CHAPTER IV
RESULTS

1. General Characteristic of the Study Subjects

The mean age of cleft subjects in this study was 12.97 ± 4.46, ranging from 5 to 29 years of age. Most subjects were in the age-group 5 to 11 years (41.90%). The total number of subjects was 105, consisting of 46 males (43.81%) and 59 females (56.19%). The numbers of subjects in different age ranges classified by gender are shown in Table 2.

Table 2 Age range classified by gender

Age range (Years)	Gender		Total n (%)
	Male n (%)	Female n (%)	
5 - 11	18 (17.14)	26 (24.76)	44 (41.90)
12 - 14	15 (14.29)	16 (15.24)	31 (29.53)
15 - 18	9 (8.57)	12 (11.43)	21 (20.00)
19 - 29	4 (3.81)	5 (4.76)	9 (8.57)
Total	46 (43.81)	59 (56.19)	105 (100.00)
Mean (SD)	12.88 (4.46)	13.09 (4.51)	12.97 (4.46)

Regarding cleft type, the cleft subjects were classified into 4 categories. These were unilateral cleft lip and palate (UCLP), bilateral cleft lip and palate (BCLP), cleft lip with or without cleft alveolus (CL/A), and cleft palate (CP) only. Table 3 shows cleft type classified by gender and age. Subjects with UCLP were the majority, consisting of 61 subjects (58.10%) from 24 males (22.86%) and 37 females (35.24%) with a mean age of 12.33 ± 4.11. The bilateral cleft lip and palate was the second largest group with 23 subjects (21.90%) with the mean age of 13.13 ± 4.63. Larger numbers of males (13.34%) than females (9.52%) were observed in this group. The number of the cleft lip with or without cleft alveolus and cleft palate only samples were 19 (18.10%) and 2 (1.90%), respectively. The oldest group of age was found in the cleft palate only group with the mean age of 17.00 ± 1.41.

Table 3 Cleft type classified by gender and age

Cleft type	Gender			Age		
	Male n (%)	Female n (%)	Total n (%)	Mean	SD	SE
UCLP	24 (22.86)	37 (35.24)	61 (58.10)	12.33	4.11	0.53
BCLP	14 (13.33)	9 (8.57)	23 (21.90)	13.13	4.63	0.96
CL/A	7 (6.67)	12 (11.43)	19 (18.10)	14.42	5.19	1.19
CP	1 (0.95)	1 (0.95)	2 (1.90)	17.00	1.41	1.00
Total	46 (43.81)	59 (56.19)	105 (100.00)	12.97	4.46	0.44

The actual treatment that the study subjects received or were receiving was categorized into orthodontic treatment alone and orthodontic treatment combined with orthognathic surgery. From 105 subjects, there were 58 subjects (55.24%) with 24 males (41.38%) and 34 females (58.62%) who received only orthodontic treatment alone. For the rest, 47 subjects (44.76%) with 22 males (46.81%) and 25 females (53.19%) were required additional orthognathic surgery. Treatment categories classified by gender are described in Table 4.

Table 4 Actual treatment categories classified by gender

Actual treatment categories	Gender		Total N=105
	Male n (%)	Female n (%)	n (%)
1. Orthodontic treatment alone	24 (41.38)	34 (58.62)	58 (55.24)
2. Additional orthognathic surgery	22 (46.81)	25 (53.19)	47 (44.76)
Total	46 (43.81)	59 (56.19)	105 (100.00)

Table 5 shows the received treatment plan classified by gender and cleft type. Considering the proportion between additional orthognathic surgery and orthodontic treatment only, in males, the BCLP group seemed to have the highest tendency for requiring of additional orthognathic surgery while the UCLP group was likely to receive only orthodontic treatment alone. On the other hand, in females, the tendency for receiving of additional orthognathic surgery and orthodontic treatment alone seemed to be equivalent in both the UCLP and BCLP groups. The CL/A and CP groups were tended to require only orthodontic treatment alone in both sexes.

Table 5 Actual treatment categories classified by gender and cleft type

Cleft type	Male n (%)			Female n (%)		
	Additional orthognathic surgery	Orthodontic treatment alone	Total	Additional orthognathic surgery	Orthodontic treatment alone	Total
UCLP	9 (8.57)	15 (14.29)	24 (22.86)	19 (18.10)	18 (17.14)	37 (35.24)
BCLP	13 (12.38)	1 (0.95)	14 (13.33)	4 (3.81)	5 (4.76)	9 (8.57)
CL/A	0 (0)	7 (6.67)	7 (6.67)	2 (1.91)	10 (9.52)	12 (11.43)
CP	0 (0)	1 (0.95)	1 (0.95)	0 (0)	1 (0.95)	1 (0.95)
			46 (43.81)			59 (56.19)

2.   Reliability Test for Cephalometric Measurements

Intra-observer reliability of the three cephalometric measurements was assessed using the intraclass correlation coefficient (ICC) and are presented in Table 6. Excellent reproducibility was found between two times of measurements in all three cephalometric values with ICC of 0.998 for ANB (degree) and U1-APog (mm) and 0.999 for L lip-Nperp. (mm).

Table 6 Reliability of the three cephalometric measurements

Cephalometric measurements	ICC
ANB (degree)	0.998
U1-APog (mm)	0.998
L lip-Nperp. (mm)	0.999

3. Evaluation of the Critical D score

After calculation of the D score for each subject, the critical D score for the FOSP was evaluated to optimally discriminate between those types of treatment need by using the receiver operating characteristic (ROC) curve (Figure 2). The area under curve was 0.796 (p-value < 0aa.001). The ROC curve was created by plotting true positive rate versus false positive rate for each cutoff. Sensitivity, specificity, accuracy, positive and negative likelihood ratios at different cutoff D scores are shown in Appendix D. The D scores that provided satisfactory sensitivity and specificity were selected and evaluated for the optimal cutoff point or critical D score. Sensitivity, specificity, accuracy, positive and negative likelihood ratios, and odds ratio of these selected points are presented in Table 7.
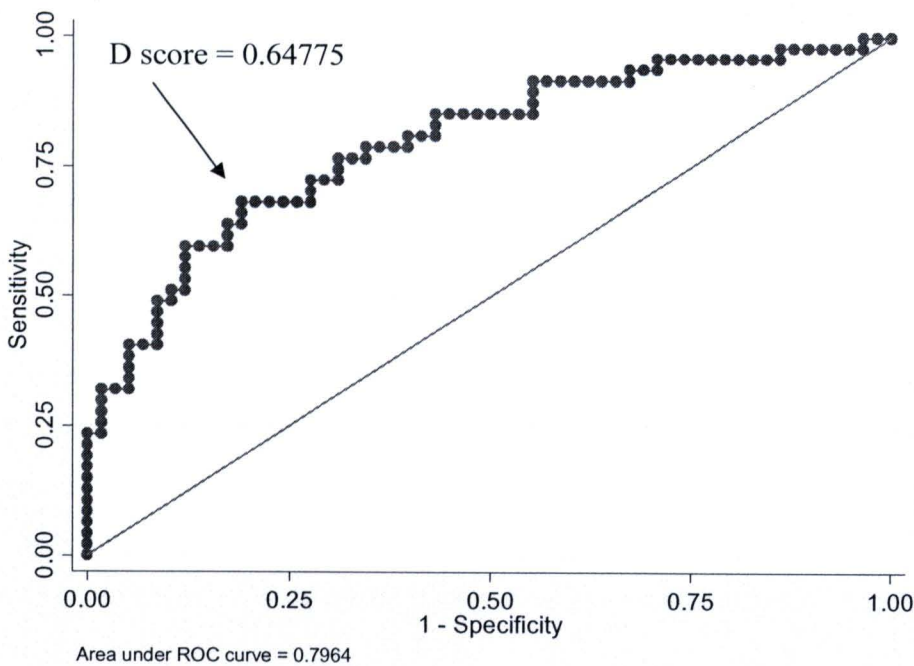

Area under ROC curve = 0.7964

Figure 2 The receiver operating characteristic (ROC) curve of the FOSP

Table 7 Sensitivity, specificity, accuracy, positive and negative likelihood ratios, and odds ratio at various selected cutoff D scores

Cutoff D Score	Sensitivity	Specificity	Accuracy	LR+	LR-	Odds Ratio
≥ -0.31750	78.72	65.52	71.43	2.283	0.325	7.03
≥ -0.27000	76.60	68.97	72.38	2.468	0.339	7.28
≥ -0.07825	72.34	72.41	72.38	2.622	0.382	6.86
≥ 0.64775	68.09	81.03	75.24	3.590	0.394	9.12
≥ 0.85800	63.83	82.76	74.29	3.702	0.437	8.47

Although the D score at ≥ -0.07825 might provide high percentage of both sensitivity and specificity (72.3% and 72.4%, respectively), the most appropriate cutoff D score was 0.64775. At this point, it provided the ROC plot closest to the upper left corner which also indicated a higher accuracy of the FOSP (75.2%).[102] It offers the best association between the sensitivity (68.1%) and specificity (81.0%) with the odds ratio of 9.12. Moreover, it minimized the rate of false positives (1 - specificity) while maximized the rate of true negatives (specificity) which was required for a good diagnostic test.

Therefore, application of the FOSP in this study was performed using the critical D score at 0.64775 as a cutoff point. If the D score calculated from the FOSP was equal or more than 0.64775, this subject would be classified in a group of needing additional orthognathic surgery. On the other hand, if the calculated D score was less than this point, an orthodontic treatment alone would be classified instead.

4.   **Reliability and Validity Test of the FOSP**

Using the FOSP with the critical D score at 0.64775, all subjects were classified into two groups of treatment prediction, the orthodontic treatment alone and orthodontic treatment combined with orthognathic surgery. Reliability and validity of the FOSP were presented in Table 8.

Table 8 Reliability and validity of the use of FOSP for comparing predicted and actually received type of treatment

Predicted treatment plan	Actual treatment plan		Total (%)
	Additional orthognathic surgery (%)	Orthodontic treatment alone (%)	
Additional orthognathic surgery	32 (30.48%)	11 (10.48%)	43 (40.95%)
Orthodontic treatment alone	15 (14.29%)	47 (44.76%)	62 (59.05%)
Total (%)	47 (44.76%)	58 (55.24%)	105 (100.00%)
Sensitivity for prediction of orthognathic surgery need (%) = 68.1% (95% CI: 52.9 - 80.9%)			
Specificity for prediction of orthognathic surgery need (%) = 81.0% (95% CI: 68.6 - 90.1%)			
Accuracy (%) = 75.24%			
Positive predictive value (PPV) = 74.4% (95%CI: 58.8 - 86.5%)			
Negative predictive value (NPV) = 75.8% (95%CI: 63.3 - 85.8%)			
Positive diagnostic likelihood ratio = 3.59 (95%CI: 2.04 - 6.33)			
Negative diagnostic likelihood ratio = 0.39 (95%CI: 0.26 - 0.61)			
Odds ratio = 9.12 (95%CI: 3.74 - 22.20)			
Kappa = 0.50 (95%CI: 0.33 - 0.66; p-value < 0.001)			

Reliability of the FOSP was determined through the agreement between the actual treatment and the predicted treatment plan interpreted from the FOSP of each subject. Kappa value was found at 0.50 with 95% of confidence intervals between 0.33 and 0.66.

When the actual treatments were used as references for assessment of validity of the FOSP, the percentage of correctly classified cases or proportion in agreement between actual treatment and predicted treatment plan was 75.2%. There were 11 patients of the nonsurgery group (n = 58) and 15 of the surgery group (n = 47) had been misclassified. Approximately 68% of sensitivity and 81% of specificity of the FOSP for prediction of the need for orthognathic surgery were observed. These findings imply that 68% of those who received or were receiving orthognathic surgery would be correctly classified with 52.9 - 80.9% of 95% CI. Similarly, a high

specificity of the FOSP indicates that 81.03% of subjects who had been treated with orthodontic treatment alone would be correctly classified with 68.6 - 90.1% of 95% CI. With the FOSP application, 79 subjects or approximately 75.2% of all study subjects were correctly identified according to their received actual treatments.

Considering the positive predictive value (PPV), among the 43 subjects who were classified from the FOSP as needing additional orthognathic surgery, 74.4% of these patients were correctly categorized. With 95% confidence, 58.8 - 86.5% of patients predicted as needing additional orthognathic surgery actually received this treatment plan. Similarly, the negative predictive value (NPV) implied that within 62 subjects of the orthodontic treatment plan alone predicted group, the percentage of a correct classification was 75.8%. With 95% confidence, 63.3 - 85.8% of patients predicted as orthodontic treatment alone group really received this type of treatment.

A positive diagnostic likelihood ratio implies that the candidates for additional orthognathic surgery would have a chance to be identified in this group with the probability of 3.6 times rather than to be classified in the group of orthodontic treatment alone. A negative diagnostic likelihood ratio of 0.39 also signifies that the negative test may be fair at ruling out a need for orthognathic surgery.[103]

Table 9 Means and standard deviations (SD) of ANB (degree), U1-APog (mm), L lip-Nperp. (mm) and the D score according to type and classification of treatment

Measurements	Actual treatment		Predicted from the FOSP	
	Orthodontic treatment alone (n = 58)	Additional orthognathic urgery (n = 47)	Orthodontic treatment alone (n = 62)	Additional orthognathic surgery (n = 43)
ANB	1.84 ± 3.74	-1.79 ± 4.77	2.74 ± 3.54	-3.42 ± 3.30
U1-APog	3.42 ± 4.11	-1.15 ± 4.39	3.66 ± 4.27	-1.92 ± 3.39
L lip-Nperp.	18.49 ± 5.40	19.89 ± 6.07	17.18 ± 5.02	21.91 ± 5.57
D score	-1.01 ± 2.01	1.66 ± 2.53	-1.58 ± 1.46	2.73 ± 1.61

The mean values of the three cephalometric values as shown in Table 9 indicate lesser positive values of the ANB and U1-APog (mm) with slightly more positive values of the L lip-Nperp. (mm) in the surgery group than those of the non-surgery group. The mean D score of the predicted nonsurgery group was more negative than that of the actual non-surgery group while in the predicted surgery group, the mean D score was higher than that of the actual surgery group.

Table 10 Means and standard deviations (SD) of ANB (degree), U1-APog (mm), L lip-Nperp. (mm) according to treatment classification from the FOSP

Measurements	Treatment classification from the FOSP			
	Correctly classified as orthodontic treatment alone group (n = 47)	Correctly classified as additional orthognathic surgery group (n = 32)	Incorrectly classified as additional orthognathic surgery group (n = 11)	Incorrectly classified as orthodontic treatment alone group (n = 15)
ANB	2.66 ± 3.52	-4.02 ± 3.38	-1.68 ± 2.43	2.98 ± 3.70
U1-APog	4.44 ± 3.67	-2.27 ± 3.53	-0.93 ± 2.88	1.23 ± 5.16
L lip-Nperp.	17.49 ± 5.24	21.63 ± 6.07	22.73 ± 3.92	16.20 ± 4.27

Means and standard deviations (SD) of the three cephalometric values according to the predicted treatment classification are presented in Table 10. The mean values of ANB (degree), U1-APog (mm), and L lip-Nperp. (mm) in the misclassified groups seem to be closer to those of the opposite groups of actual treatment need.

Table 11 shows the minimum and maximum values of ANB (degree), U1-APog (mm), and L lip-Nperp. (mm) in each predicted treatment classification. The highest values of the ANB and U1-APog and the lowest value of the L lip-N perp. were observed in the correctly classified as orthodontics alone group. In contrast, the largest distance of L lip-N perp. as well as the lowest values of the ANB were found in the correctly classified as orthognathic surgery group. The distance of U1-APog was found to be smallest in the incorrectly classified as orthodontics alone group.

Table 11 Ranges of ANB (degree), U1-APog (mm), L lip-Nperp. (mm) related to treatment classification from the FOSP

Measurements	Treatment classification from the FOSP							
	Correctly classified as orthodontic treatment alone group (n = 47)		Correctly classified as additional orthognathic surgery group (n = 32)		Incorrectly classified as additional orthognathic surgery group (n = 11)		Incorrectly classified as orthodontic treatment alone group (n = 15)	
	Min	Max	Min	Max	Min	Max	Min	Max
ANB	-3.75	11.00	-11.75	1.50	-5.25	2.25	-3.00	9.00
U1-APog	-2.75	18.00	-9.50	5.50	-5.75	4.00	-10.50	10.00
L lip-N perp.	5.00	30.50	8.00	33.50	17.25	27.25	6.75	22.00

Table 12 Means, standard deviations (SD), minimum and maximum values of the D scores according to treatment classification from the FOSP

Treatment classification	D score		
	Mean ± SD	Min	Max
Correctly classified as orthodontic treatment alone group (n = 47)	-1.70 ± 1.50	-5.77	0.56
Correctly classified as additional orthognathic surgery group (n = 32)	3.00 ± 1.71	0.65	7.43
Incorrectly classified as additional orthognathic surgery group (n = 11)	1.93 ± 0.92	0.69	3.67
Incorrectly classified as orthodontic treatment alone group (n = 15)	-1.20 ± 1.30	-4.48	-0.01

Table 12 shows the means, standard deviations (SD), minimum and maximum values of the D scores in each predicted treatment classification. Similar to those findings of the three cephalometric values in Table 10, the mean D scores calculated from the FOSP in the misclassified groups were closer to those of the opposite group of actual treatment need as well.

4.1 Influence of an Age on Reliability and Validity of the Predicted Treatment Plan from the FOSP

To observe effect of an age on reliability and validity of the FOSP application, all subjects were classified into four groups according to their chronological age; these were aged 5 - 11 years, 12 - 14 years, 15 - 18 years, and 18 - 29 years.

Reliability was determined from the kappa values, representing agreement between the actual treatment and the predicted treatment plan interpreted from the FOSP of each age group. Meanwhile, validity of the concluded treatment plan derived from the FOSP compared to the actual received treatment was also evaluated through sensitivity, specificity, accuracy, positive and negative predictive values, using the actual treatments as a reference.

Reliability and validity of the FOSP for prediction of the need for orthognathic surgery were evaluated among each group of age as are shown in Table E(1) - E(4) in Appendix E. Sensitivity, specificity, accuracy, and kappa values of these age groups are summarized in Table 13.

Table 13 Sensitivity, specificity, accuracy, and the kappa values of the FOSP for prediction of the need for orthognathic surgery according to the age, 95% CI were presented in parentheses

Age (yr)	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
5 - 11	57.9 (33.5 - 79.7)	80.0 (59.3 - 93.2)	70.5 (57.0 - 83.9)	68.8 (41.3 - 89.0)	71.4 (51.3 - 86.8)	0.39 (0.11 - 0.66)
12 - 14	86.7 (59.5 - 98.3)	81.3 (54.4 - 96.0)	83.9 (70.9 - 96.8)	81.3 (54.4 - 96.0)	86.7 (59.5 - 98.3)	0.68 (0.42 - 0.94)
15 - 18	60.0 (26.2 - 87.8)	72.7 (39.0 - 94.0)	66.7 (46.5 - 86.8)	66.7 (29.9 - 92.5)	66.7 (34.9 - 90.1)	0.33 (0.00 - 0.73)
19 - 29	66.7 (9.4 - 99.2)	100.0 (54.1 - 100.0)	88.9 (68.4 - 100.0)	100.0 (15.8 - 100.0)	85.7 (42.1 - 99.6)	0.73 (0.24 - 1.00)

* PPV = positive predictive value, NPV = negative predictive value

Reliability of the FOSP was found with substantial strength of agreement in the age groups of more than 18 years (k = 0.73, 95% CI: 0.24 - 1.00) and 12 - 14 years (k = 0.68, 95% CI: 0.42 - 0.94). Only a slight degree of agreement was observed in the age groups of 5 - 11 years (k = 0.39, 95% CI: 0.11 - 0.67) and 15 - 18 years (k = 0.33, 95% CI: 0.00 - 0.73).

The highest sensitivity of the FOSP for prediction of orthognathic surgery need was found in the group of 12 - 14 years with 86.7% (59.5 – 98.3% of 95% CI) while the highest specificity up to 100% (54.1 – 100% of 95% CI) and the highest accuracy of 88.9% (68.4 – 100.0% of 95% CI) was observed in the group of more than 18 years. Wide ranges of the 95% confidence intervals were observed in all age ranges.

4.2   Influence of Gender on Reliability and Validity of the Predicted Treatment Plan from the FOSP

Reliability and validity of the FOSP were evaluated, considering gender of the subjects as are shown in Table F(1) and F(2) in Appendix F. Sensitivity, specificity, accuracy, and kappa values of each gender are summarized in Table 14. A slightly better strength of reliability (k = 0.51, 95% CI: 0.29 - 0.73) and slightly higher percentage of specificity (82.4%, 95% CI: 65.5 – 93.2%) were observed in the females while the sensitivity was comparable in both sexes at approximately 68%.

Table 14 Sensitivity, specificity, accuracy, and the kappa values of the FOSP for prediction of additional orthognathic surgery need classified by gender, the 95% CI were presented in parentheses

Gender	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Male	68.2	79.2	73.9	75.0	73.1	0.48
	(45.1 - 86.1)	(57.8 - 92.9)	(61.2 - 86.6)	(50.9 - 91.3)	(52.2 - 88.4)	(0.22 - 0.73)
Female	68.0	82.4	76.3	73.9	77.8	0.51
	(46.5 - 85.1)	(65.5 - 93.2)	(65.4 - 87.1)	(51.6 - 89.8)	(60.8 - 89.9)	(0.29 - 0.73)

* PPV = positive predictive value, NPV = negative predictive value

4.3   Influence of Cleft Types on Reliability and Validity of the Predicted Treatment Plan from the FOSP

Type of cleft was also determined whether it had an influence on reliability and validity of the FOSP. Since the amount of subjects in the CL and CP groups were too small to calculate statistically, consequently, these two groups were discarded. Only the UCLP and BCLP categories were included for statistical analysis. Reliability and validity of the FOSP according to types of cleft are shown in (Table G(1) and G(2) in Appendix G). Sensitivity, specificity, accuracy, and kappa values of each cleft type are summarized in Table 15.

Table 15 Reliability and validity of the FOSP for prediction of the need for orthognathic surgery classified by cleft type with 95% confidence intervals presented in parentheses

Cleft type	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
UCLP	78.6	78.8	78.6	75.9	81.3	0.57
	(59.0 - 91.7)	(61.1 - 91.0)	(66.3 - 88.1)	(56.5 - 89.7)	(63.6 - 92.8)	(0.37 - 0.78)
BCLP	52.9	66.7	56.5	81.8	33.3	0.15
	(27.8 - 77.0)	(22.3 - 95.7)	(34.5 - 76.8)	(48.2 - 97.7)	(9.9 - 65.1)	(0.00 - 0.49)

* PPV = positive predictive value, NPV = negative predictive value

The UCLP had evidently higher sensitivity (78.6%, 95% CI: 59.0 - 91.7%) and specificity (78.8%, 95% CI: 61.1 - 91.0%) than those of the BCLP. Reliability of the FOSP for prediction of need for orthognathic surgery in the UCLP was also better than the BCLP with moderate strength of agreement (k = 0.57, 95% CI: 0.37 - 0.78) while only slight degree of agreement was found in the BCLP (k = 0.15, 95% CI: 0.00 - 0.49).