Many patients undergoing orthognathic surgery experience some degree of neurosensory impairment as a normal consequence of surgery. This may also occur as a complication after mandibular third molar surgery. The condition is usually reversible, but it may also be permanent. Neurosensory tests are essential in making decisions regarding the nature of the nerve injury, the potential for recovery, and the possible need for secondary microneurosurgical intervention. The aim of this study was to evaluate the repeatability of 5 clinical neurosensory tests assessing the neurosensory function of the inferior alveolar nerve (IAN). Twenty healthy subjects (9 males, 11 females) ranging in age from 21 years to 27 years participated in this study. The methods of assessment were light touch (LT), 2-point discrimination (2-P), pin tactile discrimination (PIN), thermal discrimination, and sensibility (ST) testing of the mandibular teeth with a vitality scanner. All the measurements were evaluated by 2 examiners and repeated at an interval of 6 months. In the statistical analysis, intraexaminer and interexaminer variability, as well as the variability between the 2 observations at a 6-month interval, were calculated. All the subjects reacted positively to LT and thermal discrimination in every observation. For the 2-P and PIN tests, intraexaminer variability was smaller than interexaminer variability, and repeatability of the 2 observations at 6 months was also good. Repeatability of ST was best at incisors, but it was not numerically very accurate at any site. However, all teeth that reacted positively to the vitality scanner in the first observation also reacted positively in the second observation; this instrument thus provides a more meaningful measurement of neurosensory function as a positive/negative assessment than as a numeric value. (Int J Adult Orthod Orthognath Surg 2001; 16:36-46)
Because of the lack of standardization of both terminology and observation procedures, any review of the literature regarding neurologic alteration in elective surgical procedures is inevitably somewhat confusing. A large variety of objective and subjective tests have been used to examine the neurosensory deficits of the mandible. However, it would be merely academic to identify neurosensory deficits in patients who otherwise exhibit normal sensation and function. Hence there is also a practical need to improve the methods of neurosensory tests that are easy to use and readily available in clinical practice. Taken together, the results of these tests provide important information leading to a diagnosis of nerve injury.

In previous studies, we suggested that sensibility testing (ST) of the mandibular teeth with a vitality scanner may be one practical method to evaluate postoperative sensory loss in the mandible after bilateral sagittal split osteotomy (BSSO) and to predict the recovery of the neurosensory disturbance after BSSO. In the present study, the repeatability of 5 readily available neurosensory tests (light touch, 2-point discrimination, pin tactile discrimination, thermal discrimination, and ST) designed to rapidly assess sensation was evaluated. The aim of this study was to perform assessments of intraexaminer, interexaminer, and interinterval variability for all 5 tests.

Method and materials

The subjects recruited for this study were 20 healthy dental students, 9 males and 11 females. Their ages ranged from 21 years to 27 years, with a mean of 23 years. The subjects had no craniofacial or congenital anomalies or previous trauma to the orofacial region. Subjectively, all subjects had completely normal sensibility of the lower lip and mental regions.

All subjects were tested by 2 examiners at an interval of 6 months. On both occasions, the neurosensory tests were performed by both examiners. One of the examiners was very experienced and the other was less experienced with these tests. The subjects were tested in a randomized order. During the testing, subjects were asked to close their eyes and separate their lips comfortably so that they could concentrate on the perception. The lower lip and the mental area were divided into 4 zones. The testing was performed over a 1-cm area above and beneath the labiomental fold on both sides of the chin, and each zone was measured separately by a battery of neurosensory tests: light touch (LT), 2-point discrimination (2-P), pin tactile discrimination (PIN), and thermal stimuli. Each of the 4 facial zones was stimulated 3 times, and the response was classified as correct if there were at least 2 appropriate answers to each test. Earlier studies have shown that when different sites of the face are measured, the sites on the chin show similar spatial sensitivities; therefore, the measurements of the 4 zones of the chin were recorded as separate measurements. Altogether, 160 measurements were gathered from the LT, 2-P, PIN, and thermal tests. In addition, ST of all mandibular teeth was performed using a vitality scanner (Model 2006 Vitality Scanner, Analytic Technology). Altogether, 1,088 ST measurements were performed, excluding extracted teeth and teeth that had been endodontically treated.

In LT, the lower lip and the mental area were gently touched with a cotton wisp to see whether the touch was perceptible. A positive or negative reply was the only option at each point. Two-point discrimination was measured with a sharp millimeter ruler. The subject indicated whether he or she felt 2 distinct points of contact. The test was conducted by beginning with the tips closed and by progressively opening them at 1-mm increments until the subject felt 2 distinct points of contact. This distance was then recorded. Care was taken to ensure that the tips touched the cutaneous surface at the same time. The intra- and interexaminer variation as well as the interinterval difference between measurements were estimated to be significant if they were more than 2 mm.

Pin tactile discrimination was measured with needles weighing between
0.3 g and 15 g. The needles pressed the skin of each of the 4 zones through a loop (in the framework, through which the needles could move freely) by their own weight. The lightest needle that the subject perceived as sharp was recorded. The intraexaminer, interexaminer, and interinterval differences between measurements were estimated to be significant if they were greater than 1 g.

For the thermal test, 2 small glass tubes containing water at 50°C and 15°C were used. The subject's perception of each stimulus (ie, cold versus hot) was recorded.

The sensibility of the mandibular teeth was measured with a vitality scanner with a scale of 0 to 80. Third molars were excluded from the study, because most of the subjects did not have them. The first measurement of each tooth was recorded. For the analysis, the teeth were divided into 4 different groups: incisors, canines, premolars, and molars, and repeatability was evaluated in these different groups. The intra- and interexaminer variation as well as the interinterval difference between measurements were deemed significant if they were greater than 10.

Statistical analysis

First, the measurements were recorded and the variability within each test was calculated. With LT and thermal stimuli, a positive response was deemed normal. With 2-P, PIN, and ST, the most common numeric values were calculated. For ST, the teeth were divided into 4 groups: molars, premolars, canines, and incisors, because it is already known that sensibility varies in these different groups of teeth when measured with a vitality scanner.15

Second, the results of each examiner were compared. The measurements from the first examiner were compared to the measurements from the second examiner, and the measurements from the first observation were compared to the measurements from the second observation. The degree of agreement between measurements was illustrated graphically.

For the ST test, analysis of variance (ANOVA) was used to explore the possible components of variability (subject, tooth, examiner, order of examiners, or observations) as well as the second-order interactions between components. The degree of agreement in the sensibility values of the teeth between the 2 examiners and between the 2 observations was illustrated graphically in the 4 groups: incisors, canines, premolars, and molars.

Results

Because all the subjects were healthy, no neurologic deficiencies were noted. All the subjects responded normally to LT and thermal discrimination at every observation. For these tests, there were thus no differences between the 2 examiners or between the 2 observations at an interval of 6 months.

Analysis of all 2-P data showed the most common discrimination value to be 2 mm. The minimum value was 1 mm and the maximum was 12 mm. Of the values, only 36% were 2 mm or less, 61% were 3 mm or less, and 95% were 6 mm or less. The range of the values obtained is shown in Fig 1.

With PIN, the most common value found was 0.5 g. The range of values obtained was between 0.5 g and 15 g. However, 81% of the values were 1 g or less, 88% were 1.5 g or less, and 95% were 2.5 g or less. The range of the values obtained is shown in Fig 2.

In ST of the mandibular teeth, the most common values (75% of the values) measured with the vitality scanner were between 15 and 35 for incisors, between 30 and 40 for canines, between 30 and 45 for premolars, and between 30 and 60 for molars. The values obtained for each group of teeth are shown in Fig 3.

In the 2-P test, intraexaminer repeatability seemed to be better than interexaminer repeatability. Experience with the tests also seemed to play a role in repeatability: when measurements were compared between the first and second observations of each examiner, the difference between values was within 2 mm in 88% of the measurements by the more
Fig 1  Range of “normal” 2-point discrimination test values measured over the 4 facial zones of the mental region (n = 160).

Fig 2  Range of “normal” pin tactile discrimination test values measured over the 4 facial zones of the mental region (n = 160).

Fig 3  Distribution of “normal” sensibility testing values in different groups of mandibular teeth. Boxes show mean and upper and lower quartiles. The ends of the whiskers represent the observed value that is 1.5 times the interquartile range away from the nearest quartile.
experienced examiner, but in only 78% of those made by the less experienced examiner. When the measurements of the first examiner were compared to the measurements of the second examiner, 83% of the values were within 2 mm. The maximum difference between the 2 examiners was 7 mm. The degree of agreement between the measurements of the 2 examiners is shown in Fig 4. When the measurements from the first observation were compared to the measurements from the second observation, 83% of the measurements were within 2 mm. The maximum difference between the 2 observations was 10 mm. The degree of agreement between the 2 observations at an interval of 6 months is shown in Fig 5.

For the PIN test, when the limit for a statistically significant difference was 1 g, intraexaminer repeatability was 88% by the first examiner and 91% by the second examiner. Experience of the examiner seemed, therefore, to have no effect on the repeatability of the test. When the measurements of the first examiner were compared to the measurements of the second examiner, 89% of the values were within 1 g of each other. The maximum difference between the 2 examiners was 14 g. The amount of agreement between the measurements of
When the measurements of the first and second observations were compared, the difference was within 1 g in 89% of the measurements. The maximum difference between the 2 observations was 13.5 g. The degree of agreement between the 2 observations at an interval of 6 months is shown in Fig 7.

All the teeth that reacted positively to the vitality scanner in the first testing session also reacted positively in the second testing session. In this test, intraexaminer and interexaminer repeatability was not good. Repeatability between the first and second observations was also poor. In the ST test, 3 different ANOVA models were used to evaluate whether there were second-order interactions between the dependent variables when sensibility of the tooth was an independent variable. The dependent variables were the subject, the tooth, and the examiner. The value of the ST measurement was significantly dependent on the subject (P = 0.0001), the tooth (P = 0.0001), and the examiner (P = 0.002), but not on the order of examiners or the order of observations when the limits of variation between the 2 observations were set at -10 to +10 when comparing the current result to the previous one. The percentages of

Fig 6  Degree of agreement in the pin tactile discrimination measurements between the 2 examiners (n = 160).

Fig 7  Degree of agreement in the pin tactile discrimination measurements between the 2 observations at an interval of 6 months (n = 160).
Table 1  Percentages of instances in which sensibility retest value in each group of teeth was within ± 10 versus the previous value

<table>
<thead>
<tr>
<th>Teeth</th>
<th>Examiner 1</th>
<th>Examiner 2</th>
<th>Examiner 1</th>
<th>Examiner 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incisors</td>
<td>71</td>
<td>74</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Canines</td>
<td>70</td>
<td>68</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>Premolars</td>
<td>55</td>
<td>72</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td>Molars</td>
<td>39</td>
<td>56</td>
<td>46</td>
<td>48</td>
</tr>
</tbody>
</table>

Fig 8  Degree of agreement in the sensibility values of the mandibular teeth between the 2 examiners (A = incisors, B = canines, C = premolars, D = molars).
instances in which the values were within these limits in every group of teeth is presented in Table 1. The degree of agreement in the measurements in the different groups or teeth between the 2 examiners is illustrated in Fig 8, and the agreement between the 2 observations is illustrated in Fig 9.

Discussion

In this study, 5 different methods were used to assess neurosensory function of the IAN. Four of these (LT, 2-P, PIN, thermal) have commonly been used for similar purposes in maxillofacial surgery and other medical fields, but as far as we know, only a few studies have evaluated the repeatability of these tests in healthy subjects.

This study assessed only healthy subjects, and the results were extrapolated to correlate with the results of patients with disturbances of the IAN. This may have resulted in some selection bias. However, as far as possible, the measurement procedure was always accomplished in the same manner. Care must also be taken not to overstimulate or fatigue subjects with multiple and complicated stimulus-response tests.

The fact that most neurosensory measurements are based on subjective feelings of sensation during an objective stimulus procedure explains the wide range of results obtained among healthy study
subjects. There is also usually marked biologic variability between study subjects. It is important to use methods that avoid these problems. In this study, both examiners examined the same subjects on both occasions, and this is why biologic variability had no effect on intra- or interexaminer repeatability. Furthermore, all the subjects were dental students, and their cooperation was very good.

One problem in measuring only healthy subjects is the precision of the results. In this study, both examiners examined 20 subjects but made several measurements per subject in both observations. It was therefore possible to get more measurements and more reliable results.

In testing with static LT, many investigators have used Von Frey hairs as the stimulus. The perceptions of monofilaments 1.65 and 2.36 are considered normal for trigeminal distribution. We used a cotton wisp, because it is available in every practice. However, we had only healthy subjects with normal sensation in the lower lip and mental areas, and this method did not distinguish possible differences in sensation. All the subjects responded positively to this test in every observation. Thus, although the test is rough, its repeatability is perfect.

One problem pertaining to the 2-P test is the requirement to always position the probe with the same amount of pressure on the skin. However, it is important to be aware of the immense individual variation in thresholds; in the present sample, the values between individuals ranged from 1 mm to 15 mm. Ghali and Epker reported that normal values for 2-P discrimination varied between 5 and 15 mm. Campbell et al reported that the normal measures for 2-P discrimination in trigeminal distribution varied from 7 to 14 mm, and discrimination is considered diminished at 15 to 20 mm and absent above 20 mm. Kawamura and Wessberg have shown that 2-P discrimination appears to be most sensitive near the midline of the face, with an average 2-P discrimination of 5 mm, which increases to 9 mm as one progresses laterally. Another study found a mean value of 3.9 ± 0.8 mm (SD) for 2-P discrimination. In our study, 95% of the subjects responded to a 2-P perception threshold under 6 mm, and the mean value was 3.3 ± 1.8 mm (SD). However, the variation in values was large, ranging from 1 mm to 12 mm. Because an assumption of normal distribution is unrealistic and distribution of 2-P discrimination is skewed to the right, the median more accurately reflects where the bulk of the numbers lie than does the mean. In our study, the median was 3 mm. Our results also pointed out that measurement experience plays a role in this test. The difference in measurements between the experienced and inexperienced examiners was obvious. Similar results have been reported by others; Feldman et al reported the 2-P discrimination test to be useful in evaluating trigeminal nerve-injured patients, but the training of examiners and a standard calibration protocol are particularly important when using the 2-P perception threshold.

A variety of methods have been used in clinical neurosensory observations. It is well known that bedside neurosensory observations are often crude; the testing stimuli may be poorly controlled, and many human variables may be introduced by both the examiner and the patient. It is very unlikely that different methods will agree exactly, giving an identical result for all individuals. How far apart measurements can be without causing difficulties is a question of judgment. With subjective measurements, the true value is not usually known, and the limits of agreement are only estimates that apply to the whole population.

Very little data are currently available regarding pinprick testing in subjects with normal sensation. Most studies compare postoperative thresholds individually to preoperative values. The difference between the pre- and postoperative values is then calculated and analyzed, but to our knowledge, no “normal” value for pin tactile discrimination has been given. However, some authors have suggested that 15 g is an adequate force to elicit this response. However, one of our present subjects did not respond to the pin tactile test until 15 g, although he had
subjectively normal sensitivity in the area. This shows that individual variation can be quite large.

Temperature has classically been tested with Minnesota thermal disks. They are generally made of 4 materials: polyvinyl chloride, glass, copper, and steel. The patient is required to identify the time interval at which the colder disk (i.e., the glass disk) is applied. Nishioka et al. reported that a nerve is considered to have a neurosensory temperature deficit if there are fewer than 80% correct responses at one or both of the test sites. In our study, the discrimination between cold and warm temperatures was measured with 2 glass tubes that contained cold and warm water. Again, because all the subjects were healthy and had normal sensation in the chin, they responded positively to the test in every observation. However, this indicates that this test is also perfectly repeatable.

In the sensibility testing, intraexaminer and interexaminer repeatability was not good. The vitality scanner in this study was an alternating-current instrument, and this might be why the values were not repeatable. If repeatable values are to be obtained, the instrument should be operated on a direct current.

References


