CHAPTER 3

Validation of the FAMACHA© eye colour chart on two South African sheep farms under commercial farming conditions

3.1 Introduction

Multiple anthelmintic resistance in the highly pathogenic nematode parasite *Haemonchus contortus* is a severe problem on commercial sheep farms in South Africa, and has the potential to become just as problematic on communal farms in the country. The proportion of the parasite population that escapes drug selection is at present thought to be the most important factor in influencing the rate of development of resistance (Van Wyk 2001; Leathwick *et al.* 2006). It has been largely in response to this that targeted selective treatment systems, notably FAMACHA©, were developed. Although much has been done to validate the FAMACHA© system in South Africa, it is still important that the method be tested for its operating characteristics in terms of sensitivity, specificity and predictive values under farming conditions on an ongoing basis. Once the sensitivity and specificity of a test are known, then a corrected estimate of the true prevalence of disease can be estimated. It is important to know the probability that an animal classified as test positive is truly positive and alternatively the probability that an animal classified as test negative is truly negative. These two probabilities are the predictive values of the test and they depend on sensitivity, specificity and prevalence (Thrusfield 2001). The standard 2 × 2 table method (Thrusfield 2001) was used in this investigation, to calculate the above-mentioned properties. When using the FAMACHA© system, disease management of a flock depends on accurate identification of diseased individuals, to include these individuals in the proportion of the flock that is to be treated. The FAMACHA© system has in this respect been successfully used as a stratification method, to classify individual animals requiring treatment (Bath *et al.* 2001).

The FAMACHA© system reduces the uncertainty about the state of haemonchosis in individual sheep, and can therefore be regarded as a diagnostic test (Greiner & Gardner 2000). The clinical performance of a diagnostic test can be described in terms of its diagnostic accuracy, which represents the ability of the test to correctly classify test subjects into clinically relevant subgroups (Zweig & Campbell 1993). However, as the FAMACHA©

system is based on a rating method (Hanley & McNeil 1982) with the different FAMACHA© categories from 1–5 representing the increasing probability of an abnormal test result, the results of FAMACHA© classification of a sample of sheep are required to be dichotomised into two groups, each representing the infected and uninfected proportions of the flock. This is an artificial distinction, since, due to the extra-binomial nature of the variation in worm burden infection in flocks, almost all animals are infected, but due to overdispersion, the minority of the animals harbour the highest individual number of worms (Barger 1985; Wilson, Grenfell & Shaw 1996; Herbert & Isham 2000). While diagnostic tests are subject in terms of sensitivity and specificity to arbitrary definitions (Begg 1987), the application of the FAMACHA© system provides reasonable scope to adjust for this arbitrariness because it has five categories that could potentially provide five different views of the infection status of a flock.

One limitation to the 2 × 2 table method of estimating sensitivity and specificity is that there is usually a single pre-determined criterion, referred to as a cut-off point, to indicate a true positive test result (Linden 2006). As an example, all animals in a sample could be classified as test positive, or FAMACHA© categories 2–5 could be regarded test positive with FAMACHA© category 1 test negative, FAMACHA© categories 3–5 test positive and FAMACHA© categories 1 and 2 test negative, etc. Although these dichotomized FAMACHA© test results are still arbitrarily chosen, assessment of the epidemiological risk of infection will have a strong influence on FAMACHA© categories selected as thresholds of infection status. Thus, the “resolution” of the FAMACHA© system can easily be adjusted to estimate test measures of sensitivity and specificity against a given haematocrit cut-off value.

Overdispersed worm burdens are related to a number of factors due to the fact that some animals may habitually graze in an area of pasture with higher levels of larvae, differences in immunity status of animals due to age, sex, nutrition, parturition, or previous exposure and genetic differences in animals’ ability to tolerate or expel worms. This results in extra-binomial variation, necessitating the use of a negative binomial model, rather than a poisson model. Additionally, the immune response caused by macroparasites such as *H. contortus* may depend on the number of parasites in a particular host (Bishop & Stear 2003). Barger (1985) described the negative binomial distribution of trichostrongylid nematodes in grazing lambs (n = 104), and indicated that the distribution of worm burdens in a flock would have an effect on anthelmintic treatment. Anderson & May (1982) estimated that, in helminth-infected humans, selective treatment of the most heavily infected 8 % of the population would reduce

the mean worm burden of the population by 50 % if worms were highly overdispersed, but added that the advantages of targeted treatment in terms of costs no longer incurred would have to be offset against the costs of identifying the most heavily infected individuals. It was largely also in response to the latter problem, that the FAMACHA© system was developed.

A validation study of the FAMACHA© system was conducted on a variety of commercial sheep farms in South Africa (Van Wyk & Bath 2002). Two of these farms, for which extensive data sets were generated, were selected for the present investigations. Over a period of five years, the system was tested in Merino sheep on the first farm (Farm 1; 26°40'53"S, 30°16'47"E) under routine farming conditions. The colours of the conjunctivae of sheep were scored on a 1–5 scale using the FAMACHA© chart, and blood samples were periodically collected from each animal for haematocrit determination. Only sheep that were classified into FAMACHA© categories 3, 4 and 5 were treated with anthelmintics, with approximately 260 sheep being re-evaluated at each sampling event. The farmer himself was mainly responsible for FAMACHA© scoring on Farm 1. Data for both FAMACHA© scores and haematocrit were evaluated using different criteria for anaemia. Firstly, FAMACHA© eye scores of 3, 4 and 5 and haematocrit values of ≤22 %, ≤19 %, and ≤15 % were separately considered to be anaemic. Sensitivity and specificity were maximised at a haematocrit cut-off value of ≤15 % at 0.83 and 0.85 respectively, but this haematocrit value is thought to be too low to be safe under conditions of selective anthelmintic treatment where animals are only treated when they are deemed to be anaemic. In contrast, sensitivity increased to 0.93 when eye scores of 2, 3, 4, and 5 were considered anaemic at the safer and more realistic haematocrit cut-off value of ≤19 %, but the predictive value of a positive was low indicating that many non-anaemic animals would be treated. Considerable classification bias was detected in scoring for FAMACHA© categories 1–4 on this farm, leading to the recommendation that animals in FAMACHA© category 2 should have been treated in addition to animals in FAMACHA© categories 3–5, in order to increase sensitivity and prevent the likelihood of non-treatment of sheep with a haematocrit of ≤19 %.

A second validation study was conducted on data from Farm 2 (27°23'48"S, 29°22'24"E), a commercial Merino farm, over a period of two *Haemonchus* seasons, under the so-called "Field Ram Club" system. Haematocrit determinations were done on all the rams each year, both at the beginning and at the height of the *Haemonchus* season, with approximately 200 rams being re-evaluated at each sampling event. Rams were only treated if their haematocrits were ≤15 %. In the interim every ram judged to be in FAMACHA© category 4

or 5 was bled for haematocrit determination, and only rams with haematocrit values of 15 % or lower were dewormed with effective anthelmintics. The results from Farm 2 indicated that the accuracy of anaemia estimation was higher than that of Farm 1, and that for identical haematocrit cut-off values and proportions of the sampled flock considered to be diseased, sensitivity, i.e. the conditional probability that a sampled animal has a positive test result, was always higher for Farm 2. This meant that on Farm 2, any animal defined as anaemic by the pre-determined cut-off value for the haemaocrit, had a higher probability than for Farm 1 of being detected as anaemic and treated. Sheep on Farm 2 were generally less anaemic, despite being treated at a lower FAMACHA© threshold than sheep on Farm 1, but these sheep were more accurately "scored" by the farmer into FAMACHA© categories, and were also evaluated much more frequently during periods of peak worm challenge.

3.2 Materials and methods

3.2.1 Origin of data and FAMACHA© test procedures

The data that was analysed consisted of anaemia status as evaluated by FAMACHA© score, and haematocrit values, originating from naturally infected sheep on the two farms (Fig. 3.1). The farms are situated in the summer rainfall region of South Africa. Merino sheep are predominant in this area, and *H. contortus* is the dominant nematode parasite. Climatically, the region is a part of the temperate eastern plateau, at an altitude of approximately 1 500m above sea-level, with cool, rainy summers and cold, dry winters. Over a five year period on Farm 1, from November 2000 until April 2005, two classes of animals were introduced annually into a series of FAMACHA© trials, namely replacement rams (RAMREP) and replacement ewes (EWEREP), each individually identified with a uniquely numbered ear tag. The two groups of sheep were farmed under extensive conditions, in separate flocks, according to sex. Each flock was grazed at intervals of approximately 3–5 weeks through a series of different paddocks according to available herbage. The number of sheep on the farm during the study period varied between 1 200 and 1 800, but only 130–200 sheep of each class were sampled at each FAMACHA© evaluation in the various trials per class and year. At the start of each of the five annual trials, each sheep was scored into a FAMACHA© category, its body weight and condition score determined, and it was dewormed. This was followed by a period during which only animals clinically judged to be in FAMACHA© categories 3–5 were dewormed. However, once general "severe worm challenge" was evident, usually in January or February of each year, all sheep were again dewormed. Then,

until the end of each trial, only the animals in FAMACHA© categories 3–5 were treated as before. From November to April the following year, sheep were mostly evaluated at intervals of 3–4 weeks, but in some instances the evaluation intervals were longer, at up to five weeks. A total of 7–11 sampling events took place per worm season. At the time of the "severe worm challenge", haematocrit determinations were done on all the sheep. In addition, during the 2000/2001 season, haematocrit determinations were done both initially at the start of the trials in November, and during the height of the worm season, in January or February.

On Farm 2, data collected from rams over a period of two seasons, 1998/1999 and 1999/2000, was evaluated. In this system, young rams, usually 6 months old at the start of the trial, belonging to club members were compared in the field over a period of 10–11 months. The rams were grazed on common pasture. At the end of each trial, the rams were evaluated for weight gain and wool production, and the best performing rams were sold at auction. During the course of each trial, a haematocrit determination was done for each animal judged to have had a FAMACHA© score of 4 or 5, and it was only treated if the haematocrit was ≤15 %. However, on five different occasions over the two-year trial period, the haematocrit of every ram was determined, in addition to its FAMACHA© score, for calibration purposes. This was the principal difference between the two data sets, as, even though rams on Farm 2 were scored according to FAMACHA©, only individuals in FAMACHA© categories 4 and 5 were treated, and then only if their haematocrits were ≤15 %. During the 1999 season worm challenge became intense from January on Farm 2, and FAMACHA© evaluation and sampling of individual animals for haematocrit determination and drenching was carried out more frequently, at which point the animals were evaluated weekly during the peak worm season in February instead of fortnightly as before. During the 1999/2000 season, lower levels of infection were experienced, and the increased frequency of evaluation was not repeated.

FIG. 3.1 Map of South Africa indicating position of Farm 1 (red square) and Farm 2 (blue square). Refer to text for geographical co-ordinates.

3.2.2 Statistical analysis

Data from the RAMREP and EWEREP classes on Farm 1 were pooled for comparing the accuracy of the clinical FAMACHA© scores with the haematocrit value used to determine the true presence or absence of anaemia in the trial animals, similar to the method used by Vatta *et al*. (2001) and Kaplan *et al.* (2004). For the observed haematocrit values of FAMACHA© categories 1–5, the mean, median, 5th percentile, 95th percentile, and standard deviation were calculated and tabulated against their ordinated FAMACHA© scores, using Excel spreadsheets. Two-way frequency tables were constructed, and sensitivity, specificity, predictive value of a positive and predictive value of a negative were calculated for the data. FAMACHA© scores of 3, 4 and 5 were considered to be test positive and FAMACHA© scores of 1 and 2 were considered to be test negative.

For the purposes of determination of test sensitivity and specificity on Farm 1, three different haematocrit cut-off values were considered to be anaemic. Accordingly, values were considered anaemic if ≤22 %, ≤19 % and ≤15 %, and the above test parameters were calculated separately for these three values. The haematocrit value of ≤22 % was chosen as it is the upper limit of FAMACHA© category 3, which was used on this farm as a treatment threshold. True positives were defined as sheep that were anaemic with haematocrits of ≤22 %, ≤19 %, or ≤15 % and FAMACHA© scores of 3, 4 or 5. Sheep were defined as false positives if they were not anaemic, with haematocrits of >22 %, >19 %, or >15 % but with FAMACHA© scores of 3, 4 or 5. False negatives were defined as sheep that were anaemic with FAMACHA© scores of 1 and 2 while true negatives were defined as sheep that were not anaemic but with FAMACHA© scores of 1 and 2.

A further analysis was conducted with FAMACHA© categories 2, 3, 4 and 5 considered to be test positive and FAMACHA© category 1 considered to be test negative. For this part of the analysis, haematocrit values were considered anaemic if ≤22 % or ≤19 %. Test operating characteristics were calculated as described above.

Data for Farm 2 were analysed in a similar way to Farm 1 but haematocrit values were considered anaemic only if ≤22 % or ≤19 % and only individuals in FAMACHA© 3–5 were considered to be test positive for comparison between the two farms.

3.3 Results

Farm 1

The results of Farm 1 indicated that the percentages of sheep that would be correctly treated with haematocrit cut-off values of ≤22 %, ≤19 % and ≤15 % when FAMACHA© categories 3–5 were treated were 68.3 %, 82.8 % and 65.6 %, respectively (Table 3.1a–c). The sensitivity of the FAMACHA© system to identify sheep that are anaemic with haematocrit cut-off values of ≤22, ≤19 and ≤15 when only animals that were in FAMACHA© categories 3, 4 and 5 were treated, was low for all haematocrit cut-off values (Table 3.2), with the highest sensitivity being obtained for a cut-off of ≤15 %. The specificity of the FAMACHA© method on the other hand, was highest for a haematocrit cut-off value of ≤22 %, at 96 %. The haematocrit cut-off value of ≤22 % was chosen as it is the upper haematocrit limit of FAMACHA© category 3, and would thus include treatment of FAMACHA© categories 4 and 5, in addition to FAMACHA© category 3. However,

haematocrit cut-off values of ≤19 % and ≤15 %, as described by Kaplan *et al.* (2004) were also evaluated. Sensitivity increased as the haematocrit cut-off value decreased (Table 3.2), but the predictive value of a positive decreased. Thus, using FAMACHA© categories 3–5 inclusive (with FAMACHA© category 3 as a threshold), and a haematocrit cut-off of ≤22 %, only 40 % of animals that were anaemic would have been treated due to the large number of false negatives. The proportion of animals correctly treated was highest for a haematocrit cut-off of ≤19 %, at 82.8 % (Table 3.1b), but only 58 % of sheep with a haematocrit of ≤19 % would have been detected (Table 3.2), as the majority of animals correctly left untreated would have been true negatives.

TABLE 3.1a Farm 1. Haematocrit cut-off value is ≤22 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses, for sheep with assigned ranges in haematocrit values which are based on drenching of sheep with FAMACHA© scores of 3, 4 and 5.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤22 %	201 (29.7)	-	133 (19.7)	334 (49.5)
>22 %	-	13 (1.9)	328 (48.6)	341 (50.5)
Total	201 (29.7)	13 (1.9)	461 (68.3)	675 (100)

TABLE 3.1b Farm 1. Haematocrit cut-off value is ≤19 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses, for sheep with assigned ranges in haematocrit values which are based on drenching of sheep with FAMACHA© scores of 3, 4 and 5.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤19 %	79 (11.7)	-	109 (16.1)	188 (27.8)
>19 %	-	37 (5.5)	450 (66.6)	487 (72.1)
Total	79 (11.7)	37 (5.5)	559 (82.8)	675 (100)

TABLE 3.1c Farm 1. Haematocrit cut-off value is ≤15 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses, for sheep with assigned ranges in haematocrit values which are based on drenching of sheep with FAMACHA© scores of 3, 4 and 5.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤15 %	11 (1.62)	-	56 (8.3)	67 (10)
>15 %	-	90 (13)	518 (76.7)	608 (90)
Total	11 (1.62)	90 (13)	574 (65.6)	675 (100)

TABLE 3.2 Farm 1. Sensitivity (Se), specificity (Sp), positive predictive value (Pv+), negative predictive value (Pv-), and prevalence (P) for trial data for given haematocrit cut-off values and treatment of sheep in FAMACHA© categories 3–5. The value for prevalence was calculated from standard two-way frequency tables.

Haematocrit value	Se	Sp	Pv +	Pv -	P	Confidence interval (95 %)
≤22 %	0.40	0.96	0.91	0.62	0.49	(0.458 – 0.532)
≤19 %	0.58	0.92	0.75	0.85	0.27	(0.245 – 0.312)
≤15 %	0.83	0.85	0.38	0.98	0.10	(0.089 – 0.111)

In contrast, when FAMACHA© scores of 2 – 5 (inclusive), and haematocrit cut-off values of ≤22 % and ≤19 % were considered anaemic (Table 3.3a and b), sensitivity was highest when a haematocrit value of ≤19 % was considered anaemic, at 93 % (Table 3.4). Thus, if all sheep in FAMACHA© categories 2–5 were treated, 93 % of sheep with a haematocrit of ≤19 % would have been detected, due to the small number of false negatives (Table 3.3b). The total percentage of correctly treated animals, i.e. true positives + true negatives, would have been 64 %, but this would have been due to the relatively high proportion of false positives (Table 3.3b).

The FAMACHA© scores vs. assigned and observed median haematocrit values are given in Table 3.5. Observed mean and median haematocrit values were lower than assigned mean values, indicating bias, or misclassification, on the part of the evaluators (Table 3.5). For example, the assigned minimum haematocrit value of FAMACHA© category 1 is given to be above 28 %, but the observed median for animals classed as being in this category was 23 %, and the assigned median haematocrit value of FAMACHA© category 2 (range 23–27 %) is given as 25 %, yet a relatively low value of 19.5 % was observed from the data. Similarly, the given median haematocrit value for FAMACHA© category 3 is 20 %, with an observed median value of 15 %, and the given and observed median haematocrit values of FAMACHA© category 4 are 15 % and 11 % respectively. The proportion of haematocrit values falling within assigned ranges are given in Table 3.6.

TABLE 3.3a Farm 1. Haematocrit cut-off value is ≤22 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses, for sheep with assigned ranges in haematocrit values, which are based on drenching of sheep with FAMACHA© scores of 2, 3, 4 and 5.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤22 %	56 (8.3)	-	278 (41.2)	334 (49.5)
>22 %	-	124 (18.4)	217 (32.1)	341 (50.5)
Total	56 (8.3)	124 (18.4)	495 (73.3)	675 (100)

TABLE 3.3b Farm 1. Haematocrit cut-off value is ≤19 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses, for sheep with assigned ranges in haematocrit values, which are based on drenching of sheep with FAMACHA© scores of 2, 3, 4 and 5.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤19 %	13 (1.9)	-	176 (26.1)	189 (28.0)
>19 %	-	227 (33.6)	259 (38.4)	486 (72.0)
Total	13 (1.9)	227 (33.6)	435 (64.4)	675 (100)

TABLE 3.4 Farm 1. Sensitivity (Se), specificity (Sp), positive predictive value (Pv+), negative predictive value (Pv-), and prevalence (P) for trial data for given haematocrit cut-off values and treatment of sheep in FAMACHA© categories 2–5. The value for prevalence was calculated from standard two-way frequency tables.

Haematocrit value	Se	Sp	Pv +	Pv -	P	Confidence interval (95 %)
≤22 %	0.83	0.63	0.69	0.79	0.48	(0.442 – 0.510)
≤19 %	0.93	0.53	0.43	0.95	0.28	(0.246 – 0.314)

TABLE 3.5 Farm1. FAMACHA© score vs. haematocrit: assigned values, observed values and percentiles (n = 675)

FAMACHA© score	Assigned median value of haematocrit range (%)	Observed median haematocrit value (trial data) (%)	Percentage below assigned median for observed haematocrits	Fifth percentile of observed haematocrit value	Ninety-fifth percentile of observed haematocrit value
1	30	23	23 %	19.7	30.5
2	25	19.5	22 %	15.9	27.2
3	20	15	25 %	10.6	23.9
4	15	11	26 %	6.5	18.7
5	10	10.5	-	8.6	11.5

TABLE 3.6 Farm 1. FAMACHA© categories, sample size, assigned haematocrit range and percentage of observed haematocrit values within the assigned range.

FAMACHA© category	n	Assigned haematocrit range of FAMACHA© category*	Percentage of observed haematocrit values within assigned range
1	273	≥28 %	18.8 %
2	258	23 – 27 %	27.9 %
3	126	18 – 22 %	37.5 %
4	16	13 –17 %	44 %
5	3	≤12 %	100 %

* Van Wyk &Bath (2002)

For the intermediate FAMACHA© categories 2, 3 and 4, only 27.9 %, 37.5 % and 44 % of observed haematocrit values respectively, fell within the given limits. For FAMACHA© category 1 only 18.8 % of haematocrit values were above the lower limit of 28 % for the category, while for FAMACHA© category 5, 100 % of the observed haematocrit values were below the upper limit of 12 % but note that there were only three sheep in the latter category. There was thus an increase in the accuracy of FAMACHA© classification from FAMACHA© category 1, which had the lowest overall accuracy of classification, to FAMACHA© category 5, which had the highest accuracy (Table 3.6).

Farm 2

The proportions of false negatives, false positives and correctly treated rams for this farm for individuals treated only if classified as being in FAMACHA© 4 and 5, or with a haematocrit of ≤15 %, are given in Table 3.7a and b. For a positive diagnosis of sheep in FAMACHA© categories 3–5, 86 % of sheep would have been correctly treated at a haematocrit cut-off of ≤22 % (Table 3.7a) while 88 % would have been correctly treated at a haematocrit cut-off of ≤19 % (Table 3.7b). Sensitivity, specificity, positive predictive value, negative predictive value and prevalence data are listed in Table 3.8. FAMACHA© scores versus assigned and observed median haematocrit values for Farm 2 are given in Table 3.9 and the percentage

of observed haematocrit values falling within the assigned ranges are shown in table 3.10. Sensitivity was highest for a cut-off of ≤19 % at 0.80, while specificity was highest for a cut-off of ≤22 % (Table 3.8). The observed median haematocrit values were much closer to their assigned values than was the case on Farm 1 (Table 3.9). The accuracy of FAMACHA© classification was highest for FAMACHA© category 1, with 78 % of observed haematocrit values falling within the assigned range (Table 3.10).

TABLE 3.7a Farm 2. Haematocrit cut-off value is ≤22 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses for rams with assigned ranges in haematocrit values, which are based on drenching of sheep with haematocrit ≤15 %. FAMACHA© categories 1–2 were considered test negative.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤22 %	73 (9)	-	130 (16.1)	203 (25.2)
>22 %	-	39 (4.8)	564 (70.1)	603 (75.0)
Total	73 (9)	39 (4.8)	694 (86.0)	806 (100)

TABLE 3.7b Farm 2. Haematocrit cut-off value is ≤19 %. Results of two-way frequency tables of haematocrit by FAMACHA© score. Percentage of total is given in parentheses, for rams with assigned ranges in haematocrit values, which are based on drenching of sheep with haematocrit ≤15 %. FAMACHA© categories 1–2 were considered test negative.

Haematocrit value	False negatives	False positives	Treatment correct	Total
≤19 %	22 (2.7)	-	93 (11.5)	115 (14.2)
>19 %	-	76 (9.4)	615 (76.3)	691 (85.7)
Total	22 (2.7)	76 (9.4)	708 (87.8)	806 (100)

TABLE 3.8. Farm 2. Sensitivity (Se), specificity (Sp), predictive value of a positive (Pv+), predictive value of a negative (Pv-) and prevalence (P) for trial data for given haematocrit cut-off values and proposed treatment of sheep in FAMACHA© categories 3–5. The value for prevalence was calculated from standard two-way frequency tables. FAMACHA© categories 1–2 were considered test negative.

Haematocrit value	Se	Sp	Pv +	Pv -	P	Confidence interval (95 %)
≤22 %	0.64	0.93	0.77	0.88	0.25	(0.227 – 0.273)
≤19 %	0.80	0.89	0.55	0.96	0.14	(0.116 – 0.164)

TABLE 3.9 Farm 2. FAMACHA© score of rams vs. haematocrit: assigned values, observed values and percentiles (n = 806). FAMACHA© category 5 not represented.

FAMACHA© score	Assigned median value of haematocrit range (%)	Observed median haematocrit value (trial data) (%)	Fifth percentile of observed haematocrit values	Ninety-fifth percentile of observed haematocrit values
1	30	33	23.7	40.8
2	25	26	17.4	36.3
3	20	19.5	12.6	28.3
4	15	16.5	12.5	21.2
5	10	-	-	-

TABLE 3.10. Farm 2. FAMACHA© categories, sample size, assigned haematocrit range and percentage of observed haematocrit values within the assigned range for rams. FAMACHA© category 5 not represented.

FAMACHA category	n	Assigned haematocrit range of FAMACHA© category	Percentage of observed haematocrit values within assigned range
1	365	≥28 %	78 %
2	272	23 – 27 %	40 %
3	134	18 – 22 %	39 %
4	35	13 –17 %	57 %
5	-	≤12 %	-

3.4 Discussion

The work presented here is based primarily on the clinical evaluation data collected from two farms in South Africa. For this reason it was essential to evaluate the accuracy of the clinical data in relation to the haematocrit values used for validating the on-farm use of the FAMACHA© system. In this work, the FAMACHA© diagnostic test was compared with its associated haematocrit values for the two farms where the data were gathered, and sensitivity, specificity, positive predictive values, and negative predictive values were calculated for different haematocrit cut-off values. In order to establish the sensitivity and specificity of a diagnostic test, it is important to decide which test values, or range of values, will be used to indicate a test positive individual (Thrusfield 2001). If there is a significant penalty, such as death, or severe production loss, for failing to detect a test positive individual, then it is important that test sensitivity should be maximised. Within the context of the FAMACHA© system of targeted selective treatment, this would mean that it is essential to have a test that has a high probability of correctly classifying individuals which are truly anaemic and require treatment to avoid death. This is of crucial importance both for ethical reasons and for the economic success of farmers applying the FAMACHA© system. Lowering of test sensitivity will progressively lead to increased numbers of false negatives, which are then not identified as sheep that need treatment, leading to production losses.

Farm 1

One potential drawback of maximising test sensitivity with decreasing prevalence of infection is that a large number of false positive animals are treated. However, this is relatively unimportant in relation to either selection for worm resistance or to financial implications. For instance, despite the relative inaccuracy of FAMACHA© classification on Farm 1, a maximum of 49.5 % of the animals would have been treated, which would have included true positives and false positives under all test criteria (Table 3.1a and Table 3.3a). This compares favourably with blanket treatment systems, where all animals are continually treated both before and during a given worm season. In this series of trials, only sheep scored into FAMACHA© categories 3, 4 and 5 were treated, apart from the blanket drenching events described, and when a realistic haematocrit of ≤19 % was used as a cut-off, only 58 % of sheep that were anaemic were treated (Table 3.2). If a lower haematocrit of ≤15 % were to be used as a cut-off, 83 % of sheep that were truly anaemic would have been treated, but this could potentially be catastrophic to the producer, since the remaining 17 % of sheep with a haematocrit of an already low value of ≤15 % would be in danger of succumbing to haemonchosis. It has been shown that a haematocrit drop of 7 percentage points could occur in as many days, leading to rapid death from terminal anaemia (Malan *et a*l. 2001), and for this reason, a haematocrit cutoff of ≤15 % would be unrealistic for Farm 1 in the present case. A haematocrit cut-off value of ≤19 % would therefore carry less risk under this treatment option. However, if sheep in FAMACHA© category 2 were treated in addition to FAMACHA© categories 3, 4 and 5 in this series of trials, and with a haematocrit cut-off of ≤19 %, then 93 % of sheep that were anaemic would have been detected and treated (Table 3.4). This represents a dramatic improvement over the actual situation where only 58 % of anaemic sheep with a haematocrit of ≤19 % were detected and treated. Even though 33.6 % of the total would have been treated as false positives if FAMACHA© categories 2–5 were treated (Table 3.3b), the total proportion of the animals recommended for treatment would still only have comprised a maximum of only 59 % of the flock. This would almost certainly maintain a sufficient level of refugia for large-scale reduction in selection for anthelmintic resistance while maintaining an acceptable level of parasite control for the producer.

The results from Farm 1 indicate that a high degree of misclassification occurred on the farm (Tables 3.5 and 3.6). Several reasons have been advocated for FAMACHA© misclassification. Among these are (i) wrong interpretation due to complacency and over-

confidence in estimating anaemia score without reference to the FAMACHA© card for calibration, (ii) infrequent examination, for example during the worm off-season from May to October, with resultant loss of prowess and (iii) infrequent replacement of the FAMACHA© card, the colours of which are prone to fade with age or if exposed to direct sunlight for prolonged periods.

Although it is not clear which of the above reasons or combinations of reasons are most likely to be responsible for the observed misclassification, it is clear that the observed median haematocrit values for FAMACHA© categories 1–4 were consistently lower than the expected values (Table 3.5), and that for all but FAMACHA© category 5, only a small fraction of the observed haematocrit values fell within the expected range (Table 3.6). Of all sheep represented, only 98 individuals (14.5 %) were truly in FAMACHA© category 1, leading to the conclusion that the flock was always more anaemic than what was being indicated by FAMACHA©. One possible reason for the low numbers of "healthy" sheep in FAMACHA© category 1 could be that the farmer, even during the peak of the worm season, averaged 21 days between FAMACHA© evaluations, while intervals of seven days are prescribed at the peak of the worm season. This probably resulted in the flock being much more anaemic than he was actually aware of, since the cumulative effect of worm challenge was being masked by FAMACHA© misclassification. The occurrence of a truly non-anaemic sheep in FAMACHA© category 1 during a sample would thus have been a rare event. However, as indicated in Table 3.5, it is evident that if consideration is given to the fact that these were the results of clinical evaluation compared to the laboratory determined haematocrit values over a period of five years, the percentage of deviation from the median values per FAMACHA© category of 1–4 were within the relatively narrow range of 22–26 % (Table 3.5). This indicates that although the FAMACHA© evaluations were relatively constant over the five years, they were at too low a haematocrit level throughout. The sole exception was FAMACHA© category 5, but there were only 3 sheep in this category.

The consistency of the FAMACHA© evaluation on Farm 1 was further supported by Best Linear Unbiased Prediction (BLUP) heritability analysis performed on the data collected at the height of the worm challenge during the FAMACHA© trials on the farm, made possible by the complete genealogy data that were available for the sheep in the trials (Van Wyk & Bath 2002). Every year over the trial period, almost identical heritabilities were recorded for both FAMACHA© score and haematocrit, every time at a level slightly higher than that of the heritability of the faecal worm egg counts done at the same time (Van Wyk & Bath 2002).

Albers *et al*. (1987) reported that host resistance to *H. contortus* infection as measured on the basis of faecal worm egg counts and haematocrit is a moderately heritable trait, and Barger & Dash (1987) demonstrated that, when individuals are evaluated for faecal worm egg counts and haematocrit, the same individuals tend to have the lowest haematocrit and the highest faecal worm egg counts at each evaluation. It thus seems likely that the consistent differences between the clinical FAMACHA© test and its associated haematocrit values could have been rectified by re-training evaluators at an early stage, had this been detected early enough. It is an indication that the ideal would be to evaluate the success of the FAMACHA© evaluation when a person has been applying the system for a few months after the initial training. Furthermore, it emphasizes the necessity of at least basic training of FAMACHA© evaluation and supports the decision not to allow dispersal of the FAMACHA© system without adequate training (Van Wyk & Bath 2002).

The most important finding of this study for Farm 1 is that when dosing only FAMACHA© categories 3, 4 and 5, sensitivity was highest with a haematocrit cut-off of ≤15 % (Table 3.2), and that even then it was only 83 %. A better sensitivity would have resulted if FAMACHA© categories 2, 3, 4 and 5 were treated, with a haematocrit cut-off of ≤19 %, because a sheep with a haematocrit of this value is not in immediate danger of dying unless conditions of severe pasture contamination or nutritional challenge are present. Although Kaplan *et al*. (2004) do not discuss the issue of misclassification, it would appear from their results that their observed median haematocrit values after evaluation of 847 sheep were considerably higher than assigned median values, as evidenced by box and whisper plots demonstrating the relationship between haematocrit value and FAMACHA© scores in sheep. However, data from their study was collected from a total of 39 farms in the southern United States, and involved a large number of different evaluators as well as different breeds and ages of sheep. This is in contrast to the results of the present study on Farm 1 over a five-year period, where animals were scored by the same person, and where observed median haematocrit values were lower than expected (Table 3.5). Since validation trials are continuing on the farm, it is also imperative that as a first step to correcting misclassification, the farmer is at least informed that FAMACHA© category 2 should be included in the drench as well, until the error can be rectified. Calibration of the FAMACHA© scoring procedure on the farm should then be carried out to point out anomalies in his classification process, and re-familiarization with FAMACHA© should be carried out.

Farm 2

The results for Farm 2, where only sheep scored as FAMACHA© 4 or 5, or if their haematocrits were ≤15 % were treated, indicated that application of the FAMACHA© scoring process was more accurate than on Farm 1 (Table 3.9). These sheep were scored mainly by one investigator, with the exception of the first three evaluations in the first year of trials, when FAMACHA© classifications were the combined observations of himself and 1–3 other persons. The lowest accuracy of FAMACHA© classification was obtained for FAMACHA© category 3 on this farm, where 39 % of sheep that were scored into FAMACHA© category 3 had haematocrit values in the assigned range of 18–22 % (Table 3.10), compared to 78 % for FAMACHA© category 1, and the 40 % that were correct for FAMACHA© category 2. A relatively high proportion of sheep scored as being in FAMACHA© category 4 (57 %) was correctly classified compared to Farm 1 (Table 3.6). On Farm 2, FAMACHA© category 5 was not represented in any of the samples. A factor which may have played a role in comparing the two farms is that of all the sheep sampled for haematocrit determination in addition to FAMACHA© scoring on Farm 2, 401 individuals (50 %) were truly in FAMACHA© category 1, with a haematocrit of ≥28 %, compared to only 98 individuals, or 14.5 % on Farm 1. The general level of anaemia was thus lower for sheep on Farm 2 than for Farm 1, as evidenced by these figures. This could further indicate that the higher accuracy of FAMACHA© classification, in addition to much more regular examination of the flock, was the reason that sheep in FAMACHA© category 5 were not encountered on this farm. Epidemiological differences between the two farms, however, would have been important in their own right. Salvage treatments, where blanket drenching of all sheep in a sample was undertaken, was not required on Farm 2 as was the case on Farm 1, despite the fact that a much lower threshold of treatment, i.e. a haematocrit of ≤15 %, was used on Farm 2. Sensitivity on Farm 2 for a haematocrit cut-off of ≤19 % was 80 % if sheep in FAMACHA© categories 3–5 were considered to be test positive (Table 3.8), which represents an improvement of 22 % (i.e. 80 % - 58 %) over the sensitivity obtained on Farm 1 (Table 3.2) for the same set of parameters. Under these conditions, a total of only 21 % of the flock would have been treated if all sheep in FAMACHA© categories 3–5 were treated on Farm 2. If all animals in FAMACHA© 2 were also regarded as diseased, then sensitivity would have increased to 98 % for a haematocrit cut-off of ≤19 %, but specificity would have been low at 52 %, and still only 55 % of the flock would have been treated due to the perpetually high proportion of the flock in the “healthy” FAMACHA© categories 1 and 2.

Since there were no sheep in FAMACHA© category 5 on Farm 2, and also because of the much lower prevalence of disease for equivalent cut-off values and proportions of animals considered to be diseased, a general recommendation for Farm 2 to treat only sheep in FAMACHA© categories 3–5 would have allowed a high level of safety from overwhelming haemonchosis, while still leaving a large proportion of the flock untreated. If this had been done, it is likely that the labour inputs required for FAMACHA© application could have been reduced by enabling increased intervals between evaluations. The recommendation made for Farm 1, in contrast, was that all animals in FAMACHA© categories 2–5 should be treated, and if this drenching regime had been applied on Farm 2, considerable numbers of false positive sheep would have been unnecessarily have been drenched.

3.5 Conclusion

The present results suggest that, as long as the sensitivity of the diagnosis is high enough to avoid non-treatment of a proportion of truly anaemic sheep, production losses should be minimised. This is important, as with the FAMACHA© system, non-treatment of a false negative animal could lead to death, whereas it is acceptable to treat false positive sheep, as long as a considerable proportion of the flock is left untreated (Van Wyk 2001, 2002). The fact that FAMACHA© has a resolution of five different categories, allows wide scope to adjust the sensitivity of diagnosis, and as seen in this study on Farm 1, immediate corrective action can be implemented by simply adjusting the treatment to include the “next up” FAMACHA© category of sheep, without necessarily leading to “excessive drenching” as regards the sustainability of the worm management programme. Correct classification is preferable to corrective action, but the implication is that calibration should take place at least annually on farms where the FAMACHA© system is in use.

The present analyses add further confirmation to previous inputs into validation of FAMACHA© as part of the present paradigm towards employment of targeted selective treatment for sustainable helminth control, as reviewed by Van Wyk & Bath (2002). Similar analyses to those reported here have been conducted by Vatta *et al.* (2001) and Kaplan *et al*. (2004), and all have demonstrated the practicability of on-farm application of FAMACHA© by farmers, without the need for routine laboratory intervention. The results of this study suggest that (i) the sensitivity of the FAMACHA© diagnostic system should be evaluated at more regular intervals to avoid production losses due to misclassification bias; (ii) that calibration of the FAMACHA© scoring process in terms of training is essential, and (iii) that

animals should be examined at least weekly during periods of the highest worm challenge as with previous recommendations (Van Wyk & Bath 2002).