Veterinary Parasitology xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Veterinary Parasitology

journal homepage: www.elsevier.com/locate/vetpar

Validation of the FAMACHA© eye colour chart using sensitivity/specificity analysis on two South African sheep farms

D.P. Reynecke [a,b], J.A. van Wyk [a,*], B. Gummow [a,c], P. Dorny [d], J. Boomker [a]

[a] *Department of Veterinary Tropical Diseases and Production Animal Studies, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, 0110 Onderstepoort, South Africa*
[b] *AgResearch Ltd, Hopkirk Research Institute, AgResearch Grasslands, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand*
[c] *School of Veterinary and Biomedical Sciences, James Cook University, Townsville 4811, Queensland, Australia*
[d] *Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerpen, Belgium*

ARTICLE INFO

Article history:
Received 25 October 2007
Received in revised form 20 July 2009
Accepted 26 August 2009

Keywords:
Haemonchus
Sensitivity/specificity
Sheep
Targeted selective treatment
FAMACHA© validation
Integrated parasite management

ABSTRACT

A validation study of the FAMACHA© system for clinical evaluation of anaemia due to *Haemonchus contortus* was conducted on two commercial sheep farms in the summer rainfall region of South Africa. In this region, the *Haemonchus* season lasts from October to April. On Farm 1 the system was tested over a period of five successive years in consecutive sets of young stud Merino replacement rams and ewes examined at intervals of 3–5 weeks over each *Haemonchus* season, under routine farming conditions. When FAMACHA© scores of 3, 4, and 5 and haematocrit values of ≤22%, ≤19%, and ≤15% were separately considered to be anaemic, sensitivity on Farm 1 ranged from a maximum of 83% for a haematocrit cut-off of ≤15%, to 40% for a haematocrit cut-off of ≤22%. Sensitivity increased to 93% when FAMACHA© scores of 2, 3, 4, and 5 were considered anaemic at a cut-off value of ≤19%, but the positive predictive value decreased to 0.43, indicating that many non-anaemic animals would be treated. The analysis indicated a high level of classification bias on Farm 1, with the animals consistently being classified one FAMACHA© category lower (i.e. less anaemic) than reality.

On Farm 2 the test was conducted over two successive years in yearling rams evaluated at weekly to fortnightly intervals during each worm season. Every ram judged to be in FAMACHA© category 4 or 5 was bled for haematocrit determination, and it was only dewormed with effective anthelmintics if the haematocrit was 15% or lower. When FAMACHA© scores of 3, 4, and 5 and haematocrit values of ≤22% and ≤19% were separately considered to be anaemic on Farm 2, sensitivity ranged from 64% for a haematocrit cut-off of ≤22%, to 80% for a cut-off of ≤19%.

For identical haematocrit cut-off values and proportions of the sampled flock considered to be diseased as for Farm 1, sensitivity was always higher for Farm 2. On the other hand, further analysis of the data indicated that the magnitude of the error on Farm 1 was very consistent on average over the entire trial period.

The results of this study indicate that (i) persons introduced to the system should not only be trained, but also be evaluated for accuracy of application; (ii) the sensitivity of the FAMACHA© diagnostic system should ideally be evaluated at shorter intervals to avoid production losses due to failure to detect anaemic animals which may be at risk of death; (iii) that calibration of the FAMACHA© scoring is essential per individual evaluator, and (iv) that animals should be examined at weekly intervals during periods of the highest worm challenge.

* Corresponding author. Tel.: +27 12 529 8380; fax: +27 12 529 8312.
E-mail address: jan.vanwyk@up.ac.za (J.A. van Wyk).

0304-4017/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.vetpar.2009.08.023

1. Introduction

Multiple anthelmintic resistance of 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 resource-poor communal farms in the country (Van Wyk et al., 1999; Vatta and Lindberg, 2006). Numerous populations of this parasite have developed resistance to every one of the previously effective anthelmintics in South Africa (Van Wyk et al., 1997, 1999), to the extent that on some farms even combinations are showing low levels of efficacy. The focus on management of resistance in the country has been the development of integrated methods of worm management aimed at modified use of anthelmintics, notably with the use of sustainable targeted selective treatment (*s*TST) for reducing the rate of selection for worm resistance. In addition, Van Wyk (2003, 2006) and Van Wyk et al. (2006) suggested that the present impasse with respect to anthelmintic resistance management could be overcome with dedicated drenching decision-support software.

1.1. Evaluation of the FAMACHA© system

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 systems of *s*TST, notably FAMACHA©, were developed. Although much has been done to validate the FAMACHA© system (Bath et al., 2001; Vatta et al., 2001; Van Wyk and Bath, 2002; Kaplan et al., 2004), particularly in South Africa, it is still important that the method be tested on an ongoing basis for its operating characteristics, amongst others, sensitivity, specificity and predictive values under farming conditions. Management of haemonchosis in a flock with the use of this system depends on accurate identification and treatment of diseased individuals in a given flock, while the rest are left untreated. The FAMACHA© system has in this respect been successfully used as a colour-based stratification method, with five colour categories from bright red (probably normal) to pale (probably anaemic) as an indication of the anaemia status of individual animals in relation to the haematocrit "Gold Standard" (Bath et al., 2001).

1.2. Merit of the FAMACHA© system as a diagnostic test

The FAMACHA© system has been shown to reduce the uncertainty about the state of haemonchosis in individual sheep, and satisfies the requirements of a diagnostic test (Greiner and Gardner, 2000). The test is based on the principle that the colour of mucous membranes is correlated with the anaemia status of an animal (Riley and Van Wyk, 2009), and that it therefore reflects the haematocrit. The clinical performance of a diagnostic test depends on its diagnostic accuracy, which represents the ability of the test to correctly classify test subjects into clinically relevant subgroups (Zweig and Campbell, 1993). However, as the FAMACHA© system is based on a rating method (Hanley and McNeil, 1982), with the different FAMACHA© categories from 1 (bright red) to 5 (pale) 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, with one group representing animals requiring treatment, and the other representing animals that will not be treated at a given evaluation. This is an artificial distinction, since variability in grazing, immunity, etc., results in extra-binomial variation, necessitating the use of a negative binomial rather than a Poisson model. This situation is reflected in the fact that almost all animals are infected, but due to overdispersion, the minority of the animals harbour the highest individual numbers of worms (Barger, 1985; Wilson et al., 1996; Herbert and 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 allow different views of the infection status of a flock and thus allow upward or downward adjustment of categories of animals to treat or leave untreated.

1.3. Haematocrit cut-offs of 2 × 2 table method

One limitation to the 2 × 2 table method (Thrusfield, 2001) of estimating sensitivity and specificity is that there is usually a single predetermined criterion, referred to as a cut-off point, to indicate a true positive test result (Linden, 2006). This limitation has to some extent been addressed by previous investigators (Vatta et al., 2001; Sotomaior et al., 2003a,b; Kaplan et al., 2004) as well as in the present work by evaluating several haematocrit cut-off points against different FAMACHA© infection thresholds. The haematocrit cut-off value of $\leq$22% was selected as it is the upper haematocrit limit of FAMACHA© category 3 (Bath et al., 2001), and would thus include treatment of FAMACHA© categories 4 and 5, in addition to FAMACHA© category 3. However, a cut-off value $\leq$19% was also included to provide an additional view of the data, since an animal with a haematocrit of $\leq$19% could be in grave risk of dying within 5–7 days, if not detected and treated; severe challenge from *H. contortus* could cause some animals to lose up to seven percentage points of their haematocrit in 7 days (Malan et al., 2001). A further reason for selecting $\leq$19% as a haematocrit cut-off was to compare our results with similar studies such as those by Vatta et al. (2001) and Vatta et al. (2002) in South Africa, who evaluated $\leq$19% as a cut-off for goats farmed by resource-poor farmers in South Africa, and Kaplan et al. (2004) in the southern United States, who evaluated haematocrit cut-offs of $\leq$19% and $\leq$15% in sheep and goats. A haematocrit cut-off of $\leq$15% was thus also evaluated in the present work, the aim of which was to further validate the FAMACHA© system as implemented on two commercial sheep farms in South Africa by determining the diagnostic sensitivity, specificity and predictive value of the system at differing haematocrit cut-off points, as well as at differing FAMACHA© category treatment thresholds.

Please cite this article in press as: Reynecke, D.P., et al., Validation of the FAMACHA© eye colour chart using sensitivity/specificity analysis on two South African sheep farms. Vet. Parasitol. (2009), doi:10.1016/j.vetpar.2009.08.023

This investigation was part of a series aimed at testing of further methods of application of the FAMACHA© system, particularly for deciding on levels of drenching in relation to accuracy of application of the system on one hand, and the level of worm challenge on the other.

2. Materials and methods

2.1. Parasitological methods

Faecal worm egg counts (FECs) were carried out using a sucrose flotation modification of the McMaster method (Reinecke, 1983), at a sensitivity of 100 eggs per gramme of faeces. Work over a period of more than 6 years, including post-mortem examination of sheep on Farm 1 indicated that *H. contortus* was overwhelmingly preponderant during the "*Haemonchus*" season, which annually spans more or less from November to the following April or May according to reigning climatic conditions, while *Trichostrongylus colubriformis* and *Teladorsagia circumcincta* were prevalent in late autumn and winter (April to September), and it was only over that time period that dags (soiled breech from diarrhoea) was a problem. On Farm 2 practically only *H. contortus* were recovered during the course of the 2 years of the trial. On both farms "diagnostic treatment" (Gordon, 1981) also confirmed the importance of *H. contortus* in that cases of anaemia promptly recovered after anthelmintic treatment.

2.1.1. Origin of data and FAMACHA© test procedures

A validation study of the FAMACHA© system was conducted by various workers on a variety of commercial sheep farms in South Africa (Van Wyk and Bath, 2002). Two of these farms, for which extensive data sets were generated from naturally infected sheep, were selected for the present investigations. The farms are situated in the summer rainfall region of South Africa. Climatically, the region is a part of the temperate eastern plateau, at an altitude of approximately 1500 m above sea-level, with cool, rainy summers and cold, dry winters.

Before the trials were initiated, the evaluators on both farms were trained in practice by having to evaluate a class of sheep with a variety of haematocrit levels, after the system had been explained and the trainees had been given hands-on practice in the correct method of opening and judging the colours of the eyes of the sheep.

2.1.2. Farm 1

Over a period of five *Haemonchus* seasons, the FAMACHA© system was tested in stud Merino sheep under routine farming conditions on this farm, situated on the escarpment, east of the town Ermelo in Mpumalanga Province of South Africa. The colours of the conjunctivae of sheep were scored throughout by the same person (who had been trained beforehand—Van Wyk et al., 2001), on a 1–5 scale using the FAMACHA© chart (Bath et al., 2001). In addition, blood samples were periodically collected from each animal for haematocrit determination, and only sheep that were classified into FAMACHA© categories 3, 4 and 5 were treated with anthelmintics (Table 1). From this data set, five separate sets of FAMACHA© evaluations over successive *Haemonchus* seasons were available for analysis, where haematocrit determinations were carried out at the same time (Table 1).

Table 1
Criteria for anthelmintic treatment, number of FAMACHA© evaluations and number of haematocrit determinations for the two farms in the analysis.

Farm	Criteria for treatment	Number of FAMACHA© evaluations	Number of haematocrit determinations
Farm 1	Only treated if scored into FAMACHA© 3, 4 or 5	2000: 263 2001: 160 2002: 127 2003: 125 Total = 675	675
Farm 2	Only treated if scored into FAMACHA© 4 or 5 and haematocrit ≤15%	2000: 132 2001: 319 2002: 355 Total = 806	806

At the start of each *Haemonchus* season over the period of study a new set of two classes of sheep was introduced into the 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. Approximately 130–200 sheep of each class out of a total of 1500 on the farm, were sampled at each FAMACHA© evaluation in the various trials per class and worm season. At the start of each of the five annual trials (usually October of each year), each sheep was scored into a FAMACHA© and body condition score (BCS) category, its body weight 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 detected by a substantial increase in up to 25% of animals scored into the high-risk FAMACHA© categories 3, 4 and 5, usually in January or February of each year, all sheep were again dewormed.[1] Then, until the end of each trial in April, only the animals in FAMACHA© categories 3–5 were treated as before. From October/November to the following April, sheep were mostly evaluated at intervals of 3–4 weeks, but in some instances the evaluation intervals were longer, at up to 5 weeks. A total of 7–11 sampling events occurred per worm season.

Data for both FAMACHA© scores and haematocrits 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. Secondly, FAMACHA© eye scores of 2, 3, 4 and 5

[1] These young stud sheep were primarily being subjected to evaluation of the breeding values for production, including wool characteristics (see Riley and Van Wyk, 2009). Hence this drenching occasion signified the end of the investigation of Sensitivity / Specificity analysis for evaluation of the accuracy of the FAMACHA© system.

and haematocrit values of ≤22%, ≤19%, and ≤15% were separately considered to be anaemic. The data therefore consisted of anaemia status as evaluated by FAMACHA© score, and haematocrit values, originating from naturally infected sheep.

2.1.3. Farm 2

A second validation study was conducted over a period of two *Haemonchus* seasons on data from Merino rams on Farm 2, a commercial sheep farm situated in eastern Free State Province. The rams, approximately 6 months of age at the start of each trial, had been selected from a variety of farms and brought together for comparison under the so-called Veld Ram Club system (Van Wyk et al., 1991; Bisset et al., 2001). Over a period of 10–11 months per trial, the rams were run on common pasture, ranked per production parameter and the best doers were auctioned at the end of the period, while the rest were culled.

As in the case of Farm 1, a single person was responsible for all the clinical evaluations on Farm 2, excepting for the initial three evaluations at the start of the investigations, where this person did the evaluations together with three others. Haematocrit determinations were done on all the rams on five separate occasions over the course of the two consecutive *Haemonchus* seasons, both at the beginning and at the height of the season (Table 1). In addition, each ram was clinically examined using the FAMACHA© system at intervals of 7–14 days during each trial. If judged to be in FAMACHA© category 4 or 5, it was bled for haematocrit determination, and drenched with effective anthelmintics only if its haematocrit value was 15% or lower. This was the principal difference between the data originating from the two farms—while sheep evaluated to be in FAMACHA© categories 3–5 were treated on Farm 1, without further testing, the rams on Farm 2 were treated only if FAMACHA© category 4 or 5 clinical evaluation was confirmed by haematocrit determination. After initial fortnightly clinical evaluation on Farm 2, the frequency was increased to weekly evaluation for the months of January and February in the first *Haemonchus* season on Farm 2, when worm challenge became intensive. During the second trial season there was no need for more frequent evaluation, as lower levels of infection were experienced.

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 for determining the 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 median and 5th and 95th percentile 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 purposes of determining test sensitivity and specificity on Farm 1, three different haematocrit cut-off values were considered to be anaemic, namely ≤22%, ≤19% and ≤15%, and the above test parameters were calculated separately for these three values. True positives were defined as sheep with haematocrits of ≤22%, ≤19%, or ≤15% and FAMACHA© scores of 3, 4 or 5; false positives as those with haematocrits of >22%, >19%, or >15% but with FAMACHA© scores of 3, 4 or 5; false negatives as sheep with haematocrits ≤22%, ≤19%, or ≤15%, but FAMACHA© scored as 1 or 2; and true negatives as sheep that were not anaemic according to the above haematocrit cut-offs, 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 section of the analysis, haematocrit values were considered anaemic if ≤22% or ≤19%, but due to the inclusion of FAMACHA© category 2 as test positive the haematocrit cut-off of ≤15% was not included in this part of the analysis. Test operating characteristics were calculated as described above. Prevalence of disease was calculated for all haematocrit cut-offs, and confidence limits were calculated for true prevalence.

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 scored into FAMACHA© categories 3–5 were considered to be test positive for comparison between the two farms.

3. Results

3.1. Farm 1

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 (i.e. including true positives that were treated and true negatives that were correctly left untreated). The sensitivity of the FAMACHA© system for identifying sheep that were regarded as

Table 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)

Table 3
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 4
Farm 1. 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	Nil[a]	8.6	11.5

[a] Only three sheep were recorded in this category.

Table 5
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)

anaemic with the above haematocrit cut-off values and FAMACHA© categories treated, was low for all haematocrit cut-off values, with the highest sensitivity (83%) being obtained for a cut-off of ≤15% (Table 2). Specificity was highest for a haematocrit cut-off value of ≤22%, at 96% (Table 2). Sensitivity increased as the haematocrit cut-off value decreased, but the predictive value of a positive also decreased. Thus, treating FAMACHA© categories 3–5 inclusive (with category 3 as a threshold), at a haematocrit cut-off of ≤22%, would have resulted in only 40% of animals that were defined as anaemic being 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%, but only 58% of sheep with a haematocrit of ≤19% would have been detected (Table 2), since the majority of animals correctly left untreated would have been true negatives.

In contrast to the above, when FAMACHA© scores of 2–5 (inclusive), and haematocrit cut-off values of ≤22% and ≤19% were considered anaemic, sensitivity was highest when a haematocrit value of ≤19% was considered anaemic, at 93%, while sensitivity for a cut-off value of ≤22% increased from 40% to 83% (Table 3). 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. The total percentage of correctly treated animals, i.e. true positives + true negatives, would have been 64%, but this would have been due to a relatively high proportion (34%) of false positives and a false negative percentage of only 1.9%.

The FAMACHA© scores versus assigned and observed median haematocrit values are summarised in Table 4. Observed median haematocrit values were lower than assigned median values, indicating misclassification on the part of the evaluator (Table 4). For example, the assigned minimum haematocrit value of FAMACHA© category 1 is 28%, but the observed median for all animals scored into this category was 23%. Similarly, the assigned median haematocrit value of FAMACHA© category 2 (range 23–27%) is 25%, yet a relatively low median value of 19.5% was observed. The assigned and observed median haematocrit values for FAMACHA© category 3 are 20% and 15%, and those for category 4 are 15% and 11%, respectively.

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 each category. 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 this latter category. There was therefore an increase in the accuracy of FAMACHA© classification from FAMACHA© category 1 to category 5 on Farm 1.

3.2. Farm 2

For a positive diagnosis of anaemia on Farm 2, i.e. sheep scored into FAMACHA© categories 3–5, 86% of sheep would have been correctly treated at a haematocrit cut-off of ≤22%, while 88% would have been correctly treated at a haematocrit cut-off of ≤19%. Positive predictive value was highest for a haematocrit cut-off of ≤22%, at 77% (Table 5), but prevalence was only 25%, compared to a prevalence of

Table 6
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	Nil	Nil	Nil

49% at this cut-off and treatment threshold on Farm 1. Sensitivity, specificity, positive predictive value, negative predictive value and prevalence values are listed in Table 5. FAMACHA© scores versus assigned and observed median haematocrit values for Farm 2 are reported in Table 6, and it is clear that scoring of animals into relevant categories was more accurate on Farm 2 than was achieved on Farm 1. Sensitivity was highest for a cut-off of ≤19% at 80%, while specificity was highest for a cut-off of ≤22% at 93% (Table 5). The observed median haematocrit values were much more consistent with their assigned values than was the case on Farm 1 (Table 6).

4. Discussion

An advantage of the FAMACHA© system is that it comprises five different haematocrit ranges, from the healthy, non-anaemic, to severely anaemic. Hence it is easy to adjust the treatment threshold, which need not be rigid, but can be adjusted stepwise over the course of any given worm season to allow for the level of worm challenge at any given time. From an initial appraisal of the data, it was clear that application of the FAMACHA© scoring process on Farm 1 was effective at discriminating between diseased an non-diseased individuals for given selection criteria, but that application of the system was sub-optimal compared to that achieved on Farm 2. This was, in addition to the fact that there were no animals scored into category 5 on Farm 2, the main reason that a more detailed analysis of test operating characteristics was undertaken for Farm 1.

4.1. Farm 1

4.1.1. Farm 1 FAMACHA© scoring data

Despite the relative inaccuracy of FAMACHA© scoring on Farm 1, a maximum of 49.5% of the animals would have been treated at the most conservative haematocrit cut-off of ≤22%, and this would have included the 8.3% of animals that were false negatives for a FAMACHA© treatment threshold of category 2. As regards both reduced selection for anthelmintic resistance and use of chemicals, this compares very favourably with blanket treatment systems, where all animals are commonly treated both before and repeatedly during a given worm season.

In this series of trials on Farm 1, only sheep scored into FAMACHA© categories 3, 4 and 5 were treated, apart from the blanket drenching events described, and when a haematocrit of ≤19% was used as a cut-off, only 58% of sheep that were anaemic would have been detected (Table 2). If a lower haematocrit cut-off of ≤15% was selected, 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 have been in danger of succumbing to haemonchosis. It has been shown that the haematocrit of a sheep could drop up to seven percentage points in as many days (Malan et al., 2001), with the implication that even an animal with a relatively mild level of anaemia at 19% haematocrit, could be at risk of death within a week. For this reason, a haematocrit cut-off of ≤15% would be unrealistic for Farm 1 in the present case and a haematocrit cut-off value of ≤19% would carry less risk. 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). This represents a dramatic improvement over the actual situation where only 58% of anaemic sheep with a haematocrit of ≤19% were detected and treated for a treatment threshold of FAMACHA© category 3. Even though 33.6% of the sheep would have been treated as false positives if FAMACHA© categories 2–5 were included, the total proportion of the animals recommended for treatment would still have comprised a maximum of only 59% of the flock. Such a level of treatment would almost certainly maintain a sufficient level of refugia (Van Wyk, 2001) for large-scale reduction in selection for anthelmintic resistance, while maintaining an acceptable level of parasite management for the producer.

4.1.2. Farm 1 misclassification bias

The results from Farm 1 indicate that misclassification of animals into relevant categories occurred on this farm (Table 4). For instance, of all sheep represented, only 98 individuals (14.5%) were truly in FAMACHA© category 1 (i.e. all individuals with a haematocrit >28% in all FAMACHA© categories), and even then, these sheep were spread over several sampling events, leading to the conclusion that the flock was always more anaemic than what was being indicated by clinical FAMACHA© evaluation. However, an important consideration is that the percentages of error between the assigned median haematocrit values for each FAMACHA© category and the observed median haematocrit values for FAMACHA© categories 1–4 varied within the narrow margins of only 22–26% below the assigned values (Table 4). This indicates a high level of consistency in the deviations, which were very constant over the five *Haemonchus* seasons, even though at too low a haematocrit level throughout, with only a small fraction of the clinical classifications falling within the assigned haematocrit ranges. The sole exception was FAMACHA© category 5, but there were only three sheep in this category throughout the trial period.

The low numbers of truly "healthy" sheep in FAMACHA© category 1 possibly resulted from the fact that the farmer, even during the peak of the worm season, averaged 21 days between FAMACHA© evaluations, while intervals of 7 days are prescribed at the peak of the worm season in the months of January and February (Van Wyk and Bath, 2002). This resulted in the flock being much more anaemic than the farmer concerned was aware of, since the cumulative effect of morbidity from worm challenge was being masked by FAMACHA© misclassification.

4.1.3. Heritability of FAMACHA© evaluation on Farm 1

The consistency of the FAMACHA© evaluation on Farm 1 was very strongly 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 (Bisset et al., 2001; Riley and Van Wyk, 2009). For every worm season over the trial period reported on by these authors, there were very highly significant levels of genetic and phenotypic correlation and similar levels of heritability were found between FAMACHA© score and haematocrit, as well as of log FECs done on faeces collected at the same time. The genetic correlation between FAMACHA© values and those of its gold standard, haematocrit, was close to unity (Bisset et al., 2001; Riley and Van Wyk, 2009).

Albers et al. (1987) reported that host resistance to *H. contortus* infection as measured by faecal worm egg count and haematocrit is a moderately heritable trait, and Barger and Dash (1987) demonstrated that when individuals are evaluated for FEC and haematocrit, the same individuals tend to have the lowest haematocrit and the highest FECs at each evaluation. It thus seems likely that the consistent differences between the clinical FAMACHA© test and the haematocrit ranges utilised as its gold standard could have been rectified by retraining of the evaluator at an early stage, had the classification bias been detected. 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 emphasises the necessity of at least basic training in FAMACHA© evaluation and supports the decision not to allow dispersal of the FAMACHA© system without adequate training (Van Wyk and Bath, 2002).

The most important finding of this study for Farm 1 is that when treating only FAMACHA© categories 3, 4 and 5, sensitivity was highest with a haematocrit cut-off of ≤15% (Table 2), and that even then it was at a level of 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 under conditions of severe worm and/or nutritional challenge. 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 five *Haemonchus* seasons, where animals were scored by the same person, and where observed median haematocrit values were lower than indicated by the evaluations by this person (Table 4). It is also imperative that as a first step to correcting misclassification, the farmer should, in such a situation, at least be informed that FAMACHA© category 2 should be included in the drenched group as well, until the error can be rectified. Calibration of the FAMACHA© scoring procedure on the farm should then be done to point out anomalies in his classification process, and he should be re-familiarised with FAMACHA© classification. On the other hand, because the misclassification on Farm 1 was very consistent, routine inclusion of category 2 for treatment could probably have solved the problem, without the need for retraining; in effect this would have brought him in line with current recommendation (Van Wyk and Bath, 2002).

4.2. Farm 2

4.2.1. Farm 2 clinical classification data

The results for Farm 2, where only sheep scored as FAMACHA© 4 or 5 were treated if their haematocrits were ≤15%, indicated that application of the FAMACHA© scoring process was more accurate than on Farm 1 (Table 6). As on Farm 1, these sheep were scored by only one investigator, with the exception of the first three evaluations in the first season, when FAMACHA© classifications were the combined observations of the farmer and 1–3 other persons. The lowest accuracy of FAMACHA© classification was obtained for FAMACHA© category 3 on this farm, with 39% of sheep being classified correctly into the assigned haematocrit range of 18–22%, compared to 78% for FAMACHA© category 1, and 40% for category 2. A relatively high proportion of sheep scored as being in FAMACHA© category 4 (57%) was correctly classified compared to the 44% for Farm 1. On Farm 2, FAMACHA© category 5 was not represented in any of the samples.

A factor which could have played an important role in the differences between the clinical evaluation results of the two farms is that of all the sheep sampled for haematocrit determination in addition to FAMACHA© scoring on Farm 2, 50% of individuals were truly in FAMACHA© category 1, with a haematocrit of ≥28%, compared to only 14.5% of those on Farm 1. The general level of anaemia was thus lower for sheep on Farm 2 than for Farm 1. While it should be kept in mind that farm management and pastures differed, the higher accuracy of FAMACHA© classification on Farm 2 probably contributed to more accurate detection and treatment of anaemic sheep, complemented by the fact that the sheep were evaluated more frequently, thus allowing earlier treatment and prevention of excessive levels of anaemia.

4.3. Comparison of Farms 1 and 2

Salvage treatments, with blanket anthelmintic treatment of all sheep, 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 FAMACHA© test cut point of 4 and 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 5), which is an improvement of 22% (i.e. 80% versus 58%) over the sensitivity obtained on Farm 1 (Table 2) for the same set of parameters. Under these criteria, a total of only 21% of the flock would have been treated on Farm 2 for a FAMACHA© cut point of 3. On the other hand, if all animals in FAMACHA© 2 were also to be treated on this second farm, then sensitivity would have increased to 98% for a haematocrit cut-off of ≤19%, although specificity would have been low at 52%, which would have led to more truly negative sheep being treated.

However, despite this fact, only 55% of the flock would have been treated in the latter case, due to the perpetually higher proportion of the flock in the "healthy" FAMACHA© categories 1 and 2 on Farm 2.

Because of the much lower prevalence of disease for equivalent cut-off values and proportions of animals considered to be diseased on Farm 2 (Tables 2, 3 and 5), 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. This would reduce the labour inputs required for FAMACHA© application by enabling increased intervals between evaluations. On the other hand, if the recommendation made for Farm 1 to treat all animals in FAMACHA© categories 2–5 were to be applied on Farm 2, considerable numbers of false positive sheep would have been drenched unnecessarily. 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 as defined by a selected haematocrit cut-off, 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 remains untreated (Van Wyk, 2001, 2002), or the treatment occurs at such a time and with such an anthelmintic formulation as to allow re-infection with worms in refugia, on pasture (Kenyon et al., 2009). 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 could be implemented by simply adjusting the treatment to include the "next up" FAMACHA© category of sheep, without this necessarily leading to "excessive drenching" as regards refugia and the sustainability of the worm management programme. It is also a further indication that results of evaluations carried out on different farms or under different circumstances could lead to erroneous interpretation of such data if this were pooled across regions or farms for analysis. We further suggest that this work will provide further insight into incorporating FAMACHA© data into a generalised decision-support model (Van Wyk and Reynecke, submitted for publication), which would include an estimation of test accuracy and validity for a given property as at least one of its sub-components.

The 2×2 method for estimating sensitivity and specificity is based on the diagnostic test's predetermined cut-off point. Thus Reynecke et al. (submitted for publication) investigated use of Receiver Operative Characteristic (ROC) plots (Greiner et al., 1995; Greiner and Gardner, 2000; Beck et al., 2005) for determining the overall accuracy of the FAMACHA© diagnostic test at each site in relation to the haematocrit gold standard. ROC plots have the advantage of optimising selection of FAMACHA© cut points, which comprise division into FAMACHA© categories to treat with anthelmintics or leave untreated. For instance, a cut point of 3 signifies treatment of categories 3–5, while categories 1 and 2 are left undrenched. This method proved to be valuable for complementing sensitivity/specificity analysis in order to lead evaluators as regards safe drenching practice in systems of *s*TST.

5. Conclusion

The present analyses add further confirmation to previous inputs into validation of FAMACHA© as part of the present paradigm towards employment of *s*TST for sustainable helminth management, as reviewed by Van Wyk and 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 shorter intervals to avoid losses due to misclassification bias, (ii) that regular calibration (i.e. determination of the relationship between FAMACHA© scoring and the haematocrit gold standard) of the FAMACHA© scoring process is essential, and (iii) that, in accordance with previous recommendations (Van Wyk and Bath, 2002), animals should be examined at least weekly during periods of the highest worm challenge, to reduce risk to very low levels. On the other hand, it is realised that this limits the applicability of the FAMACHA© system due to labour constraints, hence it is being addressed at present in further work aimed at development of dedicated software for obtaining an optimum balance between risk and low labour input.

Acknowledgements

This work was carried out within the Framework Agreement between the Directorate General for Development Cooperation, Belgium and the Institute of Tropical Medicine, Antwerpen, Belgium (Grant No. AG534) with the Department of Veterinary Tropical Diseases of the University of Pretoria Faculty of Veterinary Science, and with the support of the EU "PARASOL" FOOD-CT-2005-022851 FP6 project. Without the enthusiastic disregard of Mr. Johann Bührmann, owner of Farm 1, for the risks involved in evaluating such an unconventional new system as FAMACHA© under commercial farming conditions on his farm, this work could not have materialised. Mrs. Rachel Coetzee was responsible for the data capture.

References

Albers, G.A.A., Gray, G.D., Piper, L.R., Barker, J.S.F., Le Jambre, L.F., Barger, I.A., 1987. The genetics of resistance and resilience to *Haemonchus contortus* infection in young Merino sheep. Int. J. Parasitol. 17, 1355–1363.
Barger, I.A., 1985. The statistical distribution of trichostrongylid nematodes in grazing lambs. Int. J. Parasitol. 15, 645–649.
Barger, I.A., Dash, K.M., 1987. Repeatability of ovine faecal egg counts and blood packed cell volumes in *Haemonchus contortus* infections. Int. J. Parasitol. 17, 977–980.
Bath, G.F., Hansen, J.W., Krecek, R.C., Van Wyk, J.A., Vatta, A.F., 2001. Sustainable approaches for managing haemonchosis in sheep and goats. Food and Agriculture Organization (FAO) Animal Production and Health Technical Co-operation Project No. TCP/SAF/8821 (A).
Beck, R., Gašpar, A., Mihaljević, Ž., Marinculić, A., Stojčević, D., Brstilo, M., 2005. Evaluation of ELISA for detection of *Trichinella* antibodies in muscle juice samples of naturally infected pigs. Vet. Parasitol. 132, 91–95.
Begg, C., 1987. Biases in the assessment of diagnostic tests. Stat. Med. 6, 411–423.

Bisset, S.A., Van Wyk, J.A., Morris, C.A., Stenson, M.O., Malan, F.S., 2001. Phenotypic and genetic relationships amongst FAMACHA© score, faecal egg count and performance data in Merino sheep exposed to *Haemonchus contortus* infection in South Africa. In: Proceedings of the 5th International Sheep Veterinary Congress. South Africa (CD, thus no page numbers).
Gordon, H. McL., 1981. Epidemiology of helminthosis of sheep: diagnosis. Refresher course for veterinarians. Proceedings no. 58. Refresher course for sheep. University of Sydney (The Post-Graduate Committee in Veterinary Science), pp. 607–615.
Greiner, M., Sohr, D., Göbel, P., 1995. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. J. Immunol. Methods 185, 123–132.
Greiner, M., Gardner, I.A., 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. Prev. Vet. Med. 45, 3–22.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.
Herbert, J., Isham, V., 2000. Stochastic host-parasite interaction models. J. Math. Biol. 40, 343–371.
Kaplan, R.M., Burke, J.M., Terrill, T.H., Miller, J.E., Getz, W.R., Mobini, S., Valencia, E., Williams, M.J., Williamson, L.H., Larsen, M., Vatta, A.F., 2004. Validation of the FAMACHA© eye color chart for detecting clinical anemia in sheep and goats on farms in the southern United States. Vet. Parasitol. 123, 105–120.
Kenyon, F., Greer, A.W., Coles, G.C., Cringoli, G., Papadopoulos, E., Cabaret, J., Berrag, B., Varady, M., Van Wyk, J.A., Thomas, E., Vercruysse, J., Jackson, F., 2009. The role of targeted selective treatments in the development of refugia-based approaches to the control of gastrointestinal nematodes of small ruminants. Vet. Parasitol. 164, 3–11.
Leathwick, D.M., Waghorn, T.S., Miller, C.M., Atkinson, D.S., Haack, N.A., Oliver, A.M., 2006. Selective and on-demand drenching of lambs: impact on parasite populations and performance of lambs. N. Z. Vet. J. 54, 305–312.
Linden, A., 2006. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J. Eval. Clin. Pract. 12, 132–139.
Malan, F.S., Van Wyk, J.A., Wessels, C.D., 2001. Clinical evaluation of anaemia in sheep: early trials. Onderstepoort J. Vet. Res. 68, 165–174.
Reinecke, R.K., 1983. Veterinary Helminthology. Butterworth Publishers, Durban.
Reynecke, D.P., Van Wyk, J.A., Gummow, B., Dorny, P., Boomker, J., submitted for publication. Use of receiver operating characteristic curves for selection of treatment thresholds using the FAMACHA© diagnostic system for anaemia in sheep. Vet. Parasitol.
Riley, D.G., Van Wyk, J.A., 2009. Genetic parameters for FAMACHA© score and related traits for host resistance/resilience and production at differing severities of worm challenge in a Merino flock in South Africa. Vet. Parasitol. 164, 44–52.
Sotomaior, C.S., Caldas, J.S., Iark, P., Benvenutti, T., Rodrigues, N., 2003a. Use of the FAMACHA© system in goats. In: Proceedings of the III Congreso Latino Americano de Especialistas en Pequenos Rumiantes y Camelidos Sulamericanos, Vino del Mar, Chile.
Sotomaior, C.S., Milczewski, V., Iark, P., Caldas, J.S., Benvenutti, T., Sillas, R., Schwartz, M.G., 2003b. Use of the FAMACHA© system in commercial sheep flocks. In: Proceedings of the III Congreso Latino Americano de Especialistas en Pequenos Rumiantes y Camelidos Sulamericanos, Vino del Mar, Chile.
Thrusfield, M., 2001. Veterinary Epidemiology, 2nd edition. Blackwell Science.
Van Wyk, J.A., 2001. Refugia—overlooked as perhaps the most potent factor concerning the development of anthelmintic resistance. Onderstepoort J. Vet. Res. 68, 55–67.
Van Wyk, J.A., 2002. Principles for the use of macrocyclic lactones to minimise selection for resistance. Aust. Vet. J. 80, 437–438.
Van Wyk, J.A., 2003. Think refugia or lose the battle against drug resistance. In: Proceedings of the 5th International Seminar Animal Parasitology, Mexico, pp. 39–47.
Van Wyk, J.A., 2006. Face facts: drenching with anthelmintics for worm control selects for worm resistance – and no excuses! Proc. N. Z. Soc. Anim. Prod. 4–13.
Van Wyk, J.A., Bath, G.F., 2002. The FAMACHA© system for managing haemonchosis in sheep and goats by clinically identifying individual animals for treatment. Vet. Res. 33, 509–529.
Van Wyk, J.A., Bath, G.F., Groeneveld, H.T., Stenson, M.O., Malan, F.S., 2001. Extensive testing of the FAMACHA© system for accuracy of clinical evaluation of anaemia caused by ovine haemonchosis. In: Proceedings of the 5th International Sheep Vet. Congr. South Africa (CD, thus no page numbers).
Van Wyk, J.A., Hoste, H., Kaplan, R.M., Besier, R.B., 2006. Targeted selective treatment for worm management—how do we sell rational programs to farmers? Vet. Parasitol. 139, 336–346.
Van Wyk, J.A., Malan, F.S., Randles, J.L., 1997. How long before resistance makes it impossible to control some field strains of *Haemonchus contortus* in South Africa with any of the anthelmintics? Vet. Parasitol. 70, 111–122.
Van Wyk, J.A., Reynecke, D.P., submitted for publication. Blueprint for an automated specific decision-support system for countering anthelmintic resistance at farm level. Vet. Parasitol.
Van Wyk, J.A., Stenson, M.O., Van der Merwe, J.S., Vorster, R.J., Viljoen, P.G., 1999. Anthelmintic resistance in South Africa: surveys indicate an extremely serious situation in sheep and goat farming. Onderstepoort J. Vet. Res. 66, 273–284.
Van Wyk, J.A., Van Schalkwyk, P.C., Bath, G.F., Gerber, H.M., Alves, R.M.R., 1991. Die gevaar van wye verspreiding van weerstandbiedendheid teen wurmmiddels deur veldramprestasietoetssentra/The threat of wide dissemination of anthelmintic resistance by veld ram performance testing units. J. S. Afr. Vet. Assoc. 62, 171–175.
Vatta, A.F., Krecek, R.C., Letty, B.A., Van Der Linde, M.J., Grimbeek, R.J., De Villiers, J.F., Motswatswe, P.W., Molebimang, G.S., Boshoff, H.M., Hanson, J.W., 2002. Incidence of *Haemonchus contortus* spp. and effect on haematocrit and eye colour in goats farmed under resource-poor conditions in South Africa. Vet. Parasitol. 103, 119–131.
Vatta, A.F., Letty, B.A., Van der Linde, M.J., Van Wijk, E.F., Hansen, J.W., Krecek, R.C., 2001. Testing for clinical anaemia caused by *Haemonchus* spp. in goats farmed under resource-poor conditions in South Africa using an eye color chart used for sheep. Vet. Parasitol. 99, 1–14.
Vatta, A.F., Lindberg, A.L.E., 2006. Managing anthelmintic resistance in small ruminant livestock of resource-poor farmers in South Africa. J. S. Afr. Vet. Assoc. 77, 2–8.
Wilson, K., Grenfell, B.T., Shaw, D.J., 1996. Analysis of aggregated parasite distributions: a comparison of methods. Funct. Ecol. 10, 592–601.
Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577.

Please cite this article in press as: Reynecke, D.P., et al., Validation of the FAMACHA© eye colour chart using sensitivity/specificity analysis on two South African sheep farms. Vet. Parasitol. (2009), doi:10.1016/j.vetpar.2009.08.023