CHAPTER 4

Use of Receiver Operating Characteristic curves for selection of treatment thresholds using the FAMACHA© diagnostic system for anaemia in sheep naturally infected with *Haemonchus contortus*

4.1 Introduction

The purpose of this part of the study was to further compare the diagnostic accuracy of FAMACHA© classification on Farms 1 and 2 in South Africa using Receiver Operating Characteristic curve analysis. The principles and practical application of Receiver Operating Characteristic curve analysis for diagnostic tests have been reviewed by Greiner & Gardner (2000). Receiver Operating Characteristic plots provide an index of accuracy for a diagnostic test by demonstrating the limits of a test's ability to discriminate between states of health such as diseased/not diseased or infected/not infected, over the range of operating conditions for the test (Zweig & Campbell 1993).

Receiver Operating Characteristic curve analysis was used to compare the discriminating ability of the FAMACHA© diagnostic test on the above two farms by selecting two cut-off values for the haematocrit reference test, namely ≤22 % and ≤19 %, to indicate the true presence or absence of disease in four separate sets of analyses. The area under the Receiver Operating Characteristic curve (area under the curve) was calculated for the two haematocrit cut-off values and for data from both farms using STATA (Stata Statistical Software: Release 8.0. College Station, TX: StataCorp LP). Sheep on each farm were regarded as diseased if the haematocrit was ≤22 % or ≤19 %. The cut-off value of ≤22 % was chosen because it is the upper limit for FAMACHA© category 3, and this category is recommended as a treatment threshold during the application of the FAMACHA© system. The cut-off value of ≤19 % was included in the analysis to provide an additional view of the data, and to determine whether there was a difference between the two cut-off values in the diagnostic power of the FAMACHA© method. In Chapter 3, the statistical techniques of sensitivity and specificity, positive and negative likelihood ratios and prevalence of infection were used to define the accuracy of the FAMACHA© method on the two farms. In the present study a distinction is made between the term "cut-off" which is used to classify the true disease status of an animal according to a preselected haematocrit value into true diseased/non-diseased, and “cut point”, which refers to dichotomized FAMACHA© test results. The term “cut point” is also used in the STATA statistical software to designate the FAMACHA© categories as rating points (Hanley & McNeil 1982), used to calculate the area under the curve index values. For example, if all individuals in FAMACHA© categories 2–5 are considered to be test positive, then the cut point is 2, and if all individuals in FAMACHA© categories 3–5 are considered test positive, the cut point is 3, etc. The diagnostic sensitivity and specificity of a test are a function of the cut points of the test, and Receiver Operating Characteristic analysis assesses the diagnostic performance of the system in terms of sensitivity and 1 minus specificity for each possible cut point of the test. A further refinement of Receiver Operating Characteristic analysis, the two-graph Receiver Operating Characteristic curve method (Greiner, Sohr & Göbel 1995; Beck, Gašpar, Mihaljević, Marinculić, Stojčević & Brstilo 2005; Greiner, Pfeiffer & Smith 2000), was used to optimise the selection of FAMACHA© cut points for anthelmintic treatment. This method was used to plot the sensitivity and specificity as a function of the FAMACHA© cut point to maximise sensitivity under the assumptions that the disease occurs at a relatively high prevalence, as almost every animal is infected, but due to over-dispersion of worm burdens (Wilson *et al*. 1996; Herbert & Isham 2000) a few individuals harbour the majority of the parasites, and that non-detection of a truly diseased animal for treatment has a more serious consequence than the incorrect treatment of a non-diseased animal.

The aim of this work was to determine whether there would be differences in the discriminatory performance in the application of the FAMACHA© method for the two pre-determined haematocrit cut-off values and for FAMACHA© cut points of 2 and 3 for the two farms, and to select optimum FAMACHA© cut points for treatment by taking into account the relative consequences of false negative and false positive test results.

4.2 Materials and methods

4.2.1 Origin of data and FAMACHA© test procedures

The data analysed in this work consisted of anaemia status as evaluated by FAMACHA© scores and haematocrit values, collected from naturally infected sheep on the two farms in the summer rainfall area of South Africa. The origin of the data and FAMACHA© test procedures have been described in detail in Chapter 3.

4.2.2 Receiver Operating Characteristic analysis

The area under the Receiver Operating Characteristic curve, which is a summary statistic of the overall diagnostic accuracy and thus the discriminatory power of each diagnostic test (Greiner *et al*. 2000), was non-parametrically calculated for each haematocrit cut-off value and for data from each farm, using STATA. The same software was used to calculate sensitivity, specificity and likelihood ratios using the rating method (Hanley & McNiel 1982). In the present study the results of the haematocrit determination were used as the "gold standard" to validate the FAMACHA© method for two levels of infection on the two farms, namely either FAMACHA© categories 2–5 regarded as test positive, or FAMACHA© categories 3–5 regarded as test positive. The points required to produce the Receiver Operating Characteristic curve were obtained by successively considering increasingly broader categories of abnormal test results, for example by considering FAMACHA© category 5 alone as abnormal, then FAMACHA© category 5 plus 4, then FAMACHA© category 5 plus 4 plus 3, etc. Two-graph Receiver Operating Characteristic analysis was then used to select cut points for treatment for a pre-defined sensitivity of 0.90, i.e. the cut point was selected to ensure that a minimum of 90 % of animals defined as truly diseased by the FAMACHA© cut points would be detected and treated. In all analyses, the selected haematocrit cut-off values of ≤22 % and ≤19 % were used as the reference variable, while the proportional observed FAMACHA© scores were used as the classification variable in the STATA software.

4.3 Results

The performance of FAMACHA© as measured by the area under the curve for both farms is illustrated in Figs. 4.1a and 4.1b. The results of the rating method to obtain the points for the Receiver Operating Characteristic curve for Farm 1 are given in Tables 4.1 and 4.2, and for Farm 2 in Tables 4.4 and 4.5. The area under the curve index values and standard errors for Farm 1 are given in Table 4.3 and for Farm 2 in Table 4.6. The area under the curve was closest to 1 at 0.90 (0.877–0.920) on Farm 2 for a haematocrit cut-off of ≤19 % (Table 4.6 and Fig. 4.1b). The smallest area under the curve index value was obtained for a haematocrit cut-off of ≤22 % on Farm 1 (0.790), while the highest was obtained for Farm 2 for a haematocrit cut-off of ≤19 % (0.901). The results of the two-graph Receiver Operating Characteristic analysis for both farms are given in Figs. 4.2a–d. For farm 1, the FAMACHA© cut points for a sensitivity of 0.9 were FAMACHA© category 2 and category 3 for haematocrit cut-offs of ≤22 % and ≤19 %, respectively (Fig. 4.2a and b), and for Farm 2 the FAMACHA© cut points for a sensitivity of 0.9 were FAMACHA© 3 for both haematocrit cut-offs (Fig. 4.2c and d).

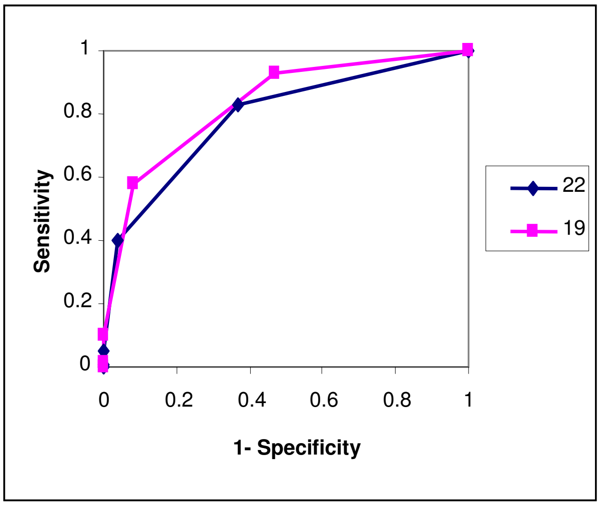

FIG. 4.1a Farm 1. Receiver Operating Characteristic curves. The area under the curve for a haematocrit cut-off of ≤22 % is 0.790, and for ≤19 % it is 0.835

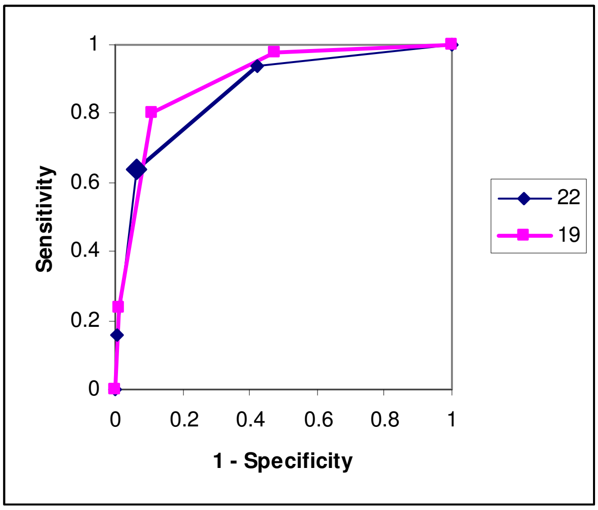

FIG. 4.1b Farm 2. Receiver Operating Characteristic curves. The area under the curve for a haematocrit cut-off of ≤22 % is 0.867, and for ≤19 % it is 0.901

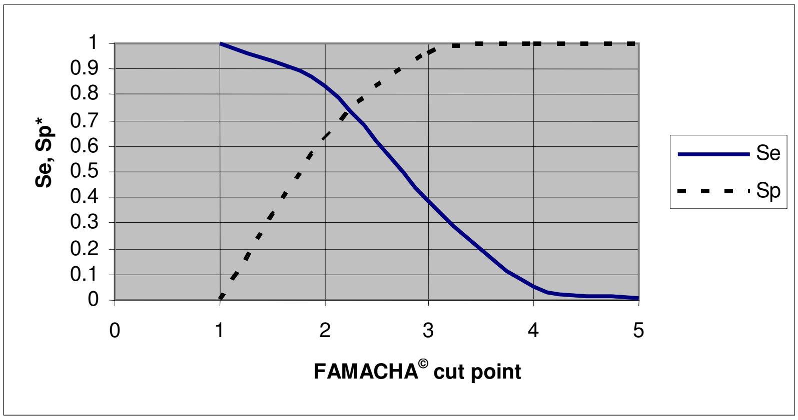

FIG. 4.2a Farm 1. Two-graph Receiver Operating Characteristic plot for a haematocrit cut-off of ≤22 %. *Se = sensitivity, Sp = specificity.

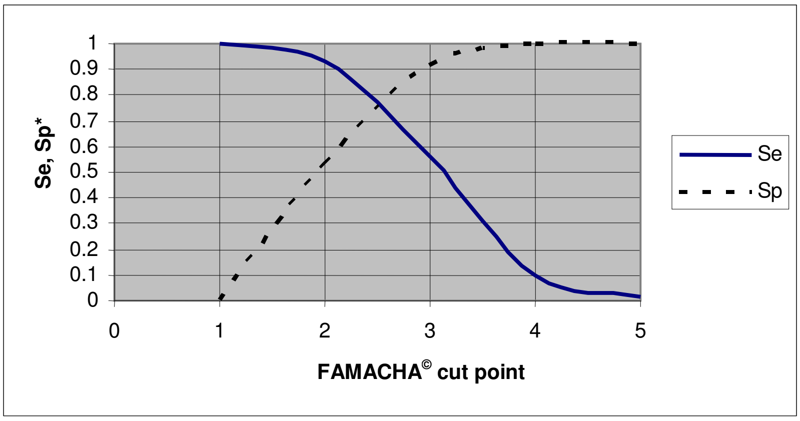

FIG. 4.2b Farm 1. Two-graph Receiver Operating Characteristic plot for a haematocrit cut-off of ≤19 %. *Se = sensitivity, Se = specificity.

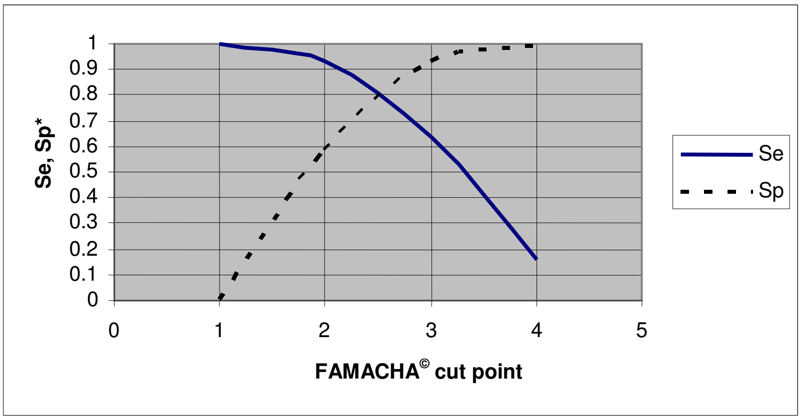

FIG. 4.2c Farm 2. Two-graph Receiver Operating Characteristic plot for a haematocrit cut-off of ≤22 %. FAMACHA© category 5 not represented. *Se = sensitivity, Sp = specificity.

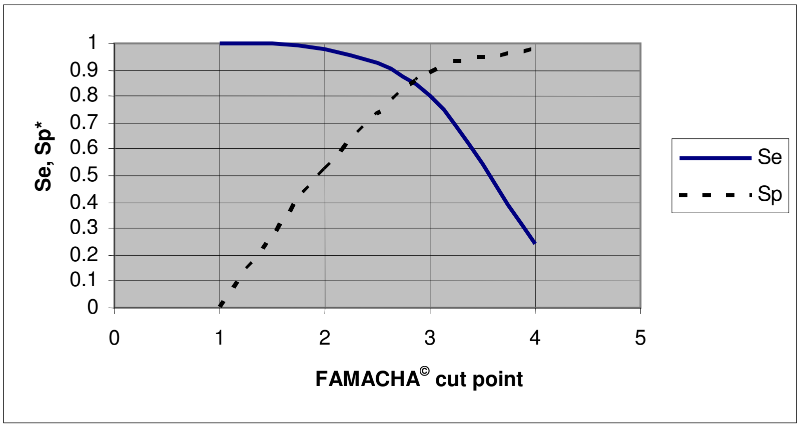

FIG. 4.2d Farm 2. Two-graph Receiver Operating Characteristic plot for a haematocrit cut-off of ≤19 %. FAMACHA© category 5 not represented. *Se = sensitivity, Sp = specificity.

TABLE 4.1 Farm 1. Haematocrit cut-off is ≤22 %. Results of the rating method for FAMACHA© cut point, sensitivity, specificity, percentage of sheep correctly classified, and likelihood ratios (LR).

FAMACHA© Cut point	Sensitivity	Specificity	Classified correctly	LR+	LR-
(≥1)	100.00 %	0.00 %	49.41 %	1.0000	-
(≥2)	83.18 %	63.05 %	73.00 %	2.2512	0.2667
(≥3)	39.94 %	96.19 %	68.40 %	10.4766	0.6244
(≥4)	5.71 %	100.00 %	53.41 %	0.9429	-
(≥5)	0.90 %	100.00 %	51.04 %	0.9910	-
(>5)	0.00 %	100.00 %	50.59 %	1.0000	-

TABLE 4.2 Farm 1. Haematocrit cut-off is ≤19 %. Results of the rating method for FAMACHA© cut point, sensitivity, specificity, percentage of sheep correctly classified and likelihood ratios (LR).

FAMACHA© Cut point	Sensitivity	Specificity	Classified correctly	LR+	LR-
(≥1)	100.00 %	0.00 %	27.89 %	1.0000	-
(≥2)	93.09 %	53.09 %	64.24 %	1.9842	0.1303
(≥3)	57.98 %	92.39 %	82.79 %	7.6156	0.4548
(≥4)	10.11 %	100.00 %	74.93 %	0.8989	-
(≥5)	1.60 %	100.00 %	72.55 %	0.9840	-
(>5)	0.00 %	100.00 %	72.11 %	1.0000	-

TABLE 4.3 Farm 1. Haematocrit cut-off, area under Receiver Operating Characteristic curve, standard error (SE) of the area under the curve and confidence limits (CL) of the area under the curve.

Haematocrit cut-off	n	Area under the curve	SE	Lower 95 % CL	Upper 95 % CL
≤22 %	674	0.7902	0.0159	0.75811	0.82092
≤19 %	674	0.8353	0.0161	0.80511	0.86254

TABLE 4.4 Farm 2. Haematocrit cut-off is ≤22 %. Results of the rating method for FAMACHA© cut point, sensitivity, specificity, percentage of sheep correctly classified and likelihood ratios (LR). FAMACHA© category 5 not represented.

FAMACHA© Cut point	Sensitivity	Specificity	Classified correctly	LR+	LR-
(≥1)	100.00 %	0.00 %	25.19 %	1.0000	-
(≥2)	93.60 %	58.37 %	67.25 %	2.2485	0.1097
(≥3)	64.04 %	93.53 %	86.10 %	9.9015	0.3845
(≥4)	16.26 %	99.67 %	78.66 %	49.0127	0.8402
(>4)	0.00 %	100.00 %	74.81 %	1.0000	-

TABLE 4.5 Farm 2. Haematocrit cut-off is ≤19 %. Results of the rating method for FAMACHA© cut point, sensitivity, specificity, percentage of sheep correctly classified and likelihood ratios (LR). FAMACHA© category 5 not represented.

FAMACHA© Cut point	Sensitivity	Specificity	Classified correctly	LR+	LR-
(≥1)	100.00 %	0.00 %	14.27 %	1.0000	-
(≥2)	98.26 %	52.53 %	59.06 %	2.0701	0.0331
(≥3)	80.87 %	89.00 %	87.84 %	7.3527	0.2149
(≥4)	24.35 %	98.99 %	88.34 %	24.0348	0.7643
(>4)	0.00 %	100.00 %	85.73 %	1.0000	-

TABLE 4.6 Farm 2. Haematocrit cut-off, area under Receiver Operating Characteristic curve, standard error (SE) of the area under the curve and confidence limits (CL) of the area under the curve.

Haematocrit cut-off	n	Area under the curve	SE	Lower 95 % CL	Upper 95 % CL
22 %	806	0.8671	0.0141	0.84185	0.88991
19 %	806	0.9012	0.0135	0.87799	0.92051

4.4 Discussion

An important part of the validation process of the FAMACHA© system has been to investigate the accuracy of the clinical evaluation in relation to the haematocrit gold standard used (Van Wyk *et al*. 2001a; Kaplan *et al.* 2004). The Receiver Operating Characteristic curve, as used in the analysis of the sensitivity of the FAMACHA© system, is a graphical representation of the trade-off between the rates of false negatives and false positives for every possible cut point, and thus represents the trade-offs between sensitivity (Se) and specificity (Sp). Conventionally, the Receiver Operating Characteristic curve plot depicts 1 minus specificity (1-Sp) on the X-axis and sensitivity (Se) on the Y-axis. The area under the Receiver Operating Characteristic curve is a quantitative measure of how rapidly the curve rises to the upper left corner of the plot. The closer the area is to 1, the better the diagnostic test, since, if an area of 1 were to be obtained for the area under the curve, it would imply that the test is perfectly able to discriminate between healthy and diseased individuals, as Se and Sp would both be 100 %. If, on the other hand, the area under the curve is 0.5 or close to 0.5, then the test would have the same discriminatory power as the binomial probability obtained by tossing a coin to determine the infection status of an animal. In practice, useful diagnostic tests will have an area between these two extremes, that is between perfect discrimination (area under the curve = 1) and no discrimination (area under the curve = 0.5) (Greiner *et al*. 2000).

In this work the rating method was used to calculate the area under the curve (Hanley & McNeil 1982). The first rows in Tables 4.1 and 4.2, and Tables 4.4 and 4.5 represent the extreme case where all individuals are classified as test positive, or alternatively stated, that all individuals in FAMACHA© categories 1–5 are regarded as test positive. This would lead to a situation where there would be a 100 % true positive rate among the truly diseased, and implies that Se = 1, but that there would also be a 100 % false positive rate among individuals which are truly non-diseased with Sp = 0. The last row in the above mentioned tables represents the other extreme where all individuals are classified as test negative, but in this case, there is a 0 % true positive rate among the truly diseased (Se = 0) and a 0 % false positive rate among the truly non-diseased (Sp = 1). At a practical level, if an individual is always regarded as test positive or test negative, it would mean that test results are ignored or no testing has taken place. An example of the first instance where all individuals are always regarded as test positive is commonly encountered in conventional blanket drenching systems when all individuals are treated at any given time. The diagnostic Se and Sp are always a function of the selected cut point of the test (Greiner *et al*. 2000).

The smallest area under the curve was calculated for Farm 1 data (0.79), for a haematocrit cut-off of ≤22 %, while the largest was for Farm 2 (0.90) for a haematocrit cut-off of ≤19 %. All area under curve values obtained in this study were much larger than 0.5, indicating that the discriminating power for the FAMACHA© test was good for both of the haematocrit cut-off values for FAMACHA© data from both farms. On Farm 1 the area under the curve for both haematocrit cut-off values (Table 4.3) was lower than for the same cut-off values on Farm 2 (Table 4.6). However, the accuracy of FAMACHA© classification was also lower on Farm 1 than on Farm 2. All calculated under-curve areas fell within the moderately accurate range of 0.7<area under the curve<0.9 (Greiner *et al*. 2000), while the area under the curve for Farm 2 was highly accurate at a value of 0.9 (Table 4.6) for a cut-off of ≤19 %. The smallest area under the curve was for Farm 1 (0.79), and this value is higher than the accepted value for a moderately accurate test. Thus, the smallest calculated probability in this study, that a randomly drawn individual that is truly diseased as defined by the haematocrit cut-off value has a higher haematocrit value than a randomly drawn non-diseased individual is 0.79 (Table 4.3; Farm 1), while the highest probability was 0.9 (Table 4.6; Farm 2). Both of the haematocrit values used as cut-offs in this study therefore yielded a moderate to high probability of discriminating between diseased and non-diseased individuals, as defined by their respective under-curve areas.

Kaplan *et al*. (2004), in the southern United States, reported an Se of 100 % in sheep when FAMACHA© categories 3, 4 and 5 were considered test positive with a haematocrit cut-off of ≤15 %. However, they also found that Se decreased to 92 % if the haematocrit cut-off was increased to ≤19 %. Bath *et al.* (2001) found that in goats farmed under resource-poor conditions in South Africa an Se of 80 % was obtained when FAMACHA© categories 3–5 were considered test positive, and recommended that the treatment threshold under these conditions should be FAMACHA© category 3.

In the present work a low proportion (1.9 %) of false negatives for Farm 1 was obtained for drenching sheep in FAMACHA© categories 2–5, with a haematocrit cut-off of ≤19 % (Table 3.3b; Chapter 3). These results are clearly supported by the two-graph Receiver Operating Characteristic plots for Farm 1 (Fig. 4.2a and b), where it can be read directly off the plot that 0.9<Se<1 for a FAMACHA© cut point of 2 and a haematocrit cut-off of ≤22 %, and for a haematocrit cut-off of ≤19 % the same Se could be obtained for a FAMACHA© cut point of 3. Taking into account the findings in Chapter 3, where it was recommended that all sheep in FAMACHA© categories 2–5 be treated, the results of the two-graph Receiver Operating Characteristic analysis for Farm 1 indicate that an Se of at least 90 % would be obtained for a haematocrit cut-off of ≤22 %. Thus, in the diagnostic context of the use of the FAMACHA© system, it would be realistic to specify an Se of at least 90 % in order to ensure that at least 90 % of truly diseased individuals are detected and treated for the defined cut-off values. The selected Se value of 90 % could have been higher, but it was selected as a realistic value for the sheep on the farms concerned. The Se value, however, will vary according to the FAMACHA© cut points chosen as treatment thresholds, as well as according to the haematocrit cut-off values selected as the reference variables for the test. The maximum accuracy for a given haematocrit cut-off value can be read directly off the two-graph Receiver Operating Characteristic graphs in Fig. 4.2 a–d and following this approach, the FAMACHA© cut point value where the two curves cross for all analyses is at FAMACHA© 3, indicating that all sheep in FAMACHA© categories 3–5 should be treated under the assumption that maximum accuracy should be used to select FAMACHA© cut points (Fig.

4.2 a–d). However, simply maximizing the accuracy, defined as the point where the Se and Sp curves intersect (Greiner & Gardner 2000) does not reflect the epidemiological risk situation for FAMACHA© implementation. The relative consequences of false negative test results are potentially much more serious than those for false positive diagnoses due to the selective nature of FAMACHA© treatment (Bath *et al*. 2001).

The main goal for FAMACHA© implementation should be to maximise Se while still leaving a large proportion of the flock undrenched to maintain a proportion of the parasite population in refugia (Van Wyk 2001). For Farm 1 Se was maximised at the pre-selected value of 0.9 at a FAMACHA© cut point of 2 and a haematocrit cut-off of ≤22 % (Fig. 4.2a), while for a haematocrit cut-off of ≤19 %, an Se of 0.9 was achieved at a FAMACHA© cut point of 3 (Fig. 4.2b). However, because the accuracy of FAMACHA© diagnosis was low on Farm 1 it is recommended that unless the primary problem of misclassification can be overcome through re-training of the evaluator on the farm, a FAMACHA© cut point of 2 be selected together with a haematocrit cut-off of ≤22 %. This will ensure that at the more conservative cut-off of ≤22 %, 90 % of sheep with a haematocrit of ≤22 % will be detected as diseased and treated, while the total proportion of the animals recommended for treatment would still only comprise a maximum of (278+124)/675, or 59 % of the flock. This can be observed from Table 3.3a, where there were 278 true positives and 124 false positives among the 402 individuals classified into FAMACHA© categories 2–5, and furthermore, only 56/675 or 8 % were false negatives for a haematocrit cut-off of ≤22 % if FAMACHA© categories 2–5 were to be treated. These findings are further supported by the likelihood ratio of a positive test result (LR+), which is a prevalence-independent combined measure of Se and Sp that is a representation of the odds of the pre-test and post-test probability of disease, which in turn is conditional to a positive test result (Greiner & Gardner 2000). For Farm 1 the LR+ was higher for a haematocrit cut-off of ≤22 % and a FAMACHA© cut point of 2 (Table 4.1) at 2.25, than for a haematocrit cut-off of ≤19 % where the LR+ was 1.98 for the same FAMACHA© cut point (Table 4.2). This also means that the total area under the Receiver Operating Characteristic curve for a FAMACHA© cut point of 2 is larger for a haematocrit cut-off of ≤22 % than for a cut-off of ≤19 %, with a resulting higher probability of discrimination between diseased and non-diseased individuals. These two-graph Receiver Operating Characteristic results support the findings in Chapter 3, but the increased refinement of two-graph Receiver Operating Characteristic analysis indicated that, because of the high proportion of misclassified sheep on Farm 1, a FAMACHA© cut point of 2 is the most appropriate treatment threshold.

For Farm 2 Se was maximised at the pre-selected value of 0.9 for a FAMACHA© cut point of 3 for both haematocrit cut-off values of ≤22 % and ≤19 % (Figs. 4.2c and d). However, because of the higher accuracy of FAMACHA© evaluation on Farm 2 compared to Farm 1, the proportion of sheep classified correctly according to FAMACHA© on Farm 2 was much higher than that achieved on Farm 1 (Chapter 3). The results of the two-graph Receiver Operating Characteristic analyses for Farm 2 indicated that a FAMACHA© cut point of 3 could be therefore be recommended for both haematocrit cut-off values of ≤22 % and ≤19 % (Fig. 4.2c and d). The LR+ for Farm 2 was 9.9 and 7.3 for the haematocrit cut-offs of ≤22 % and ≤19 %, respectively, for a FAMACHA© cut point of 3 (Tables 4.4 and 4.5). This value is much higher than the LR+ for a FAMACHA© cut point of 2, at 2.2 and 2.07, for the haematocrit cut-offs of ≤22 % and ≤19 %, respectively, indicating a higher cumulative area under the curve for a FAMACHA© cut point of 3 than for a cut point of 2 on Farm 2, and therefore also a higher degree of discrimination between diseased and non-diseased individuals. Further support is also provided for the recommendation that a FAMACHA© cut point of 3 would be safe to implement on Farm 2.

4.5 Conclusion

The calculation of the area under the curve for the FAMACHA© system for haematocrit cut-off values of ≤22 % and ≤19 % indicated that the diagnostic accuracy of the FAMACHA© system was moderate to high, indicating that the system as implemented on the two farms examined here is effective in discriminating between diseased and non-diseased individuals. The area under the curve index values ranged from a minimum value of 0.79 on Farm 1 to a maximum value of 0.90 on Farm 2, and since the area under the curve represents the probability that a randomly selected individual with the disease will have a lower haematocrit value than a randomly selected individual without the disease for a given haematocrit cut-off, the FAMACHA© test is clinically relevant and useful. The selection of suitable FAMACHA© cut points as thresholds for anthelmintic treatment, where it is assumed that Se must have a certain minimum selected value, could readily be achieved with two-graph Receiver Operating Characteristic curve analysis. On Farm 1 a FAMACHA© cut point of 2 and a haematocrit cut-off of ≤22 % was found to be epidemiologically most suitable for a selected Se of 90 %. On Farm 2, where the risk of disease was lower, a FAMACHA© cut point of 3 and a haematocrit cut-off of ≤19 % could be safely recommended for a required Se of 90 %.

The results of this work show that implementing the FAMACHA© system with haematocrit cut-off values of ≤22 % and ≤19 % is sensitive and adequately specific and that these values can safely be used to monitor haemonchosis in sheep on the two farms.