*Onderstepoort J. vet. Res.* 57, 123–127 (1990)

# ON SAMPLING TICK POPULATIONS: THE PROBLEM OF OVERDISPERSION

T. N. PETNEY, H. VAN ARK and A. M. SPICKETT, Veterinary Research Institute, Onderstepoort 0110


## ABSTRACT

PETNEY, T. N., VAN ARK, H. & SPICKETT, A. M., 1990. On sampling tick populations: the problem of overdispersion. *Onderstepoort Journal of Veterinary Research,* 57, 123–127 (1990)

Data collected on both free-living and parasitic tick populations are likely to be overdispersed. The use of means from few replicate samples of overdispersed data as quantitative estimators of tick population density is in turn likely to lead to inaccurate interpretations which may be scientifically misleading. In this paper ways of estimating overdispersion are listed and suggestions for the use of correct statistical tests for handling overdispersed data are given.


## INTRODUCTION

The way in which parasites are dispersed within their host population is of considerable significance if an accurate, quantitative assessment of the parasites' pathological role is to be determined (Crofton, 1971; Anderson & May, 1985). There is considerable evidence available suggesting that ticks are overdispersed within host populations, meaning that some hosts have many ticks but most other hosts have only few ticks (Taylor, Woiwood & Perry, 1981; Petney & Fourie, 1989). In this paper we give additional information supporting this view and point out the pitfalls of interpretation and analysis of such data. Both non-parametric tests and parametric tests on suitably transformed overdispersed data can give an accurate picture of the significance of the data for sufficiently large samples sizes.

## DISPERSION PATTERN: THEORY

There are three basic dispersion patterns (Fig. 1). Underdispersion occurs when there is very little variation in the number of ticks on individual hosts, i.e. the ticks are evenly distributed between hosts. Ticks may also occur randomly on hosts in which situation the number of ticks on individual hosts clusters evenly about a mean value. Random distribution of arthropods are, however, very rare in nature (Taylor *et al.*, 1981). Lastly, ticks may be overdispersed, in which case most of the ticks occur on only a few hosts while the majority of hosts harbour only few ticks (Fig. 1 & 2). This is the most commonly occurring distribution for arthropods.

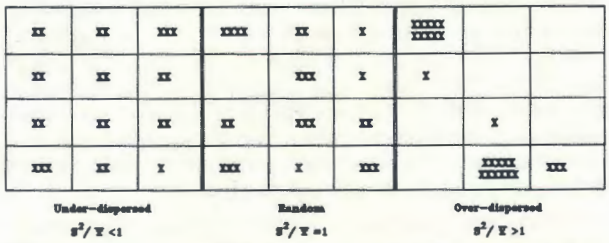


FIG. 1 The ways in which ticks may be distributed among drags or hosts. The ratio $S^2/x$ is a measure of dispersion pattern

The pattern of dispersion can be determined by comparing a sample's variance to its mean (Taylor, 1961) (Fig. 1). If the variance is approximately equal to the mean (i.e. the variance to mean ratio, $s^2/\bar{x}$, is approximately 1) then the pattern is random, conforming to a Poisson distribution. If the ratio is less than 1 the pattern is underdispersed and if it is greater than 1 the pattern is overdispersed. The significance of $s^2/\bar{x}$ for a number of sampling units or hosts (preferably >10) can be calculated using either a t-test:

$$\left( t_{n-1} = \frac{S^2/\bar{x} - 1}{\sqrt{2/n-1}} \right)$$

(Kershaw, 1973) or by using a $\chi^2$ test:

$$\left( \chi^2_{n-1} = \frac{S^2_{n-1}}{\bar{x}} \right)$$

(Elliot, 1977). It should be noted that the t-test will indicate overdispersion even when the latter is not very marked and could for biological purposes be regarded as random (Kershaw, 1973).

Should a substantial number of sampling units have few ticks (e.g <5), the variance to mean ratio will approximate zero even if the underlying distribution is an overdispersed one. For such a sample the interpretation of the ratio may be meaningless.

There are a variety of ways in which different dispersion patterns may be generated (Table 1). In the case of ticks the most likely reasons are the non-ran-

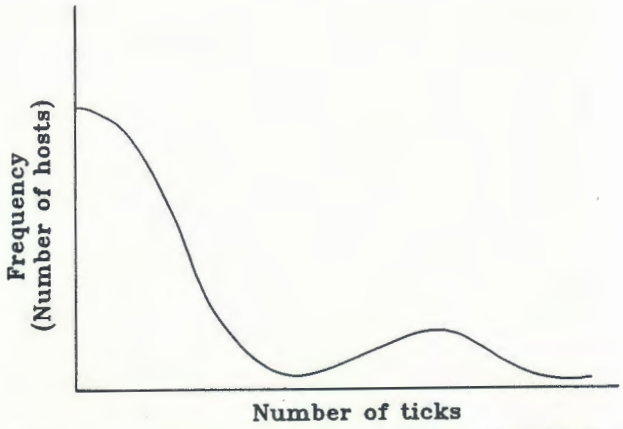


FIG. 2 Overdispersion occurs when a few drags or hosts harbour a high proportion of the ticks sampled

TABLE 1 Possible reasons for overdispersion

| | Reasons |
|---|---|
| 1 | Free living ticks are not randomly distributed with the host habitat |
| 2 | The presence of ticks on a particular host predisposes that host to attract more ticks |
| 3 | Heterogeneity in hosts ability to reduce or limit tick burden by immune response or other means |
| 4 | Sampling of heterogeneous subpopulations |

Received 14 December 1989 — Editor

TABLE 2 The numbers of larval ticks found on replicate drags from the Kruger National Park

| Data set | Replicates | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | $\bar{x}$ | $s^2/\bar{x}$ | $t_2$ | P |
| *A. hebraeum* | | | | | | | |
| 1 | 1 | 201 | 264 | 155,3 | 121,4 | 120,4 | <0,001 |
| 2 | 20 | 69 | 701 | 263,3 | 547,8 | 546,8 | <0,001 |
| 3 | 200 | 594 | 4329 | 1707,7 | 3040,6 | 3039,6 | <0,001 |
| *B. decoloratus* | | | | | | | |
| 1 | 3 | 6 | 52 | 20,3 | 37,1 | 36,1 | <0,001 |
| 2 | 8 | 10 | 74 | 33,3 | 38,0 | 37,0 | <0,001 |
| 3 | 1 | 4 | 331 | 112,0 | 321,2 | 320,2 | <0,001 |

dom pattern of dispersion of free-living stages and the differences in immune response between hosts.

*Dispersion of free-living larvae*

Female ixodid ticks lay their eggs in a single batch. When the larvae hatch there is little lateral dispersal (Rechav, 1979) hence in the field larvae tend to occur in clusters which may number many thousands depending on the species.

The usual way to sample free-living larvae is by drag sampling (Petney & Horak, 1987). On an individual drag it is therefore possible that an aggregation of larvae may or may not be encountered. Replicate samples are therefore likely to contain drags in which such aggregations both have and have not been sampled. Most of the larvae are likely to be found on only a few drags if aggregations have been hit. Such drag data is therefore likely to be overdispersed. If very few aggregations are present in an area dragged, it is likely that the mean for the sample will be underestimated. If a large number of aggregations are present the reverse is likely.

Three sets of drags, each of 3 replicates, are given in Table 2 for both *Amblyomma hebraeum* and *Boophilus decoloratus* larvae. These drags were selected randomly from data collected in the Kruger National Park by I. G. Horak and A. M. Spickett. In each case the ratio of $s^2/\bar{x}$ is very much greater than 1 indicating overdispersion. This pattern is typical for drag data.

The importance of this overdispersion is demonstrated in Table 3. This takes two sets of drag data for *A. hebraeum* with widely differing means and shows additional possible drags generated first by substituting the highest real drag count with the lowest and then by the converse substitution. Both of the newly generated drags are possible as the substitution is with a real count obtained for that sampling area and date.

TABLE 3 Examples of the possible error associated with data from drag samples

| | Replicates | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | $\bar{x}$ |
| (A) | | | | |
| Real drag | 20 | 69 | 701 | 263,3 |
| Max. replaced by min. | 20 | 69 | 20 | 36,3 |
| Min. replaced by max. | 701 | 69 | 701 | 490,3 |
| (B) | | | | |
| Real drag | 200 | 594 | 4329 | 1707,7 |
| Max. replaced by min. | 200 | 594 | 200 | 331,3 |
| Min. replaced by max. | 4329 | 594 | 4329 | 3084,0 |

In the first data group the value of the mean number of larvae per drag varies from 36,3 to 490,3 and in the second from 331,3 to 3084,0. Thus, depending on where the random sample is taken, the mean generated may vary by up to an order of magnitude. It is impossible to tell what the inferred mean should be from such data. Any interpretation of such means is likely to be inaccurate and may be highly misleading. Such data would therefore make the quantitative assessment of larval densities unreliable and allows only the interpretation of very conspicuous trends. Increasing the number of drags, however, would increase the accuracy of the estimation of the mean.

*Dispersion of ticks on hosts*

A similar situation pertains to infestation pattern of ticks on hosts. The data in Table 4 represents the number on engorged larval *A. hebraeum* ticks detaching from 5 goats over 4 infestations. Each infestation was of approximately 2 000 larvae. The first 3 infestations occurred over a 10 week period and the 4th 6 months later. The aim of the experiment was to determine if an immune response occurred (data supplied by B. Fivaz and D. Adamson).

TABLE 4 The response of 5 Boer goats to repeated infestations of *A. hebraeum* larvae

| Infestation number | Goat number | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 2499 | 1528 | 3942 | 539 | 193 |
| 2 | 1183 | 1369 | 107 | 2901 | 240 |
| 3 | 16 | 84 | 13 | 77 | 344 |
| 4 | 894 | 1382 | 130 | 1283 | 875 |

The first 3 goats show consistently decreasing successful larval detachment over the first 3 infestations followed by an increase on the 4th. However, the relative decreases and increase for each successive infestation differ for each goat. Goat 4 shows an increase in larval detachment on the second infestation while goat 5 shows an continual increase in detachment throughout the infestation period.

Clearly the goats react in different ways to infestation with *A. hebraeum* larvae. This is indicated by the significant value of $s^2/\bar{x}$ for each infestation (Table 5). Mean tick burden for a given infestation the-

TABLE 5 Analysis of data on Boer goat infestation for overdispersion

| Infestation number | $\bar{x}$ | $s^2/\bar{x}$ | $t_4$ | P |
|---|---|---|---|---|
| 1 | 1740,2 | 1336,8 | 1889,2 | <0,001 |
| 2 | 1160,0 | 1084,2 | 1531,8 | <0,001 |
| 3 | 106,8 | 174,9 | 245,9 | <0,001 |
| 4 | 912,8 | 266,1 | 375,0 | <0,001 |

refore has little quantitative value as these goats do not represent a homogeneous sample. The use of only a mean to represent such data will be misleading by ignoring the great variation between goats which is present. Clearly, the more heterogeneous the response is, the more goats must be used for inference to all goats.

To summarize: a great deal of data which is collected on tick populations, both from free-living stages as well as from collections off wild and experimental hosts, is not distributed in a way allowing for normal methods of statistical analysis. There is likely to be very great variation around the mean. Quantitative statements made on such data without recognition of the underlying problem may well be wrong and therefore scientifically misleading.

*How to handle these problems*

It is possible to reduce the effects of these problems by using the correct sampling procedures and the correct method of statistical analysis (Table 6).

TABLE 6 What to do when overdispersion is expected

| | |
|---|---|
| (A) | When possible carry out a preliminary experiment to determine the variablity present |
| (B) | Ensure that the sample size is sufficiently large to enable correct statistical tests to be performed. Use larger or repetitive samples. Replicate data for seasonal abundance between years |
| (C) | Use non-parametric statistical tests or use parametric statistical tests on transformed data |

Firstly it is important to sample a sufficient number of times. With low sample sizes the chances of getting a non-representative sample are high as the effect of a single highly divergent individual will be greater. Moreover, it is also more likely that highly parasitised individuals will be underrepresented as they are uncommon. If at all possible preliminary experiments should be conducted to determine the variability of an effect to be measured and the type of underlying distribution present. If this information if available, the number of samples needed for an acceptable accuracy or difference between means can be estimated, before the real experiment is carried out (Harris, Horvitz & Mood, 1948; Karandinos, 1976). The number of sampling units (hosts or drags) needed is entirely dependent on the variability of the measured effect or response and no specific recommended number can be given. However if fewer than 10 individuals are used per "treatment" the researcher may find it difficult, if not impossible, to accurately interpret his result. If any analysis of variance is to be carried out on normalized data (see section on transformations), 18 degrees of freedom or more usually represents an analysis of acceptable accuracy.

If large sample sizes cannot be obtained at a given time then a repeat of the experiment should enable the consistency of the result to be determined. Consistency between years in surveys to obtain seasonal trends is an indication that the qualitative trend is real. However, as yearly variation in tick density is possible, mean values calculated from small sample sizes may have little value. Again, the larger the sample size the more accurate the interpretation is likely to be.

It is usually impossible to analyse raw, overdispersed data using parametric methods because parametric tests assume that the data is distributed randomly. Alternative methods of analysis do however exist. Non-parametric statistical methods do not assume a normal distribution of data. These methods use ranked data so that real differences between the numbers of ticks in different replicates are reduced to differences in ranks. Information on variability is therefore lost. With somewhat larger sample sizes however, non-parametric tests should give a good indication of real differences between samples. In the data presented in Table 4 a suitable non-parametric test (Wilcoxon paired signed ranks test) was unable to show differences between infestations (except for infestations 3 and 4 where Tc = 0; $P < 0{,}05$) because goat 5 responded differently to the other goats. A larger sample size (more goats) would have increased the chances of finding the decrease in larval attachment expected over the first 3 infestations.

It may also be possible to transform data so that the assumptions for parametric tests can be met. The transformation used will depend on the way in which the data is dispersed. Fig. 3 and 4 plot log x against log $s^2$ for larvae collected on a number of replicate drag samples from a Maroela Knobthorn Savanna vegetational zone in the Kruger National Park. A regression analysis can be carried out on this data and the slope calculated. This slope can be used to determine the type of transformation needed (Table 7). Graphs for both *A. hebraeum* and *B. decoloratus* larvae have slopes of approximately 2 and hence a log (or ln) transformation is indicated (Table 7; Taylor, 1961). This type of transformation is most often suitable for overdispersed data.

TABLE 7 Transformations which can be used to normalize data
When log s is plotted against log x the slope of the regression is b

| b approx. | Dispersion pattern | Transformation |
|---|---|---|
| 0 | Underdispersed | $x^2$ |
| 1 | Random | $\sqrt{x}$ |
| 2 | Overdispersed | log x; log (x+1) alternately ln |
| 3 | Overdispersed | $1/\sqrt{x}$ |

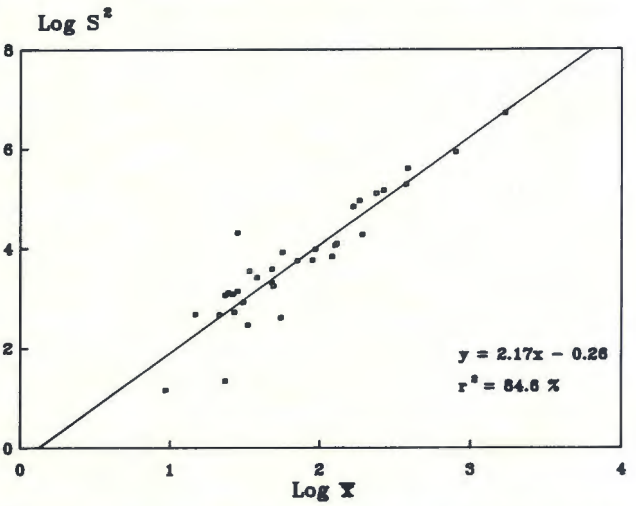


FIG. 3 Regression of log $\bar{x}$ *A. hebraeum* larvae with log $S^2$ for drag samples taken from Maroela Knobthorn Savanna in the Kruger National Park. The regression equation and the coefficient of determination ($r^2$) are given

The goat data (Table 4) can be used as an example of the possible satistical treatment of overdispersed

TABLE 8 Results for statistical analyses of goat data

| Original data | | | | Data transformed to log (x) | | | |
|---|---|---|---|---|---|---|---|
| (A) **Bartlett's test** | | | | | | | |
| $X^2_{(3)} = 15,09$; P<0,01 | | | | $X^2_{(3)} = 3,04$; ns | | | |
| variances are heterogeneous | | | | variances are homogeneous | | | |
| (B) **Parametric ANOVA** | | | | | | | |
| $F_{(3,16)} = 2,39$; ns | | | | $F_{(3,16)} = 6,11$; P<0,01 | | | |
| C.V. = 100,06 % | | | | C.V. = 20,54 % | | | |
| means are: | | | | | | | |
| 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 |
| 1740 | 1160 | 912 | 107 | 3,039 | 2,851 | 2,816 | 1,733 |
| Test invalid | | | | | | | |
| (C) **Kruskal-Wallis test** | | | | | | | |
| $X^2_{(3)} = 9,15$; P<0,05 | | | | as for original data | | | |
| means are: | | | | | | | |
| 1 | 2 | 4 | 3 | | | | |
| 1740 | 1160 | 912 | 107 | | | | |

Means underlined are not significantly different at 5 %

TABLE 9 Results for the statistical analysis of goat data omitting goat 5

| Original data | | | | Data transformed to log (x) | | | |
|---|---|---|---|---|---|---|---|
| (A) **Bartlett's test** | | | | | | | |
| $X^2_{(3)} = 16,80$; P<0,05 | | | | $X^2_{(3)} = 0,77$; ns | | | |
| variances are heterogeneous | | | | variances are homogeneous | | | |
| (B) **Parametric ANOVA** | | | | | | | |
| $F_{(3,12)} = 3,23$; ns | | | | $F_{(3,12)} = 9,53$; P<0,01 | | | |
| C.V. = 86,36 % | | | | C.V. = 18,49 % | | | |
| means are: | | | | | | | |
| 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 |
| 2127 | 1390 | 922 | 47 | 3,227 | 2,925 | 2,816 | 1,532 |
| Test Invalid | | | | | | | |
| (C) **Kruskall-Wallis test** | | | | | | | |
| $X^2_{(3)} = 9,62$; P<0,05 | | | | same as original | | | |
| means are: | | | | | | | |
| 1 | 2 | 4 | 3 | | | | |
| 2127 | 1390 | 922 | 47 | | | | |

Means underlined are not significantly different at 5 %

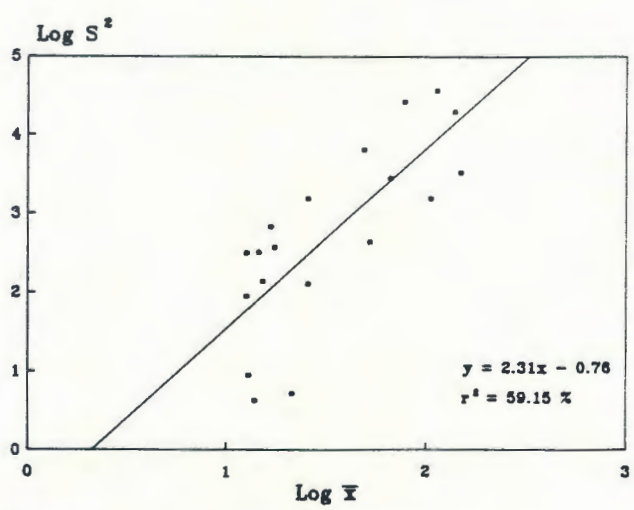


FIG. 4 Regression of log $\bar{x}$ *B. decoloratus* larvae with log $S^2$ for drag samples taken from Maroela Knobthorn Savanna in Kruger National Park. The regression equation and the coefficient of determination ($r^2$) are given

data. Plotting the raw data indicated extensive variation but plotting the data after transformation to logarithms showed a clear trend that the detachment for the third infestation was considerably lower than that for the other infestations (Fig. 5). These plots also show that goat 5 reacted completely differently to the other 4 goats.

The second step was to subject the data to an analysis of variance. Although the same goats were used for successive infestation resulting in dependency after each successive infestation, the four infestations were considered to be independent for the purpose of this exercise. The results of the analysis are given in Table 8.

Considering the analyses on all 5 goats it is clear that the parametric analysis of variance, which does not indicate significant differences between infestations, is invalid. Bartlett's test for homogeneity of variances is sensitive for non-normality (due to large variances) and the significant chi-squared value indicates that the assumptions for the ANOVA are not satisfied. The non-parametric test (here the Kruskal-

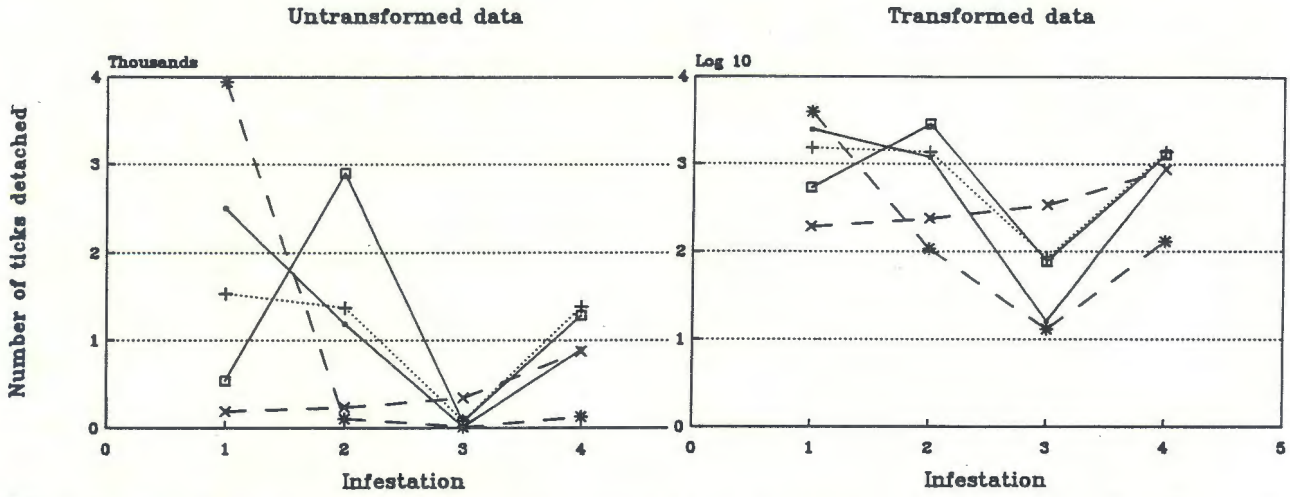


FIG. 5 The effect of log transformation on the infestation pattern of *A. hebraeum* larvae on goats. Each symbol represents a different goat

Wallis test) is, however, valid because the normality assumption is not applicable. With this test the third infestation shows a significantly lower detachment than the first infestation.

After transformation of the data to logarithms (base 10), the variances were homogeneous and the ANOVA can be regarded as valid. A highly significant, smaller detachment for the third infestation compared to all the other infestations was observed. Note that the separation of means is more prominent than for the Kruskal-Wallis analysis. The power of the non-parametric tests is somewhat less than that of the parametric tests when the same number of observations are used.

As goat 5 reacted completely differently to the other 4 goats, the data for this goat was omitted and the same analysis again applied (Table 9). The same results were obtained as for all 5 goats.

It is clear that correct statistical treatment of data is of the utmost importance for reliable interpretation of that data. Incorrect treatment can lead to unclear or incorrect interpretations.

## REFERENCES

ANDERSON, R. M. & MAY, R. M., 1985. Helminth infections of humans: mathematical models, population dynamics and control. *Advances in Parasitology*, 24, 1–101.

CROFTON, H. D., 1971. A quantitative approach to parasitism. *Parasitology*, 63, 343–364.

ELLIOT, J. M., 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Scientific Publication 25. Freshwater Biological Association. 2nd ed.

HARRIS, M., HORVITZ, D. G. & MOOD, A. M. 1948. On the determination of sample sizes in designing experiments. *Journal of the American Statistical Association*, 43, 391–402.

KARANDINOS, M. G., 1976. Optimum sample size and comments on some published formulae. *Bulletin of the Entomological Society of America*, 22, 417–421.

KERSHAW, K. A., 1973. Quantitative and dynamic plant ecology. Edward Arnold. 2nd ed.

PETNEY, T. N. & FOURIE, L. J., 1989. The dispersion of the Karoo paralysis tick, *Ixodes rubicundus*, within a naturally infested population of sheep in South Africa. *Veterinay Parasitology*, in press.

PETNEY, T. N. & HORAK, I. G., 1987. The effect of dipping on parasitic and free-living populations of *Amblyomma hebraeum* on a farm and on an adjacent nature reserve. *Onderstepoort Journal of Veterinary Research*, 54, 529–533.

RECHAV, Y., 1979. Migration and dispersal patterns of three african ticks (Acari: Ixodidae) under field conditions. *Journal of Medical Entomology*, 16, 150–163.

TAYLOR, L. R., 1961. Aggregation, variance and the mean. *Nature*, 189, 732–735.

TAYLOR, L. R., WOIWOOD, I. P. & PERRY, J. N., 1981. The density dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology*, 47, 383–406.