How good are we at estimating crowding and how does it affect our treatment decisions?

*The Specialist Orthodontic Practice, Epping, **School of Oral and Dental Science, University of Bristol, UK, ***Eastman Dental Institute, University of London

Correspondence to: Anthony J. Ireland, School of Oral and Dental Science, University of Bristol, Bristol, UK. E-mail: tony.ireland@bristol.ac.uk

SUMMARY
OBJECTIVES: To assess the estimation of crowding by orthodontists and their subsequent extraction choices.
MATERIALS AND METHODS: Sixty-two orthodontists were asked to quantify crowding on eight lower arch study models using their preferred method and also to indicate possible extraction choices. For each model, the intermolar widths, intercanine widths, and clinical scenarios were identical, but the true crowding varied from 0.1 to 7.5 mm across the eight models as to a lesser extent did the curve of Spee. Eleven orthodontists repeated the exercise after 9 months to assess reliability.
RESULTS: The preferred method of space estimation by all of the orthodontists was direct visualization. However, the estimates of crowding were very variable. For the most crowded lower model with 7.5 mm of crowding, the estimates ranged from 5 to 20 mm. Extraction choices were less variable than estimates of crowding and shifted from second to first premolars as crowding increased. Estimates of crowding and treatment decisions changed with time in 28 of 33 repeat measures. Estimates of crowding were unrelated to clinical experience.
LIMITATIONS: The principal limitation of this study is that it was a laboratory-based study and utilized just the lower arch model for estimation and treatment planning.
CONCLUSIONS: Extraction decisions and estimates of crowding tended to vary both initially and over time but were less varied in the case of the extraction decisions. Although this may have been a reflection of the limited treatment options, perhaps reassuringly, as the degree of crowding increased, so did the likelihood of prescribing extractions and the decisions generally shifted from second to first premolars. How orthodontists estimate crowding and make subsequent extraction choices is important and has potential medico-legal implications.

Introduction

In order to determine the degree of crowding during orthodontic diagnosis, it is necessary to measure two features, mesio-distal tooth widths and arch length. Although these would seem straightforward things to measure, this is not always the case as crowding can make the visualization of the contact points difficult and the arch length is dependent on three factors: arch form, arch symmetry, and the curve of Spee. Which arch form to choose is not straightforward and there are a large number available including the Bonwill-Hawley arch (Germane et al., 1992), Catenary curve (MacConaill and Scher, 1949; Wheeler, 1950), and trifocal ellipse (Brader, 1972). More recently, the Elan system has been introduced, which takes into account arch symmetry, another factor affecting arch length (Andreiko, 1994). The curve of Spee will also have an effect since its presence signifies crowding and therefore space will usually be required for arch levelling (Baldridge, 1969).

In an attempt to make the diagnosis and quantification of crowding more accurate, a number of space analyses have been developed over the years, including the Tweed-Merrifield total face analysis (Merrifield, 1994), the Howes analysis (Howes, 1954), and the more comprehensive Royal London Space Analysis (Kirschen et al., 2000). Although some experienced clinicians may question the value of this type of formalized numerical data collection, a comprehensive space analyses may be helpful for the less experienced clinician. A previous study has shown that most orthodontists in the UK are taught a formal space analysis during their specialist training, although relatively few then continue to use it once qualified and working as a specialist orthodontist (Wallis, 2011). Possible reasons for this might include the time required to complete the analysis and the complexity and also the assumption that with increasing experience, diagnostic decisions are made just as well by direct visualization. This might have medico-legal implications if a
there was ever a challenge as to whether a treatment should or should not involve extractions. With this in mind, it was decided to investigate whether the estimation of crowding by clinicians is both accurate and reliable and how this might be translated into diagnostic decision making, with particular reference to extractions.

**Materials and methods**

In order to test estimates of crowding, eight lower arch typodonts were fabricated using eight identical sets of custom made acrylic teeth. In each case, the intercanine and intermolar widths were identical, but each typodont was set up with varying degrees of crowding and curves of Spee. The mesio-distal widths of the teeth used in each typodont were measured using a digital Vernier calliper on two separate occasions before being set up in the typodonts. Once set up, alginate impressions were taken and immediately poured up in dental stone to create eight lower arch stone models with varying degrees of crowding (Figure 1). The arch length of each model was determined using digital 3D scanning (3Shape, Denmark). The degree of crowding in each typodont was then determined by subtracting the total mesio-distal tooth widths of each set of teeth from the arch lengths measured on the scanner. The depth of the curve of Spee was also determined in each case. The crowding for the eight lower models ranged from 0.1 to 7.5 mm and the curve of Spee from 0 to 2.5 mm (Table 1).

Sixty-two orthodontists were then asked to estimate the degree of crowding in millimetres in each of the eight lower stone models using their preferred method. They were also asked to choose whether this was an extraction or non-extraction case, or whether they considered this decision to be borderline. If they felt extractions were required, they were also asked to indicate which teeth they would extract.

For each lower model, they were told to assume it was from a patient with a class I incisor relationship on a class I skeletal base with normal soft tissues. The reason only a lower model was provided along with this class I patient scenario is that traditional teaching in the UK is that treatment planning should begin in the lower arch in such cases and the upper arch is then planned to fit around the treatment planned lower.

Within the 62 orthodontists, there were five levels of experience, namely postgraduate orthodontic students, dentists with a special interest, senior orthodontic specialist trainees, high street specialists, and hospital specialists in orthodontics. In addition to the models, a range of measuring devices (e.g. rulers, dividers, callipers, brass wire) were also provided so that each orthodontist could estimate the crowding according to their own preferences and usual practice. In order to test reproducibility, 11 of the orthodontists repeated the exercise using the same eight lower models approximately 9 months later.

**Statistical methods**

The data were analysed using Stata Version 12 (Stata Corp, College Station, Texas, USA) with a predetermined level of significance of $\alpha = 0.05$. Univariate summary statistics for the estimates of crowding were determined for each model. The effect of operator experience on estimations of crowding was assessed using the Kruskal–Wallis one-way analysis of variance (ANOVA), which was used in preference to a parametric ANOVA due to the ‘discrete’ nature of the estimates. Estimated crowding was also compared with the exact value using the sign rank test.

Only models 4–8 were considered to be extraction cases by the 62 orthodontists and so the data for these models were tested using the Jonckheere–Terpstra test for double ordered data, where the intention to extract is ordered as follows: non-extraction, borderline extraction, and extraction and where the models were ordered with increasing amounts of crowding.

To test reliability in the estimation of crowding and treatment choices, 11 orthodontists repeated the exercise 9 months later on the same eight lower models. Reliability of the crowding estimates was assessed using the
intra-class correlation coefficient (ICC) and Lin’s coefficient of concordance (Table 2; Lin, 2000). The intra-class coefficient used was ICC[2,1], which considers the raters to be a representative sample of a population of similar raters. The bias correction factor $C_b$ is a measure of accuracy, unity indicating perfect accuracy, and this is multiplied by the Pearson correlation coefficient to produce Lin’s correlation coefficient and is a measure of both accuracy and precision. A value of $\rho_c < 0.9$ is considered to be poor (McBride, 2005).

**Results**

Univariate summary statistics for the estimates of crowding are shown in Table 1 and further illustrated in Figure 2 for each model, along with the true crowding in parentheses.

The Kruskal–Wallis one-way ANOVA showed there to be no statistically significant effect of operator experience on the estimates of crowding. As might be expected, all of the orthodontists estimated in either whole or half millimetre measures. What was also surprising is that none of the orthodontists used any of the tools provided to do a space analysis, instead opting to estimate crowding using visualization.

As there was no significant effect of experience on the estimates of crowding, data were pooled over experience for subsequent calculations. With the exception of model 2, there was a significant difference between estimated crowding and the exact value. However, crowding was underestimated with model 1, but overestimated with models 3–8. Indeed, overestimation was seen in one case, model 8, by more than 12 mm.

The large range in the estimates of crowding is illustrated in Figure 2, which would appear to suggest that it is the presence of crowding and or the curve of Spee that makes its estimation difficult, i.e. as the degree of crowding or curve of Spee increased so did the potential range of its estimation.

The frequency of treatment choices (non-extraction, borderline, extraction) for models 4–8 is shown in Table 3. When considering each of the models in turn, models 1–3 were deemed to be non-extraction and models 6–8 were chosen as extraction cases by all 62 orthodontists. As might be expected, the probability for increasing extraction with increasing crowding was statistically significant ($P = 0.01$) and this effect is clearly shown in Figure 3, which is a plot of the logit transformed cumulative distribution function of the intent to extract for models 4–8. This transformation is used for its linearizing properties (Cox, 2001) and it can be seen that there is a distinct separation between the levels of crowding and a shift to intention to extract with increased crowding. Interestingly, in total, six clinicians considered the least crowded model (model 4 with 0.1 mm of true crowding) to be a borderline extraction case. By contrast, all 62 clinicians considered the most crowded model (model 8), with 7.5 mm of crowding, to be an extraction case. For models 6–8, where other teeth were considered for extraction, only one clinician considered the extraction of a lower incisor to relieve the crowding. In all other cases, as the degree of crowding increased so the extraction choice changed from being largely second premolars to predominantly first premolars.

---

**Table 1** Univariate summary data for the estimated degree of crowding (mm) by the 62 orthodontists for the eight models using their preferred methods and the true values for crowding and depth of curve of Spee. IQR, interquartile range; SD, standard deviation.

<table>
<thead>
<tr>
<th>Model</th>
<th>True crowding</th>
<th>Curve of Spee</th>
<th>Obs</th>
<th>Mean estimated crowding</th>
<th>SD</th>
<th>Median estimated crowding</th>
<th>IQR</th>
<th>Mode of estimated crowding</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.0</td>
<td>62</td>
<td>0.1</td>
<td>0.35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>1.5</td>
<td>62</td>
<td>0.4</td>
<td>0.71</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1.3</td>
<td>1.5</td>
<td>62</td>
<td>1.9</td>
<td>1.01</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>1.5</td>
<td>62</td>
<td>3.1</td>
<td>1.06</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>2.5</td>
<td>1.5</td>
<td>62</td>
<td>4.1</td>
<td>1.31</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>4.2</td>
<td>2.0</td>
<td>62</td>
<td>5.4</td>
<td>1.70</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>2.5</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>5.0</td>
<td>2.0</td>
<td>62</td>
<td>7.1</td>
<td>2.15</td>
<td>7</td>
<td>2</td>
<td>6/7</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>7.5</td>
<td>2.5</td>
<td>62</td>
<td>9.6</td>
<td>2.90</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>

**Table 2** Frequency table of treatment choices for each of the five models 4–8 in which the treatment decision with respect to extractions was considered to be at least borderline.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Model 4 (0.1 mm crowding)</th>
<th>Model 5 (2.5 mm crowding)</th>
<th>Model 6 (4.2 mm crowding)</th>
<th>Model 7 (5.0 mm crowding)</th>
<th>Model 8 (7.5 mm crowding)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-extraction</td>
<td>56</td>
<td>32</td>
<td>10</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Borderline</td>
<td>6</td>
<td>22</td>
<td>24</td>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>Extraction</td>
<td>—</td>
<td>8</td>
<td>28</td>
<td>51</td>
<td>62</td>
</tr>
</tbody>
</table>
The reliability of the crowding estimates assessed using the intra-class correlation coefficient (ICC) and Lin’s coefficient of concordance. The bias correction factor $C_b$ is a measure of accuracy (unity indicating perfect accuracy). Lin’s correlation coefficient is a measure of both accuracy and precision and a value of $\rho_c < 0.9$ is considered to be poor. CI, confidence interval.

<table>
<thead>
<tr>
<th>Model</th>
<th>ICC</th>
<th>95% CI</th>
<th>$\rho_c$</th>
<th>$C_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.931</td>
<td>0.893 to 0.956</td>
<td>0.931</td>
<td>0.994</td>
</tr>
<tr>
<td>1</td>
<td>-0.236</td>
<td>-0.759 to 0.413</td>
<td>-0.210</td>
<td>0.920</td>
</tr>
<tr>
<td>2</td>
<td>0.605</td>
<td>0.086 to 0.873</td>
<td>0.582</td>
<td>0.836</td>
</tr>
<tr>
<td>3</td>
<td>0.267</td>
<td>-0.234 to 0.708</td>
<td>0.249</td>
<td>0.718</td>
</tr>
<tr>
<td>4</td>
<td>0.575</td>
<td>0.045 to 0.861</td>
<td>0.552</td>
<td>0.825</td>
</tr>
<tr>
<td>5</td>
<td>0.549</td>
<td>-0.012 to 0.853</td>
<td>0.525</td>
<td>0.770</td>
</tr>
<tr>
<td>6</td>
<td>0.732</td>
<td>0.285 to 0.919</td>
<td>0.713</td>
<td>0.973</td>
</tr>
<tr>
<td>7</td>
<td>0.831</td>
<td>0.503 to 0.951</td>
<td>0.817</td>
<td>0.985</td>
</tr>
<tr>
<td>8</td>
<td>0.780</td>
<td>0.213 to 0.941</td>
<td>0.763</td>
<td>0.891</td>
</tr>
</tbody>
</table>

When looking at reliability, that of the entire data set was higher than for each typodont. It was with models 4–6 that the original 62 orthodontists recorded some borderline extraction cases and for this reason, the intra-operator estimation of crowding is plotted as pair plots in Figure 4. Of the 33 repeat estimations, only 5 were the same over the two time periods. The remainder ranged from 1 mm different to a maximum of 3 mm different with time. In the case of model 6, the extraction choices altered for 6 of the 11 orthodontists between T1 and T2 and for model 7, three extraction choices

changed. For model 6, the estimated crowding altered by as much as 2 mm between T1 and T2 and for model 7 by as much as 3 mm. The range of estimated crowding between orthodontists for the same models and the subsequent treatment choices was also interesting. For example, with model 6, one orthodontist estimated the crowding as 12 mm (T2), but favoured a non-extraction approach and yet on the same model, 10 orthodontists estimated the crowding as 4 mm (T1) and advocated extraction.

**Discussion**

In this study, the 62 orthodontists were asked to estimate the degree of crowding on eight lower arch models and all made their estimations using direct visualization. This was despite the fact that instruments such as callipers, rulers, dividers, and brass wire were provided, should they wish to undertake a more formalized space analysis. The results indicate that in most cases, the orthodontists were not good at accurately estimating the amount of crowding present using direct visualization, and when they did so it was in 0.5 mm increments. In the case of models 3–8, crowding was overestimated in each case and for model 8, with 7.5 mm of true crowding, there was a 15 mm range in the estimation of crowding from 5 to 20 mm. Some of this apparent overestimation may have been due to the curve of Spee, which ranged from 0 to 2.5 mm, but it could not be wholly accounted for by this. In order to level the curve of Spee, Braun et al. (1996) have suggested 1 mm of extra space is required to level a 2 mm curve of Spee and 2 mm for a 2–3 mm curve of Spee. Therefore, looking at all eight models, it would seem that as the degree of true crowding increased, so did the estimated range (Table 1 and Figure 2).

Within the group of 62 orthodontists who took part in this study, there were five levels of clinical experience and it might have been expected this would have an effect on estimations of crowding. However, there was no statistically observed effect of clinical experience on the accuracy of the estimation of crowding when compared with the true degree of crowding present.

Since the results of this study appear to show that orthodontists are not always good at assessing the degree of crowding, it was felt important to explore how their estimates might translate into impact on treatment, and specifically extraction choices. To date, there has been relatively little published work on the degree of crowding and subsequent extraction decisions. A comprehensive analysis by Baumrind et al. (1996) focussed on the multifactorial nature of the decision-making process and although crowding was cited as the main consideration in 49 per cent of cases under investigation, the extent of crowding was not mentioned. In the current study, any other factors that might affect extraction choices were minimized by providing only a lower model and by specifying it was a class I case with a normal skeletal and soft tissue pattern. Interestingly, the
results suggest that although the estimation of crowding is often variable, it may still lead to the same treatment/extraction decision, since orthodontists who tended to overestimate crowding did not show a greater tendency to prescribe extractions. For example, in the case of model 8 with 7.5 mm of true crowding, those orthodontists who visually estimated less than 6 mm of crowding still made the same decision to extract two premolars as those orthodontists estimating over 15 mm of crowding on the same model. This may be due to the relatively limited number of options available to the orthodontist to relieve crowding, namely expansion, extraction(s), inter-proximal reduction, or distal movement of buccal segment teeth. It begs the question as to whether we really need to know exactly how much crowding exists if the treatment options are relatively limited? Nevertheless, the results do suggest that clinicians relate the amount of crowding to treatment decisions, e.g. non-extraction versus extraction and second versus first premolar extractions. Therefore, as the amount of true crowding increases so does the tendency for a non-extraction approach to be replaced with an extraction approach and the extraction of second premolars to change to the extraction of first premolars. It would be interesting to speculate whether any of the decisions might have been different if the orthodontists had been provided with the figures for the true amount of crowding for each model. For example, eight clinicians decided that extractions were indicated in the case of model 5, having estimated the crowding to be between 5 and 8 mm. If they had been informed that the amount of true crowding was in fact only 2.5 mm, would they have made the same choices? Similarly with model 6, of the 28 clinicians who made a firm extraction decision, 24 overestimated the crowding. Had they known that the true crowding was 4.2 mm, would it have affected their extraction decision? Whether knowledge of the precise degree of crowding via the routine use of measurement tools and analyses would alter treatment choices is the subject of further ongoing investigations.

The intra-operator assessment of crowding and extraction choices at T1 and 9 months later at T2 also highlighted some interesting findings. Reassuringly, with the exception of one orthodontist using model 5, all of the orthodontists agreed that no extractions were indicated in the case of models 1–5 at both T1 and T2. With model 6 (4.2 mm of crowding), only 3 of the 11 clinicians made the same extraction decision on the repeat questionnaire at T2; 4 changed from a borderline decision to a definite extraction decision; 2 changed from extractions to borderline; 1 changed from non-extraction to borderline and interestingly 1 clinician changed their opinion from borderline to non-extraction, this was despite their estimating 12 mm of crowding at T2. This particular clinician was the only person to consider model 7 to be a borderline rather than an extraction case, despite their estimation of 15 mm of
crowding. The results indicate that for some clinicians, this particular aspect of diagnosis is generally repeatable, but not necessarily valid. All 11 clinicians recorded that extractions were indicated for model 8 on both occasions, although estimates of crowding ranged on the second occasion from 7 to 18 mm. It would appear that crowding in excess of 7 or 8 mm is usually considered sufficient to warrant extractions, even among those clinicians who had a lesser tendency to extract. All clinicians showed a general trend towards extracting first premolars in the more crowded cases. For example, the majority of the extractions planned for model 6 (4.2 mm of crowding) were second premolars, yet virtually all extractions for model 8 were first premolars (7.5 mm of crowding). As the crowding increased, there was a marked shift of extraction pattern from first to second premolars.

Although this study was designed to eliminate the wide range of complexities that may affect the assessment of crowding, the observed variation among the 62 orthodontists was surprising. According to a recent review of the psychological aspects of orthodontics (Hicks and Kluemper, 2011), difficult clinical decisions are often unwittingly flawed by a range of biases, which might include personal treatment philosophies or experience of previous litigation. Certainly, in this present study, several clinicians were known to be users of appliances that carry a ‘non-extraction philosophy’ and this and other biases may have influenced at least their treatment decisions, if not their estimations of crowding. Even though the variability in the estimates of crowding was reassuringly tempered by a lesser variability in the choice whether or not to extract, if there were litigation as to whether or not extractions should be used as part of a course of orthodontic treatment, on the evidence of this study, it is not obvious that such decisions are made purely on the visual estimate of crowding. We are currently undertaking further research to see if knowledge of the true rather than estimated crowding would affect extraction decisions.

Conclusions

The following conclusions can be made as a result of this investigation:

- All 62 clinicians utilized a visual method when estimating crowding.
- The estimates of crowding varied considerably between clinicians and were not related to the level of training/clinical experience.
- Although some clinicians were poor at estimating the true amount of crowding, the extraction decisions were less varied than the estimates of crowding.
- As crowding increased, so did the likelihood of prescribing extractions and it generally shifted from second to first premolars.
- Estimates of crowding may vary by up to 3 mm for the same orthodontist over time.
- In some cases, extraction choices also altered over time.

Acknowledgements

The authors would like to thank Mr Rob Edge for setting up the typodonts and creating the stone models.

References

Baldridge D W 1969 Leveling the curve of Spee: its effect on mandibular arch length. The Journal of Practical Orthodontics 3: 26–41

Hicks E P, Kluemper G T 2011 Heuristic reasoning and cognitive biases: are they hindrances to judgments and decision making in orthodontics? American Journal of Orthodontics and Dentofacial Orthopedics 139: 297–304
MacConaill M A, Scher E A 1949 The ideal form of the human dental arcade, with some prosthetic application. The Dental Record 69: 285–302