPR achieved increasingly differed from the amount planned, the likelihood of reaching the initial treatment goal decreases. The ability of teeth to remain seated within an aligner, when their mesiodistal size is greater than the dimensions of the aligner is limited. IPR should be completed with a high degree of accuracy to avoid undesired side effects such as failure to adequately align the teeth.

When comparisons were made between IPR performed in upper and lower arches, anterior and posterior teeth, and the subgroupings of upper and lower anterior and posterior groups, a significant difference between the actual and proposed IPR amount per tooth surface was seen in all groups. Nevertheless, most groups were found to fall close to the overall mean of 44.0%. While statistically significant differences were observed between the overall anterior and posterior groups \( p < 0.01 \), and also between maxillary anterior and posterior groups \( p < 0.01 \), with mean shortfall differences of only 0.05 and 0.07 mm, respectively, these could not be considered clinically significant. Despite the consistency of orthodontic practices to complete the same amount of IPR per tooth surface, the results demonstrate a universal inability to reach the treatment goal of completing an accurate amount of IPR.

The finding of consistent, generalised under-achievement of IPR per tooth surface in the present study necessitates consideration of possible reasons. Since the periodontal ligament behaves thixotropically when teeth are placed under pressure, one explanation is that, as an IPR implement is passed through the contact point, pressure on the teeth moves them laterally into the periodontal ligament space and they only slowly re-approximate again. Furthermore, as the subjects in this study were undergoing active orthodontic treatment with the inherent inflammatory process and widened periodontal ligaments associated with tooth movement, a degree of tooth mobility could be expected. Therefore, when measuring the amount of IPR, it may appear to the clinician that the correct amount of reduction has been achieved. While it has been thought that using hand abrasive strips produces a greater amount of tooth displacement compared to motor driven or oscillating strips/discs, Johner et al.\(^{25}\) found that the stripping technique was not a significant predictor of the actual amount of enamel reduction in vitro.

Conclusions

- The present study demonstrated that the clinical performance of IPR in 10 orthodontic practices consistently fell short of the prescribed amounts. The effect of this under-performance on clinical outcomes remains to be quantified.
- Overall IPR was deficient by 56.0% per tooth surface, which may be deemed to be of clinical significance, especially when substantial amounts of IPR are prescribed per arch. IPR may not be completed with a high degree of accuracy and the amount completed is generally smaller than what was intended. This was a consistent finding in both arches, and between posterior and anterior regions of each arch.
- It is possible that underperformance of IPR, combined with other known deficiencies in treatment outcomes with aligners, may contribute to less optimal treatment outcomes or at least the need for longer and more complex refinement orders.
- Large variations were noted between orthodontic practices in regard to the completion of prescribed IPR, perhaps reflecting different methodologies.
- Clinicians should acknowledge these findings, and accordingly, use the demonstrated methodology to accurately critique their own performance of conducting IPR to ensure the highest standard of treatment outcome.

Corresponding author

Tony Weir
Honorary Senior Lecturer
Discipline of Orthodontics, School of Dentistry
The University of Queensland
Oral Health Centre, 288 Herston Road
Herston, QLD 4006
Email: anthony.weir@uq.edu.au

Conflict of Interest
The authors declare no conflict of interest.

Acknowledgements
This study was supported by a grant from the 2019 EMEA and APAC Align Research Award.

References