MURDOCH RESEARCH REPOSITORY

http://researchrepository.murdoch.edu.au

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination.

http://researchrepository.murdoch.edu.au/3859

Rapid Communication.

Geometric means provide a biased efficacy result when conducting a faecal egg count reduction test (FECRT).

R.J. Dobson[a,*], N.C. Sangster[b], R.B. Besier[c], R.G. Woodgate[c]

[a]Murdoch University, School of Veterinary & Biomedical Sciences, South Street Murdoch, WA 6150 Australia.

[b]School of Animal & Veterinary Sciences, Charles Sturt University, Locked Bag 588 Wagga Wagga, NSW 2678 Australia.

[c]Department of Agriculture & Food WA, 444 Albany Highway, Albany, WA 6330 Australia.
* Corresponding author. Tel. +61 8 93607423; Fax. +61 8 93104144; Email R.Dobson@murdoch.edu.au

Abstract

The process of conducting a faecal egg count reduction test was simulated to examine whether arithmetic or geometric means offer the best estimate of efficacy in a situation where the true efficacy is known. Two components of sample variation were simulated: selecting hosts from the general population which was modelled by the negative binomial distribution (NBD), and taking an aliquot of faeces from the selected host to estimate the worm egg count by assuming a Poisson distribution of sample counts. Geometric mean counts were determined by adding a constant (C) to each count prior to log transformation, C was set at 25, 12 or 1. Ten thousand Monte Carlo simulations were run to estimate mean efficacy, the 2.5% (lower) and the 97.5% (upper) percentile based on arithmetic or geometric means.

Arithmetic means best estimated efficacy for all different levels of worm aggregation. For moderate levels of aggregation and with C=1 the geometric mean substantially overestimated efficacy. The bias was reduced if C was increased to 25 but the results were no better than those based on arithmetic means. For very high levels of aggregation (over-dispersed populations) the geometric mean underestimated efficacy regardless of the size of C. It is recommended that the guidelines on anthelmintic resistance be revised to advocate the use of arithmetic means to estimate efficacy.

1. Introduction

As anthelmintic resistance emerges in a parasite population it has been recommended that resistance be declared when a faecal egg count reduction test (FECRT) result is less than 95% efficacy or if the lower confidence limit is below 90% (Anon,1989; Coles et al., 1992). Both these publications advocate the use of arithmetic means in preference to geometric means to estimate efficacy and provide methodology for determining the 95% confidence interval for the estimate of efficacy. However, more recent publications have advocated the use of geometric means (Wood et al., 1995; Smothers et al., 1999; Vercruysse et al., 2001) for determining efficacy in controlled slaughter test and FECRT.

There is little disagreement that when conducting an ANOVA or testing the differences between two means of a parasite population the data should be transformed (e.g. using logs or roots) so that variances between groups are more homogenous. However it cannot be assumed that the best transformation to stabilise variances is also the best to determine efficacy (Dash et al. 1988). We set out to explore this question using Monte Carlo simulation

2

where the true efficacy was set at the "critical" point of 95% (Miller et al., 2006). If efficacy is 100%, or so low that all treated animals exhibit positive counts, then the choice of either geometric or arithmetic mean is of little consequence. However, when a proportion of the pre- or post-treatment counts are zero the choice of mean can make a substantial difference to the resulting efficacy. To estimate the geometric mean where some zero counts are present in a data set, a constant (C) must be added prior to a log transformation as otherwise the geometric mean becomes zero. Donald et al., (1978) found that log (count+25) was most effective for stabilizing variances when analysing tracer sheep worm counts. Dash et al. (1988) suggested that the value added to each count should be half the minimum detection level. We therefore explored different values for C for estimating the geometric mean for comparison with the arithmetic mean and assumed a detection level of 25. Vidyashankar et al., (2007) present methods for determining efficacy, particularly for relatively small samples, that do not require transformation of the data. They achieve this by developing a statistical model for the change in pre to post treatment counts which is independent of distributional assumption for the raw data.

The negative binomial distribution (NBD) is considered to adequately model egg and worm count data (Morgan et. al., 2005; Barger, 1985), and the aggregation parameter k for the NBD typically varies from 0.2 to 2.3 for commercial flocks. If k is large (say greater than 10) the NBD of parasite counts within a flock begins to approach the normal distribution where the mean is a good measure of central tendency. When k is small then the NBD is skewed towards the vertical (left) axis with many animals having a relatively low count and a few animals having very high counts (i.e. highly aggregated or "over-dispersed" populations). Separate Monte Carlo simulations were run to explore the impact of a range of k values on the appropriate mean to use in a FECRT. The Poisson distribution is used to describe counts

that arise as a result of a random process or if objects are randomly distributed, such as worm eggs in a sample volume of liquid drawn from a larger agitated volume. When the expected number of counts is low the Poisson is skewed to the left (like the NBD) but as the expected number of counts increases (say greater than 15) the Poisson distribution tends to become symmetric about the mean.

2. Materials and methods

2.1 Data Generation.

In Table 1 the 20 "true" counts were a random sample drawn from a negative binomial distribution (NBD) with a mean of 300 and a dispersion parameter k of 2. The "observed" count was obtained from the true count by taking a random Poisson sample (Morgan et. al., 2005) for each true value as follows. Each true count was divided by the detection or multiplication factor (set here to 25); this result was set as the expected number of eggs to be found in the Poisson distribution from which the observed sample count was drawn (total eggs counted); the randomly sampled eggs counted were then multiplied by the detection factor to give the observed count. To obtain a 95% reduction in an FECRT the true post treatment count was set to 5% of the true pre treatment count, the observed post treatment count was obtained by taking a Poisson sample as described above. For example, in line 1 of Table 1:

146/25=5.84 expected eggs; however only 4 eggs were drawn for this particular random sample (assuming a Poisson distribution with mean 5.84) to yield an observed count of 100. Random variables (NBD and Poisson) and Monte Carlo simulations were generated using PopTools (CSIRO, Australia) within Excel (Microsoft Inc., USA).

2.2 Efficacy Estimates.

4

For n animals with count Xi for the i-th animal, efficacy was determined by:

Arithmetic mean count, $\mu = (\sum Xi)/n$ for i=1…n

Geometric mean count, $\underline{\mu} = \{10^\wedge([\sum \log(Xi + C)]/n)\} - C$ for i=1…n & C≥1

Percent Efficacy, $E = 100(1 - \mu_t/\mu_u)$ where μ_t & μ_u are the mean for

treated and untreated counts respectively. Note C is an arbitrary constant added to all counts

prior to log transformation and should be greater than 1.

2.3 Monte Carlo Simulation.

Monte Carlo analysis was conducted by successively re-estimating Table 1, i.e. for each

iteration 20 new random samples from the NBD (with mean 300 and set k) were taken, the

new observed counts made by drawing Poisson random samples, and resulting efficacies

calculated. 10,000 Monte Carlo iterations were used to estimate the mean efficacy, 2.5%

(lower) and the 97.5% (upper) percentile for efficacy. Efficacy at each iteration was

determined using arithmetic or geometric means with C set at 1, 12 and 25 prior to log

transformation (as shown in Table 1). Separate Monte Carlo simulations were run with k for

the NBD set to 0.2, 0.5, 1, 1.5, 2 and 2.5.

3. Results

Figure 1 shows the first 1000 efficacies (iterations) based on arithmetic or geometric means

(with C=1) and NBD with mean 300 and k=2. Table 2 shows the mean efficacy, upper and

lower percentiles (2.5%) based on arithmetic or geometric means with varying C and k for

the NBD set at 2 or 0.5, for 10,000 iterations. Figure 2 shows, for C=1 and C=25, how the

mean efficacy based on arithmetic or the geometric means varies for different levels of

overdispersion (k). Table 3 shows the results for one Monte Carlo iteration where the 20

sampled sheep were drawn from a highly aggregated population, more typical of older

grazing animals, i.e. a negative binomial distribution (NBD) with a mean of 300 epg and a dispersion parameter k of 0.2.

4. Discussion

For relatively homogenous populations (e.g. k for the NBD ranges from 1 to 2.5), Figure 1 and 2 clearly demonstrate that the use of geometric means will mask the emergence of anthelmintic resistance by overestimating efficacy if an inappropriate constant (C=1) is chosen for the log transformation. For the same range of k, if C were set at 25 (Table 2 and Figure 2) then the efficacy has a smaller bias, however this result is not better than that obtained from the arithmetic mean. Given that it is difficult to estimate the optimum value of C for a particular efficacy study, it is simpler and suffers less from bias to estimate efficacy from the arithmetic means. For highly over-dispersed populations (say k for the NBD less than 1) the estimate of efficacy by geometric means becomes unstable regardless of the value chosen for C (Figure 2). Morgan et. al., (2005) showed that younger sheep with more uniform counts across the flock generally have a higher k (above 0.5), and such sheep are considered most appropriate for FECRT. In this situation setting C=1 provides a disproportionately high weight to post-treatment zero values, leading to an inflated estimate of efficacy from geometric means. Figure 2 and Table 3 demonstrate that for k approximately below 0.5, i.e. in older sheep with less uniform counts, the geometric mean was likely to underestimate efficacy. This result dispels the common belief that arithmetic means generally provide lower estimates of efficacy than geometric means (Dash et al., 1988; Coles et al., 1992; Vercruysse et al., 2001). This problem arises when pre-treatment or control counts include some zeros. The geometric means then reduced the control count by a greater proportion than the post-treated count was reduced, leading to an under estimate of efficacy (e.g. see Table 3). Further arithmetic means better estimate the true levels of pasture

contamination with worm eggs (Dash et al., 1988 and see Table 3). Despite this it is usually

necessary to transform parasite data, in an attempt to make the variances between treatment

groups of an experiment approximately equal, prior to other routine statistical procedures

(e.g. ANOVA or t-test).

In summary geometric means are likely to underestimate efficacy results when there are

some zero counts in the control/pre-treatment counts (i.e. when k is low) and also

overestimate efficacy when all pre-treatment counts are positive but post-treatment counts

contain some zero counts (i.e. when k is high and efficacy is relatively high but not 100%).

For unbiased results the distribution of efficacy estimates would be less than the true efficacy

for approximately 50% of the time. The upper and lower percentiles given in Table 2 define

the bounds for the central 95% of the efficacy results. When k=2 for the NBD it is alarming

that the geometric mean for C=1 only yields an efficacy result below 96% for less than 2.5%

of the time (Figure 1 & Table 2), thus while being precise (i.e. having the closest upper and

lower bounds) these results are inaccurate. Despite arithmetic and geometric means being

calculated from the same data (e.g. see Tables 1 & 3) the bounds for 95% of the efficacy

results tend to be wider for efficacies based on geometric means particularly when k=0.5

(Table 2) indicating the improved accuracy and precision obtained by using arithmetic means

to estimate efficacy.

The same problems with geometric means apply equally to worm count data used to estimate

efficacy from a controlled slaughter test as described by Wood et al., (1995). This is because

the sampling issues are the same; the slaughtered animals are drawn from a NBD and

aliquots of intestinal contents used to estimate worm numbers have Poisson sampling

variability. If there are some zero values in control or treated groups then efficacy based on geometric means are likely to be unreliable.

In comparisons of mathematical techniques for estimating worm egg count reductions obtained for different drugs in a FECRT, Torgerson et al., (2005) and Schnyder et al., (2005) found an instance where geometric means failed to declare resistance while arithmetic means indicated suspect resistance, although they did not indicate how the geometric mean was estimated or discuss this issue. The methodology presented here is a simple efficient way to explore the various factors associated with precision and accuracy of FECRTs and controlled slaughter tests, it is in part similar to that used by Torgerson et al., (2005). Parameters such as sample size (n) (Torgerson et al., 2005), flock mean, detection level and methodology (e.g. control vs. treated groups; pre vs. post treatment counts) can be analysed for a variety of situations. The effects of some of these factors have been explored in the field (Miller et al., 2006), but to evaluate these systematically in a variety of commercial flocks would be impractical and expensive.

5. Conclusion

When conducting FECRTs only the efficacy estimated from the arithmetic mean consistently provides unbiased results where the expected efficacy is close to that of the true efficacy. In contrast, efficacies determined by geometric means from the same data often yield biased results. Geometric means should not be used to estimate efficacy because the appropriate constant added to counts prior to log transformation, to provide an unbiased result, will vary for different situations. The use of arithmetic means should be advocated as the basis to estimate efficacy and anthelmintic resistance detection in future guidelines.

References

Anonymous, 1989. Report of the Working Party for the Animal Health Committee of the Standing Committee on Agriculture. SCA Tech. Rep. Ser. No. 28, Anthelmintic Resistance, CSIRO, Australia.

Barger, I.A., 1985. The statistical distribution of trichostrongylid nematodes in grazing lambs. Int. J. Parasitol. 15, 645–649.

Coles, G.C., Bauer, C., Borgsteede, F.H.M., Geerts, S., Klei, T.R., Taylor, M.A., Waller, P.J., 1992. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. Vet. Parasitol. 44, 35-44.

Dash, K.M., Hall, E., Barger, I.A., 1988. The role of arithmetic and geometric mean worm egg counts in faecal egg count reduction tests and in monitoring strategic drenching programs in sheep. Aust. Vet. J. 65, 66-68.

Donald, A.D., Morley, F.H.W., Waller, P.J., Axelsen A., Donnelly J.R., 1978. Availability to Grazing Sheep of Gastrointestinal Nematode Infections Arising from Summer Contamination of Pasture. Aust. J. Agric. Res. 29, 189-205.

Miller, C.M., Waghorn, T.S., Leathwick, D.M., Gilmour, M.L., 2006. How repeatable is a faecal egg count reduction test? New Zealand Vet. J. 54, 323-328.

Morgan E. R., Cavill L., Curry G. E., Wood R., Mitchell, E. S. E., 2005. Effects of aggregation and sample size on composite faecal egg counts in sheep. Vet. Parasitol. 131, 79-87.

Schnyder, M., Torgerson, P.R., Schonmann, M., Kohler, L., Hertzberg, H., 2005. Multiple anthelmintic resistance in Haemonchus contortus isolated from South African Boer goats in Switzerland. Vet. Parasitol. 128, 285–290.

Smothers, C.D., Sun, F., Dayton, A.D., 1999. Comparison of arithmetic and geometric means as measures of a central tendency in cattle nematode populations. Vet. Parasitol. 81, 211–224.

Torgerson, P.R., Schnyder, M., Hertzberg, H., 2005. Detection of anthelmintic resistance: a comparison of mathematical techniques. Vet. Parasitol. 128, 291–298.

Vercruysse, J., Holdsworth, P., Letonja, T., Barth, D., Conder, G., Hamamoto, K., Okano, K., 2001. International harmonisation of anthelmintic efficacy guidelines. Vet. Parasitol. 96, 171–193.

Vidyashankar, A.N., Kaplan, R.M., Chan, S., 2007. Statistical approach to measure the efficacy of anthelmintic treatment on horse farms. Parasitol. 134, 2027-2039.

Wood, I.B., Amaral, N.K., Bairden, K., Duncan, J.L., Kassai, T., Malone, J.B., Pankavich, J.A., Reinecke, R.K., Slocombe, O., Taylor, S.M., Vercruysse, J., 1995. World association for the advancement of veterinary parasitology (WAAVP). Second edition of guidelines for evaluating the efficacy of anthelmintics in ruminants (bovine, ovine, caprine). Vet. Parasitol. 58, 181–213.

Table 1. Sample data from one Monte Carlo iteration for NBD with mean 300 and k=2. Efficacy was determined for observed counts from arithmetic (AM) and geometric means (GM). For the latter the constant (C) of 1, 12 or 25 was added to each count prior to log transformation.

	Pre Treatment Count		Post Treatment Count	
	True Count	Observed Count	True Count	Observed Count
	146	100	7	50
	354	225	18	25
	420	325	21	50
	188	175	9	0
	250	300	13	0
	255	300	13	0
	125	100	6	0
	313	350	16	0
	422	450	21	0
	496	325	25	0
	292	275	15	25
	220	175	11	50
	156	100	8	0
	367	275	18	0
	185	125	9	0
	219	150	11	0
	178	250	9	0
	428	675	21	0
	336	300	17	0
	41	50	2	0
AM	270	251	13	10
GM C=1	238	213	12	1
Efficacy from Arithmetic Mean				96.0%
Efficacy from Geometric Mean C=1				99.3%
Efficacy from Geometric Mean C=12				97.6%
Efficacy from Geometric Mean C=25				96.9%

Table 2. Mean efficacy, upper and lower percentiles from 10,000 Monte Carlo simulations where efficacy from observed counts was determined from arithmetic or geometric means with constant (C) of 1, 12 or 25 added to the counts prior to log transformation. Samples were drawn from a NBD with mean of 300 and k set to 2 or 0.5, true efficacy was set at 95%.

Type of Mean	Efficacy	2.5 Lower percentile	97.5 Upper percentile
k = 2 Arithmetic	95.0%	91.8%	97.7%
Geometric C=1	98.3%	96.0%	99.5%
Geometric C=12	95.9%	92.7%	98.3%
Geometric C=25	95.2%	91.9%	97.9%
k = 0.5 Arithmetic	95.0%	91.6%	97.8%
Geometric C=1	96.2%	90.6%	99.0%
Geometric C=12	93.2%	87.2%	97.3%
Geometric C=25	92.7%	87.0%	97.0%

Table 3. Sample data from one Monte Carlo iteration for NBD with mean 300 and k=0.2. Efficacy was determined for observed counts from arithmetic (AM) and geometric means (GM). For the latter the constant (C) of 1, 12 or 25 was added to each count prior to log transformation.

	Pre Treatment Count		Post Treatment Count	
	True Count	Observed Count	True Count	Observed Count
	0	0	0	0
	0	0	0	0
	171	150	9	0
	410	500	21	25
	149	125	7	0
	1996	2100	100	50
	200	150	10	25
	43	25	2	0
	1174	1025	59	25
	948	875	47	50
	0	0	0	0
	241	300	12	0
	2726	2550	136	50
	1	0	0	0
	0	0	0	0
	142	75	7	50
	1	0	0	0
	929	950	46	75
	0	0	0	0
	5	0	0	0
AM	457	441	23	18
GM C=1	42	33	5	3

Efficacy from Arithmetic Mean	96.0%
Efficacy from Geometric Mean C=1	89.5%
Efficacy from Geometric Mean C=12	88.2%
Efficacy from Geometric Mean C=25	88.9%

Figure 1. Distribution of first 1000 FECRT results from Monte Carlo simulation where efficacy was determined from observed counts by arithmetic or geometric means. The 2.5% and the 97.5% percentile for efficacy are indicated on each histogram by "**V**". True counts were drawn from an NBD with mean 300 and k equals 2, true efficacy was set at 95%. Observed counts were drawn assuming a detection factor of 25 and a Poisson sampling of the true count. For geometric means the constant C=1 was added to the counts prior to log transformation.

Figure 2. Mean efficacy from 10,000 Monte Carlo simulations where efficacy was determined from observed counts by arithmetic mean (AM) and geometric means with constant of 1 (GM C=1) and 25 (GM C=25) added to the counts prior to log transformation. Samples were drawn from a NBD with mean of 300 and k ranging from 0.2 to 2.5, true efficacy was set at 95%.
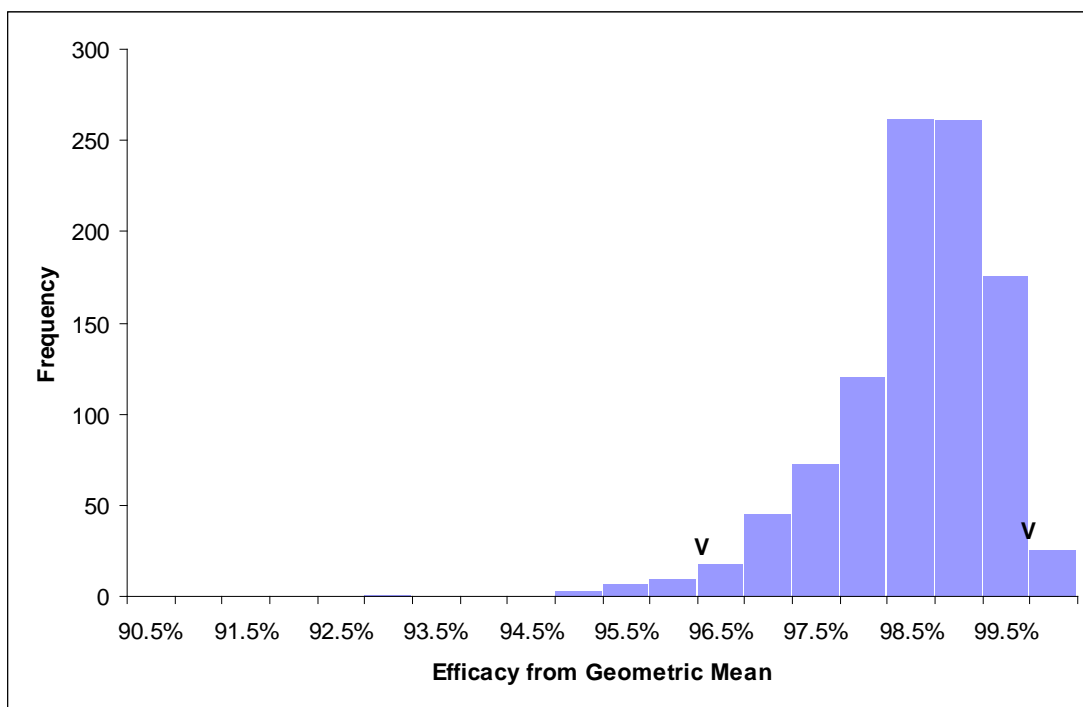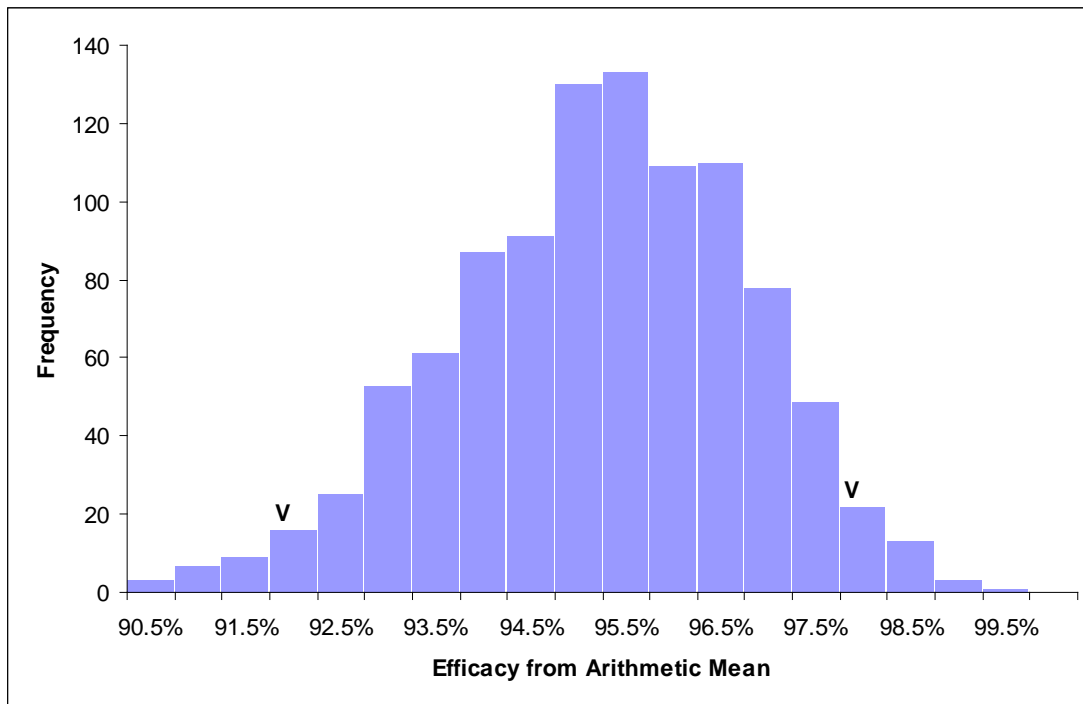
Figure 1

Figure 2