

# Analysis of Parasite and Other Skewed Counts

Neal Alexander

London School of Hygiene and Tropical Medicine

## Abstract

**OBJECTIVE** To review methods for the statistical analysis of parasite and other skewed count data.

**METHODS** Statistical methods for skewed count data are described and compared, with reference to those used over a ten year period of *Tropical Medicine and International Health*. Two parasitological datasets are used for illustration.

**RESULTS** Ninety papers were identified, 89 with descriptive and 60 with inferential analysis. A lack of clarity is noted in identifying measures of location, in particular the Williams and geometric mean. The different measures are compared, emphasizing the legitimacy of the arithmetic mean for skewed data. In the published papers, the  $t$  test and related methods were often used on untransformed data, which is likely to be invalid. Several approaches to inferential analysis are described, emphasizing 1) non-parametric methods, while noting that they are not simply comparisons of medians, and 2) generalized linear modelling, in particular with the negative binomial distribution. Additional methods, such as the bootstrap, with potential for greater use are described.

**CONCLUSIONS** Clarity is recommended when describing transformations and measures of location. It is suggested that non-parametric methods and generalized linear models are likely to be sufficient for most analyses.

MeSH Keywords: Statistical Data Analysis; Parasitology; Statistics, Nonparametric;

Regression Analysis

82, 83, 84...

*87 weevils today, Mr. Victor.*

*That's above average, but the trend is down.*

In a wartime detention camp, a boy counts the vermin in his food rations (from the film *Empire of the Sun*, based on the autobiographical novel by JG Ballard).

## **INTRODUCTION**

Counting items such as parasites results in non-negative integer data: 0, 1, 2, etc. The aim of this paper is to review methods for statistical analysis of such data in medical research. The focus will be on parasite counts although most of the same methods are applicable to other count data such as numbers of insects, plaques or disease episodes.

Descriptive analysis of such data usually includes summary numbers which aim to convey average values of the data. In statistics these are called ‘measures of location’. For count data even these simple measures prove problematic. This is because the data are often skewed: their ‘long tail’ (Mumpower and McClelland 2002) means that a small proportion of people can, for example, harbour a large proportion of the parasites. In basic statistics it is often taught that such skewness precludes the use of the arithmetic mean because it is ‘overly’ influenced by the high values (Kirkwood and Sterne 2003). A possible alternative is to analyse the logarithms of the data instead of the original values. However, for count data, this is only feasible in the absence of zeros, because the logarithm of zero is not a finite

number. This has led to controversy over what measures of location are appropriate, and to basic terms such as ‘geometric mean’ being used inconsistently.

Much of this controversy results from a reliance on statistical methods which use the normal (Gaussian) distribution, whose symmetry usually precludes a good fit to count data. However, this dependence is now largely outdated, due to the availability of approaches based on skewed distributions such as the negative binomial (Anderson and May 1991).

The current paper aims to review the statistical methods currently being used for such data, and describe their advantages and disadvantages. It will concentrate on the most common kinds of analyses, leaving aside more specialized ones such as spatial patterns, repeatability and reproducibility. Descriptive methods will be addressed first, then inferential methods, the latter being those which compare counts between groups or correlate them with other variables.

## **METHODS AND RESULTS**

### **Descriptive methods**

#### *Measures of location: median and means*

Location is “the notion of central or ‘typical value’ in a sample distribution” (Everitt 1995). The most commonly measures of location are the median, and various types of mean. The *median* is the middle value of the sorted data. If the number of values is even, then there are two middle values, and the convention is to define the median as the value half way between them. The median is zero if more than half of the values are zero.

By default ‘the mean’ usually refers to the *arithmetic mean* — the sum of the values divided by  $n$ , the number of values — although there are other types of mean which may be useful for count data. In particular, the *geometric mean* is obtained by multiplying together all the data values, then taking the  $n$ th root (Borowski and Borwein 1989, Everitt 1995). The geometric mean is always less than the arithmetic mean, unless all values in the dataset are identical, in which case the two are equal (Borowski and Borwein 1989). The geometric mean is zero whenever any of the individual values is zero. This makes it rather meaningless for datasets with zeros, e.g. uninfected people.

The geometric mean is related to logarithmic transformation of the data. If there are no zero values, then the geometric mean equals the exponential of the mean of the log-values. Any exponent can be used, but it must equal the logarithm base, e.g. 10 to the power of the mean of the  $\log_{10}$ -values. This suggests the following modification of the geometric mean to accommodate zero values: add 1 to all the data values, then take the geometric mean, then subtract 1 again. This is known as the *Williams mean* (Williams 1937). There are also other types of mean. For example, the *square mean root* is the square of the mean of the square root-transformed data. This measure has applications in meta-analysis (Bushman and Wang 1995) and in assessing the agreement of counts between observers or methods (Alexander et al. 2007). Other powers can also be used: in general these are called power means or algebraic means (Bonferroni 1950).

*Confusion between geometric and Williams means*

In the medical research literature, the Williams mean is sometimes called the geometric mean, as in the onchocerciasis community microfilarial load (CMFL) (Remme et al. 1986). Nevertheless, the two are not equal. One disadvantage of the Williams mean is its dependence on the choice of units. For example, if skin snips weigh 2mg, the CMFL per snip is not double the CMFL per mg, although both units are used in the literature (Marshall et al. 1986) (Remme et al. 1986). This scale dependence can be avoided by adding and subtracting a value with units e.g. 1/mg rather than a dimensionless 1 (Alexander et al. 2005).

*The geometric mean is not a clever way to estimate the arithmetic mean*

Fulford explains that the Williams mean (he calls it the geometric mean) should not be used to estimate the arithmetic mean (Fulford 1994). Similarly, Dobson et al (Dobson et al. 2009) define helminth vaccine efficacy as a fold reduction in arithmetic mean egg count, then show that this is not well estimated by a ratio of Williams means. The points made by these papers are valid although may seem rather tautological. The different types of mean differ mathematically so it should not be a surprise that one of them cannot properly estimate another. Their values are no more commensurable than, for example, weight-for-height and weight-for-age in nutrition.

*Choice of measure of location: arithmetic versus geometric mean*

Compared to the arithmetic mean, the geometric mean is said to be ‘not overly influenced by the very large values in a skewed distribution’ (Kirkwood and Sterne 2003). How much is ‘overly’? We can try to answer this question objectively by going back to whichever health outcomes motivated the study.

A clear-cut example would be noise pollution. Human perception of sound is approximately logarithmic in that orders of magnitude change in sound pressure (measured in Pascals) are perceived as approximately uniform increments on a linear scale. This is known as Weber's law (Wojtczak and Viemeister 2008) and is why sound volumes are expressed in decibels (log-ratio of pressure over baseline). Hence, averaging is more aligned with human perception if done on the logarithmic scale: geometric mean of Pascals or arithmetic mean of decibels (Olayinka and Abdullahi 2009). This reasoning is independent of the statistical distribution of the data: even if the decibel values were skewed, their arithmetic mean would still be more relevant than the geometric mean.

The choice of measure of location may not always be so straightforward. If we consider malaria parasite density, then we could imagine either the arithmetic or the geometric mean having more interest, depending on the purpose of the study. If the intensity of transmission is of interest, it may be reasonable to assume that this is proportional to parasite density, in which case the arithmetic mean would be relevant. This would correspond to the use of the annual transmission potential (ATP) in filarial diseases, which is based on the arithmetic mean number of filariae per mosquito (Bockarie et al. 1996). If, by contrast, the objective of the study relates to number of clinical cases, the geometric mean may be more appropriate because the relation of parasite density to the probability of being a case (as opposed to asymptomatic) is non-linear, the slope reducing at higher densities (Smith et al. 1994). In this case, we would need to exclude non-parasitaemic people from the calculation of the geometric mean, although this is not problematic for this purpose because, by definition, they cannot have malaria.

In other situations another measure of location, such as the median, may be preferable. The general point is to ‘deconstruct’ the problem (Hand 1994) so that the choice of measure is subordinated to the study’s objectives, with purely statistical concerns — for example achieving a normal distribution — being secondary.

### *Measures of dispersion*

The most commonly used measures of dispersion are probably the *standard deviation*, *range*, and *interquartile range*. The standard deviation is the square root of the *variance*, which is the average squared deviation from the mean. More specifically, the standard deviation equals  $\sqrt{\sum_i (x_i - \bar{x})^2 / (n - 1)}$ , where  $\sum_i$  means the sum over the data values  $x_i$ , and  $\bar{x}$  is the sample mean. The range is the interval between the minimum and maximum values in the data. This tends to increase with larger sample sizes. Percentiles are values below which a certain percentage of the data lie, after they have been sorted into ascending order. The median, for example, is the 50<sup>th</sup> percentile. The interquartile range is the interval between the 25<sup>th</sup> and 75<sup>th</sup> percentiles: this interval contains the middle half of the data.

Since the distributions of count data are usually asymmetric, using the standard deviation for error bars, or in  $\pm$  notation about the mean, are not usually helpful. This is because they will often imply negative values whereas, of course, for count data the lowest possible value is zero. For example, looking at the mean and standard deviation of the hookworm egg counts in Table 1, quoting ‘126  $\pm$  322’ would be rather meaningless. The interquartile range would be one preferable option.

Similarly, error bars based on standard deviations are likely to go below zero, as in,



for example, the figure showing schistosomiasis faecal eggs per gram by age in one of the papers in the review described below (Seto et al. 2007). Again, plotting the interquartile range, or superimposing box and whisker plots (Kirkwood and Sterne 2003), would be better options.

For count data, the greatest utility of the standard deviation or variance may be to assess how homogeneous a distribution the data have. The Poisson distribution results from homogeneous processes, which are the simplest models for count data. The mean and variance of the Poisson distribution are equal whereas, if the variance of data is considerably greater than the mean, there is said to be *overdispersion* (Elliott 1977, Mwangi et al. 2008). The variance of the negative binomial distribution, for example, equals  $\mu + \mu^2/k$ , where  $\mu$  is the mean and  $k$  is a positive-valued parameter. For small values of  $k$  the variance is much greater than the mean, reflecting a high degree of overdispersion. Such distributions are sometimes also called ‘contagious’ — even if no infectious disease is involved — because they reflect clustering in the data (Elliott 1977). These two situations — homogeneity and overdispersion — are illustrated in Figure 1. It is also possible for the variance to be less than the mean. This occurs when distances between events tend to be similar, with both smaller and larger distances being rare. This is rare in practice but would be the case, for example, for spatial events on an even grid.

## **Example datasets**

### *Hookworm eggs*

These are baseline data from a longitudinal study of hookworm in Minas Gerais state, Brazil(Cundill et al. 2011). Table 1 shows summary statistics for the total egg count from two Kato-Katz thick smears, each pair being prepared from a single stool sample. Those with missing data on age, sex, or either egg count have been excluded.

#### *Plasmodium falciparum* asexual blood stages

Children participating in a trial of anti-malarial drugs had a thick blood film read microscopically(Dunyo et al. 2006). The parasite counts from a single high power field are shown in Table 1.

### **Literature review**

Statistical methods used for parasite count data were reviewed over ten years of *Tropical Medicine and International Health*. The following search was done in the PubMed online database:

(parasit\* OR malaria\* OR helmint\* OR filar\*) AND (count\* OR intensit\* OR densit\*) AND trop med int health

For the years 2001-2010 (volumes 6-15), the search returned 156 papers. They were retained in the review if they were found to include either i) descriptive analysis of parasite count data, or ii) inferential analysis in which count data constituted an outcome variable. Papers using correlation coefficients — which do not distinguish between outcome and explanatory variables — were also included under the second category. Of the 156 papers, 90 were found to include such analyses: 89 descriptive and 60 inferential.

Methods for descriptive and inferential analysis are shown in Table 2 and Table 3. The former includes measures of location, of which the geometric mean was the most commonly used (51% of papers). In some papers, the Williams mean was calculated but presented as the geometric mean. It's likely that there were other instances of this which could not be unambiguously identified from the text. In other words, the Williams mean category (13%) was difficult, in practice, to separate from the geometric mean. The arithmetic mean was also commonly used (35%) and the median less so (7%).

### **Inferential analysis**

#### *t test and related methods*

The *t* test compares means between two samples. Although it assumes that the samples are drawn from normal distributions with equal means, the results are surprisingly robust to departure from these assumptions (Heeren and d'Agostino 1987). Nevertheless, it is liable to break down when 'skew is severe or when population variances and sample sizes both differ' (Stonehouse and Forrester 1998, Boneau 1960). At least one of these two circumstances is likely to pertain with count data. Skewness has already been mentioned, while a difference in means implies a difference in variances if the data are actually drawn from, for example, a Poisson or negative binomial distribution. Hence the *t* test cannot be recommended for untransformed parasite data. The same caveats apply to regression and analysis of variance, because these are all mathematically similar. Nevertheless, such analyses were done on untransformed data in at least 10% of the papers in the literature review (Table 3).

The performance of these methods may be improved by first transforming the data. In particular, if there are no zeros in the data, then a logarithmic transformation may suffice. Moreover, the  $t$  test and related techniques will then yield ratios of geometric means and so can easily be interpreted. Such methods were used in at least 15% of the papers in Table 3. As noted above, their robustness means that the normal distribution does not need to be a perfect fit for the results to be reliable. Figure 2 shows a histogram of malaria parasite densities from Dunyo *et al.* (2006) on a log scale. The shape is similar to that of the superimposed fitted normal curve, suggesting that the  $t$  test and related methods are likely to be applicable to the log-transformed data.

The logarithmic transformation is not applicable, however, in the presence of zeros. Various other options are available although, in most cases, they lack the same ease of interpretation. One option is to add 1, or another value, before taking logarithms. Table 2 shows that at least 13% of papers in the literature review used a transformation other than the simple logarithmic, while in 15% it was not clear what, if any, transformation was used. For negative binomial distributions, the value  $k/2$  can be added instead of 1 before taking logarithms, or an inverse hyperbolic sine transformation can be used (Beall 1942, Anscombe 1948, Laubscher 1961, Elliott 1977). Apart from the basic logarithmic transformation, these have at least two disadvantages. With few exceptions, the results cannot be interpreted as easily, and the transformation ‘cannot remove the clump’ of zeros, if present, and so the data will not be normalized (Hallstrom 2010). In general, if zeros are present, then other techniques, in particular generalized linear models (see below), are likely to be preferable to normal-theory methods on transformed data (O’Hara and Kotze 2010).

### *Non-parametric methods*

These methods generally use the ranks of the data rather than the original values. This means that, for example, a single data value much greater than the others will not greatly affect the results as it would for a  $t$  test. One of the simplest non-parametric methods is the Mann-Whitney test, which compares data from two independent samples. In the literature review, non-parametric methods were used in 23% of those papers which used any inferential method (Table 3). Nevertheless, they do have some disadvantages:

If the sample size is low, and the distribution of the data is close to normal, then they are likely to have considerably lower power than parametric methods such as the  $t$  test (Bland 1995). In these — rather limited — circumstances, a parametric method is likely to be preferable.

Contrary to common conception they cannot, in general, be interpreted as comparisons of medians (Hart 2001). For example, the Mann-Whitney test (also known as the Wilcoxon test) assesses whether two distribution functions are unequal for at least one value (Conover 1980). The test can only be interpreted as a comparison of medians if the samples are drawn from populations whose distributions have the same shape, i.e., when represented graphically, they can be laid exactly on top of each other simply by shifting one along the horizontal axis by an amount  $\Delta$ : a ‘shift alternative’ (Bauer 1972). This is unlikely to apply to count data but, if it were, it’s likely that a  $t$  test would have been applicable in the first place (Bland 1995). When software

provides a confidence interval, it's not necessarily for the difference in medians.

They are not generally capable of complex analyses, in particular those with multiple explanatory (predictor) variables.

Overall, non-parametric methods are suitable for simple analyses of skewed count data, although are not as foolproof as they may appear.

### *Generalized linear models*

We have seen techniques which assume that the distribution of the data is normal (Gaussian) and others which do not depend on it having any particular form. Another option is to find a distribution which does fit the data adequately: this is done in generalized linear modelling. This can be thought of a kind of regression in which:

The distribution of the data is not necessarily normal. For count data the Poisson or negative binomial, for example, may be suitable.

We model a function of the mean, not necessarily the mean itself. This function is called the *link function* and, for present purposes, is likely to be the logarithm. It is conventional here for statistical packages to use the natural (base  $e$ ) logarithms

Parasite counts are usually overdispersed relative to Poisson, requiring a distribution such as the negative binomial which can accommodate this. If a Poisson distribution

is used regardless, the results can be misleading (Barker and Cadwell 2008).

However, it is possible to use the ‘sandwich’ estimator (e.g. with the ‘vce(robust)’ option in STATA) or ‘quasi-likelihood’ models, both of which effectively multiply the standard errors by a factor estimated from the data (White 1980, Sileshi 2006, Noe et al. 2010), but which do not fully characterize a specific distribution for the data.

Once the distribution has been chosen, a generalized linear model is fitted by maximizing the likelihood — i.e. the probability of the data, considered as a function of the model parameters (Everitt 1995) — over possible values of the regression coefficients. A couple of points are worth stressing.

First, it is the fitted mean, not the data, which is transformed in this approach. Hence, a logarithmic link function models the logarithm of the arithmetic mean, not the geometric mean as with an ordinary regression on the log-transformed data.

Secondly, the data must have values which are feasible for the chosen distribution. In the case of the Poisson or negative binomial distributions, for example, which apply to whole-number (integer) data, then, we must model the actual counts. If we have, for example, a set of parasite counts based on varying numbers of replicates, e.g. Kato-Katz slides, we should not input the average values to the GLM, since these are not necessarily whole numbers. Rather, the log number of replicates should be included as an *offset* in the model. This is an explanatory variable with its regression coefficient constrained to equal 1. In the example of multiple slides per person, if there is a single explanatory variable  $x$  with coefficient  $\beta$ , then we model the log mean count per person as  $\log_e(\mu) = \alpha + \beta x + \log_e(n)$ , so  $\log_e(\mu/n) = \alpha + \beta x$ . We can then

obtain  $\exp(\beta)$  as the ratio of the average count per slide, even though we specified the total count as the response variable. Similarly, when analysing the incidence of events when the follow-up time varies between people, using log-time as an offset enables the analysis to yield rate ratios. Figure 3 shows mean hookworm egg count fitted as a function of age. As with other regression models, different forms of dependence can be included (e.g. polynomial), as can multiple explanatory variables.

O'Hara & Kotze compared GLMs with transformation-based methods and found that the latter perform poorly — in terms of bias and sampling error — unless dispersion is small and the mean counts large (O'Hara and Kotze 2010). Goodness of fit measures for generalized linear models, and negative binomial regression in particular, include checking that the residual deviance is similar to the residual degrees of freedom (Hilbe 2007), and plotting the residuals against fitted values to check that they have random scatter with constant variance (McCullagh and Nelder 1983). Chi-squared tests can also be done to compare the numbers in various categories of the outcome variable with those expected under the given distribution. However, it should be borne in mind that, with a large sample size, a departure from the assumed distribution may be large enough to be detected by the goodness-of-fit test, but too small to affect materially the results of the main analysis. The present author tends to rely on visual assessment of the overall distribution as shown in Figure 4.

Other distributions for skewed data include versions of the Poisson and negative binomial in which the proportion of zeros is defined by additional parameters. These are called zero-inflated distributions. The zero-inflated negative binomial may be a



better fit than the negative binomial in some cases(Walker et al. 2009). If zeros were not recorded in the dataset then a zero-truncated distribution may be suitable, while if counting stopped at a maximum value — for example no more than 500 *Ascaris* eggs were counted per slide by Cundill et al (2011) — then censored distributions are available(Hilbe 2007). There are yet other distributions which have been little (Grenfell et al. 1990, Shoukri et al. 2004) (Hoshino 2005), if at all, (Dobbie 2001, Massé and Theodorescu 2005, Shmueli et al. 2005) used for parasite data.

*Other methods with potential for greater use*

There are several statistical methods which have not been described above but which have scope for greater use with count data. The bootstrap was notable by its absence in the articles reviewed. This approach is based on re-sampling the data and observing the sampling variation in any outcome of choice, such as a difference or ratio in medians or means(Efron and Tibshirani 1993). For simple comparisons this would seem to have some advantages over non-parametric methods. However, it is not recommended for small sample sizes because the discreteness of the sampling distribution may make the inference unreliable.

The median or other percentile of the counts can be modelled as a function of explanatory variables using quantile regression(Yu et al. 2003). Transformation of the counts may be advisable but, since the usual functions maintain rank order (they are monotonic), this will not affect the interpretation of the results.

In many cases the negative (zero) instances may warrant different treatment, whether for purely statistical reasons or due to clinical or scientific distinctions between them

and the positive group. There is a well-developed literature on models which treat these two groups distinctly within a single overall analysis. These may be referred to as two-part, mixture, or bivariate models (Moulton et al. 2002, Lachenbruch 2002, Hallstrom 2010). When the proportion of zeros is modelled as a function of covariates, then zero-inflated distributions are in this category. The results for the zero and non-zero parts of the model can be presented separately or, depending on the kind of model, it may be possible to synthesize them to report the overall mean (Burton et al. 2003).

Finally, when no single parameter, such as the median or mean, is found to be an adequate summary of the data (Montresor 2007), infection intensity classes (eg negative, mild, moderate, heavy) can be analysed as ordered categories (Agresti 1999).

## **DISCUSSION**

The statistical challenges of count data result not so much from extreme skewness *per se* as from its combination with zero values. If there are no zero values then analysis of log-transformed data by common methods such as the *t* test may be applicable, yielding results in terms of geometric means. However, when zeros are present, the log transformation is not readily applicable, and the geometric mean is zero, complicating the choice of measure of location for count data.

The Williams mean is an attempt to retrieve the geometric mean by applying it to the counts plus one, then subtracting one again. This approach is commonly used in analysis of parasite counts, although more for expediency than for clinical or

biological relevance. Moreover, since it is sometimes misleadingly called the geometric mean, it is often difficult to decide from the text of a paper what was actually done.

There is some resistance to using the arithmetic mean as a measure of location, given that basic statistics teaching usually brands it unsuitable for skewed data.

Nevertheless, in some circumstances the mean is the most relevant outcome measure. In health economics, for example, the mean cost per patient is proportional to the total cost, and for that reason it is more relevant than other measures such as the median (Barber and Thompson 2000). This should not, and need not, be trumped by concerns over difficulty of analysis. Similarly, in parasitology, the arithmetic mean per person may be of interest because, for example, the total community burden in terms of parasite numbers is proportional to the arithmetic mean. Methods are available which model the arithmetic mean while accommodating skewness in the data. The median can also be used, although will be zero if the prevalence is less than 50%. Non-parametric methods are commonly said to assess differences between medians, although this is not generally the case. It is possible to do sample size calculations and statistical analysis in terms of arithmetic means via suitable distributions such as the negative binomial (Alexander et al. 2011).

Choosing the most relevant measure of location will affect how any further analysis is done. Given the powerful statistical software now readily available, it should be possible to choose a method not solely on statistical considerations — finding one whose assumptions are justified — but based on how closely it responds to the original research question (Hand 1994).

The current paper emphasizes generalized linear models (GLMs), in particular with the negative binomial distribution. Fuller explications of GLMs are available elsewhere (Gardner et al. 1995, Wilson and Grenfell 1997, Coxe et al. 2009, McElduff et al. 2010). The negative binomial often provides a good fit to parasite data (Anderson and May 1991) although other distributions may be better for particular datasets (Walker et al. 2009). One review of analysis methods for entomological data recommended negative binomial and quasi-likelihood approaches (Sileshi 2006), while another of aquatic organisms favoured the latter (Noe et al. 2010). One concern with the negative binomial is that, empirically, the dispersion parameter ( $k$ ) is often found to increase with the mean (Alexander et al. 2011), yet basic models assume it to be constant. On the other hand, a simulation study has found likelihood-based analysis to be robust to this (Aban et al. 2008). There are also alternative parameterizations of the negative binomial with different variance-mean dependence (Hilbe 2007). GLM techniques for taking account of clustering can also be used in cohort studies with multiple disease episodes or repeated measures per person, although the issue of time dependence is likely to arise and this is beyond the scope of the current paper.

When describing any numerical information there is a trade-off to be made between the conciseness of any summary measure and the information it conveys. For parasite data, the case has been made for summarizing intensity in the form of categories rather than any single value (Montresor 2007). This is indeed more informative, but less concise. For some purposes, such as clinical trials, a single measure per arm — such as a mean — is likely to be preferred. If categories are used in clinical trials,

they will probably be further summarized into a single measure, such as the proportion with heavy infection.

**Table 1 Descriptive statistics of example datasets.**

	Hookworm eggs	<i>Plasmodium falciparum</i> asexual blood stages
Sample size	1237	477
(male:female)	(603:634)	(247:230)
Mean age in years	26	4.9
(range)	(0-95)	(1-10)
Mass or volume of each sample	two Kato-Katz faecal slides, totalling approximately 1/12g	1 high power field of a thick blood film, approximately 0.001-0.0025 $\mu$ L (Warrell and Gilles 2002)
Percent positive	76	100 <sup>a</sup>
Mean	126	139
Median	22	74
Geometric mean	0 <sup>b</sup>	59
Williams mean	18	61
Range	0 – 4,803	1 - 4,941
Inter-quartile range	1 – 103	23 – 181
Standard deviation	322	266
Reference	Cundhill et al (2011)	(Dunyo et al. 2006)

<sup>a</sup>A positive malaria slide was an entry criterion for the study.

<sup>b</sup>The geometric mean is necessarily zero because the dataset contains zero values (i.e. negative people).

Table 2. Descriptive measures of location for parasite count data used in 89 papers in *Tropical Medicine and International Health*, 2001-2010.

Descriptive measure of location	Number of papers (%) <sup>a</sup>
Arithmetic mean	31 (35)
Geometric mean <sup>b</sup> , or arithmetic mean of logarithms	
zeros absent	25 (28)
zeros present	7 (8)
unclear whether zeros present or not	13 (15)
Williams mean or similar <sup>b,c</sup>	12 (13)
Median	6 (7)
Prevalence by category of infection intensity	20 (22)
Other <sup>d</sup>	1 (1)

<sup>a</sup>The percentages add to more than 100 because some papers used more than one measure.

<sup>b</sup>The term ‘geometric mean’ is sometimes used in these papers to refer to the Williams mean. Unambiguous examples of this, e.g. evidenced by an equation in the paper, are included under ‘Williams mean or similar’. However, it is likely that more of the ‘geometric means’ are, in fact, Williams means, especially where the data included zeros.

<sup>c</sup>Williams mean = (geometric mean of  $(x+\delta)$ )-  $\delta$ , when  $\delta=1$ . Here, ‘similar’ measures include those which: used values of  $\delta$  other than 1; added 1 without then subtracting it; or took the mean of  $\log(x+1)$  without then exponentiating.

<sup>d</sup>The arithmetic mean restricted to parasite-positive individuals.

Table 3. Inferential statistical methods used for parasite count data in 60 papers in *Tropical Medicine and International Health*, 2001-2010.

Inferential Method	Number of papers (%) <sup>a</sup>
Distribution-based:	
Normal (Gaussian) <sup>b</sup>	
on untransformed data	6 (10)
after transformation of the data:	
logarithmic	9 (15)
other	8 (13)
unclear what, if any, transformation was used	9 (15)
Negative binomial	6 (10)
Poisson with allowance for overdispersion	1 (2)
Non-parametric	14 (23)
Other <sup>c</sup>	8 (13)
Unclear	7 (12)

<sup>a</sup>The percentages add to more than 100 because some papers used more than one method.

<sup>b</sup>Including *t* test, analysis of variance (ANOVA), and ordinary least squares regression.

<sup>c</sup>Comprising: ratios of parasite densities as response variable (2);  $\chi^2$  test for trend on infection categories (1); normal distribution-based analysis of village-level indices (1); inference on overlap of confidence intervals (1); Wagstaff index of aggregation of parasites among people (1); logistic regression of high versus low parasite intensity (1); review citing previous analysis (1).



Legend for Figure 1.

Simulated spatial data showing homogeneous and clustered processes in the left and right panels, respectively. The dashed lines divide each area into a 5×5 grid. For the homogenous process, the variance and mean of the 25 counts are similar: 3.8 and 4.0 respectively, consistent with a Poisson distribution. For the clustered process, the variance is much larger than the mean: 19.3 versus 3.8, indicating overdispersion. Although these data are spatial, the same principle applies to counts per person, for example. The data were generated using the ‘rpoispp’ and ‘rMatClust’ functions of the ‘spatstat’ package in the R software.

Legend for Figure 2.

Histogram, on a log scale, of numbers of *Plasmodium falciparum* asexual blood stages on one high power field of a thick blood film in a clinical trial (Dunyo et al. 2006). The dashed line is the fitted normal distribution.

Legend for Figure 3.

Negative binomial regression of faecal egg count on age, the sloped solid line showing the fitted values. The × symbols are the means by age group, with vertical solid lines showing the 95% confidence intervals. Although age groups are shown, the fitted model uses the exact age for each person. The curved dashed lines are the pointwise 95% confidence intervals based on the whole dataset.

Legend for Figure 4.

Histogram of numbers of hookworm eggs counted on two Kato-Katz slides per person in the baseline survey of a prospective study in Brazil. The axis labels show the

original values although the scale has been eighth-root transformed. The dashed line is the fitted negative binomial distribution and the dotted line the fitted Poisson.

## **Acknowledgements**

This work was supported financially by United Kingdom Medical Research Council grant number G7508177 to the Tropical Epidemiology Group, and by the Human Hookworm Vaccine Initiative (HHVI) of the Sabin Vaccine Institute, which receives funding from the Bill and Melinda Gates Foundation. The studies which generated the hookworm and malaria datasets presented here were supported respectively by 1) the HHVI and 2) the Wellcome Trust (Project 061910), the Gates Malaria Partnership, and the United Kingdom Medical Research Council. I am grateful to Christian Bottomley, Paul Milligan and an anonymous referee for useful comments on the manuscript.

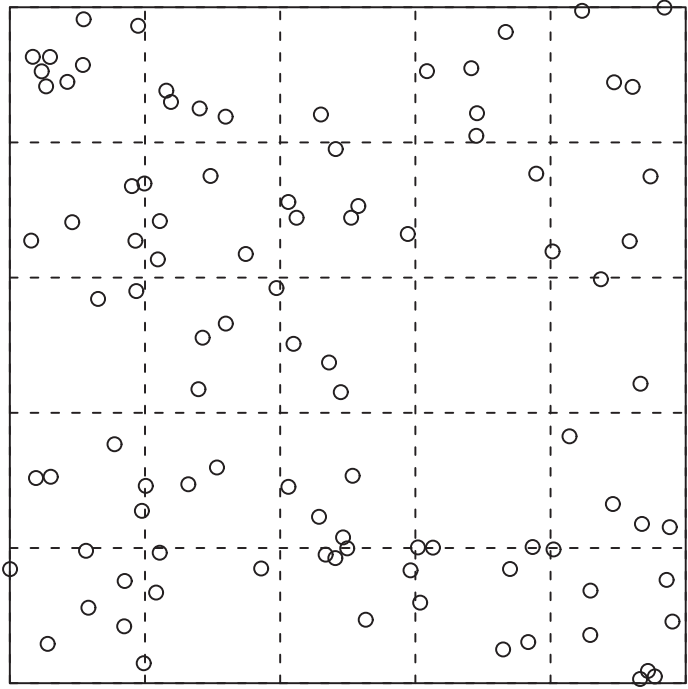
## References-

- Aban, I. B., Cutter, G. R. & Mavinga, N. (2008) Inferences and Power Analysis Concerning Two Negative Binomial Distributions with An Application to MRI Lesion Counts Data. *Comput Stat Data Anal*, 53, 820-833.
- Agresti, A. (1999) Modelling ordered categorical data: recent advances and future challenges. *Statistics in Medicine*, 18, 2191-2208.
- Alexander, N., Bethony, J., Corrêa-Oliveira, R., Rodrigues, L. C., Hotez, P. & Brooker, S. (2007) Repeatability of paired counts. *Statistics in Medicine*, 26, 3566 - 3577.
- Alexander, N., Cundill, B., Sabatelli, L., *et al.* (2011) Selection and quantification of infection endpoints for trials of vaccines against intestinal helminths. *Vaccine*, 29, 3686-94.
- Alexander, N. D., Solomon, A. W., Holland, M. J., *et al.* (2005) An index of community ocular *Chlamydia trachomatis* load for control of trachoma. *Trans R Soc Trop Med Hyg*, 99, 175-7.
- Anderson, R. M. & May, R. M. (1991) *Infectious Diseases of Humans: Dynamics and Control*. 1st edn. Oxford University Press Oxford.
- Anscombe, F. J. (1948) The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35, 246-254.
- Barber, J. A. & Thompson, S. G. (2000) Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Statistics in Medicine*, 19, 3219-3236.
- Barker, L. & Cadwell, B. L. (2008) An analysis of eight 95 per cent confidence intervals for a ratio of Poisson parameters when events are rare. *Statistics in Medicine*, 27, 4030 - 4037.
- Bauer, D. F. (1972) Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67, 687-690.
- Beall, G. (1942) The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*, 32, 243-262.
- Bland, M. (1995) *An Introduction to Medical Statistics*. 2nd edn. Oxford University Press Oxford.
- Bockarie, M., Kazura, J., Alexander, N., *et al.* (1996) Transmission dynamics of *Wuchereria bancrofti* in East Sepik Province, Papua New Guinea. *American Journal of Tropical Medicine and Hygiene*, 54, 577-581.
- Boneau, C. A. (1960) The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57, 49-64.
- Bonferroni, C. E. (1950) Sulle medie multiple di potenze [On multiple algebraic means]. *Bollettino dell'Unione Matematica Italiana, serie 3*, 5, 267-270.
- Borowski, E. J. & Borwein, J. M. (1989) *Dictionary of Mathematics*. 1st edn. HarperCollins Publishers London.
- Burton, M. J., Holland, M. J., Faal, N., *et al.* (2003) Which members of a community need antibiotics to control trachoma? Conjunctival *Chlamydia trachomatis* infection load in Gambian villages. *Investigative Ophthalmology and Visual Science*, 44, 4215-4222.
- Bushman, B. J. & Wang, M. C. (1995) A procedure for combining sample correlation coefficients and vote counts to obtain an estimate and a confidence interval for the population correlation coefficient. *Psychological Bulletin*, 117, 530-546.
- Conover, W. J. (1980) *Practical Nonparametric Statistics*. 2nd edn. John Wiley & Sons New York.

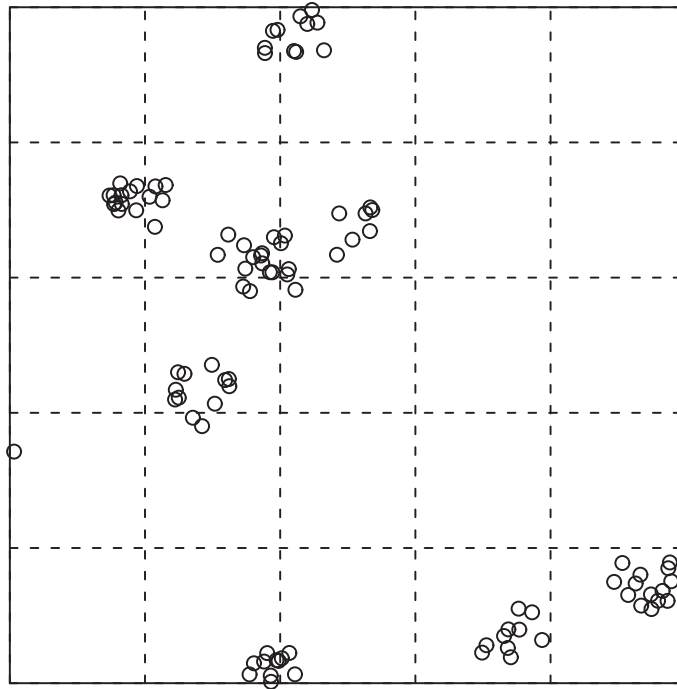
- Coxe, S., West, S. G. & Aiken, L. S. (2009) The analysis of count data: a gentle introduction to Poisson regression and its alternatives. *J Pers Assess*, 91, 121-36.
- Cundill, B., Alexander, N., Bethony, J., Diemert, D., Pullan, R. L. & Brooker, S. (2011) Rates and intensity of re-infection with human helminths after treatment and the influence of individual, household, and environmental factors in a Brazilian community. *Parasitology*, 138, 1406-16.
- Dobbie, M. J. (2001) Modelling correlated zero-inflated count data. Canberra, Australian National University.
- Dobson, R. J., Sangster, N. C., Besier, R. B. & Woodgate, R. G. (2009) Geometric means provide a biased efficacy result when conducting a faecal egg count reduction test (FECRT). *Vet Parasitol*, 161, 162-7.
- Dunyo, S., Ord, R., Hallett, R., *et al.* (2006) Randomised trial of chloroquine/sulphadoxine-pyrimethamine in Gambian children with malaria: impact against multidrug-resistant *P. falciparum*. *PLoS Clin Trials*, 1, e14.
- Efron, B. & Tibshirani, R. (1993) *An Introduction to the Bootstrap*. 1st edn. Chapman and Hall New York.
- Elliott, J. M. (1977) *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*. 2nd edn. Freshwater Biological Association Ambleside.
- Everitt, B. (1995) *Cambridge Dictionary of Statistics in the Medical Sciences*. edn. Cambridge University Press Cambridge.
- Fulford, A. J. C. (1994) Dispersion and bias: can we trust geometric means? *Parasitology Today*, 10, 446-448.
- Gardner, W., Mulvey, E. P. & Shaw, E. C. (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118, 392-404.
- Grenfell, B. T., Das, P. K., Rajagopalan, P. K. & Bundy, D. A. P. (1990) Frequency distribution of lymphatic filariasis microfilariae in human populations: population processes and statistical estimation. *Parasitology*, 101, 417-427.
- Hallstrom, A. P. (2010) A modified Wilcoxon test for non-negative distributions with a clump of zeros. *Stat Med*, 29, 391-400.
- Hand, D. J. (1994) Deconstructing statistical questions (with discussion). *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 157, 317-356.
- Hart, A. (2001) Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ*, 323, 391-3.
- Heeren, T. & d'Agostino, R. (1987) Robustness of the two-independent samples *t*-test when applies to ordinal scale data. *Statistics in Medicine*, 6, 79-90.
- Hilbe, J. M. (2007) *Negative Binomial Regression*. 1st edn. Cambridge University Press Cambridge.
- Hoshino, N. (2005) Engen's extended negative binomial model revisited. *Annals of the Institute of Statistical Mathematics*, 57, 369-387.
- Kirkwood, B. R. & Sterne, J. A. C. (2003) *Essentials of Medical Statistics*. 2nd edn. Blackwell Scientific Publications Oxford.
- Lachenbruch, P. A. (2002) Analysis of data with excess zeros. *Statistical Methods in Medical Research*, 11, 297-302.
- Laubscher, N. F. (1961) On stabilizing the binomial and negative binomial variance. *Journal of the American Statistical Association*, 56, 143-150.
- Marshall, T. F., Anderson, J. & Fuglsang, H. (1986) The incidence of eye lesions and visual impairment in onchocerciasis in relationship to the intensity of

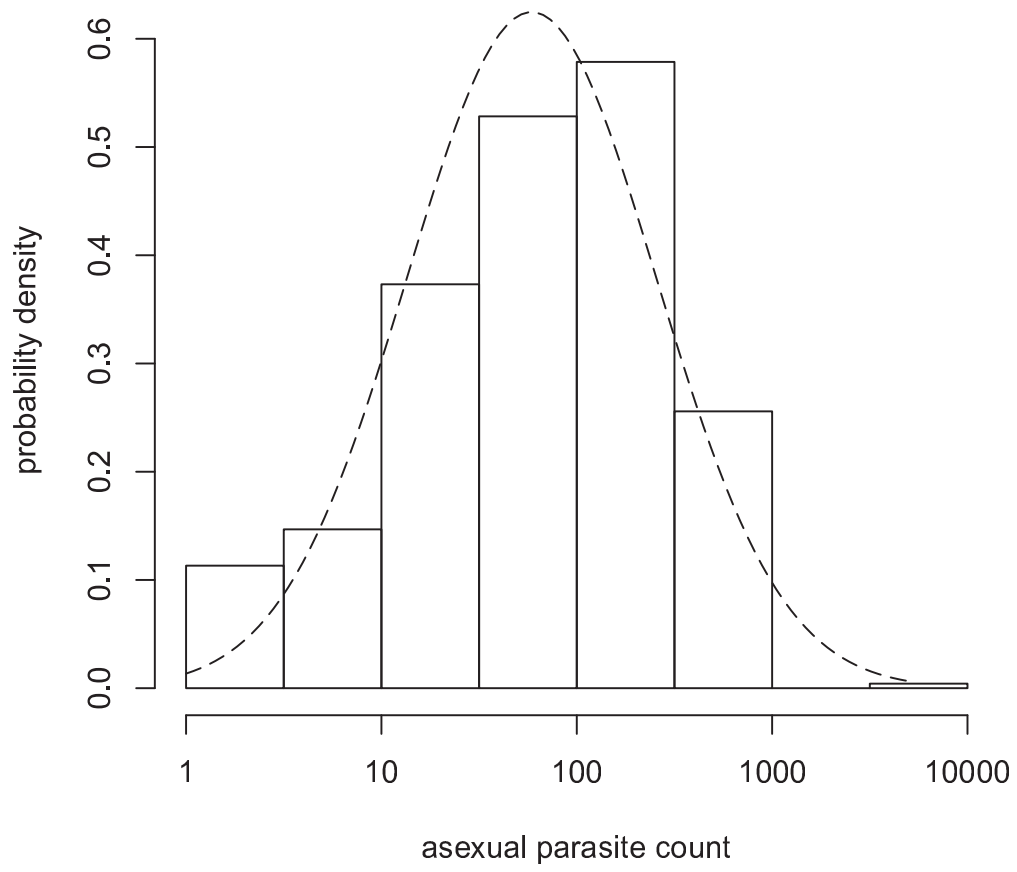
- infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 80, 426-434.
- Massé, J. C. & Theodorescu, R. (2005) Neyman type A distribution revisited. *Statistica Neerlandica*, 59, 206–213.
- McCullagh, P. & Nelder, J. A. (1983) *Generalized Linear Models*. 1st edn. Chapman and Hall London.
- McElduff, F., Cortina-Borja, M., Chan, S.-K. & Wade, A. (2010) When *t*-tests or Wilcoxon-Mann-Whitney tests won't do. *Advances in Physiology Education*, 34, 128-133.
- Montresor, A. (2007) Arithmetic or geometric means of eggs per gram are not appropriate indicators to estimate the impact of control measures in helminth infections. *Trans R Soc Trop Med Hyg*, 101, 773-6.
- Moulton, L. H., Curriero, F. C. & Barroso, P. F. (2002) Mixture models for quantitative HIV RNA data. *Statistical Methods in Medical Research*, 11, 317-325.
- Mumpower, J. L. & McClelland, G. (2002) Measurement error, skewness, and risk analysis: coping with the long tail of the distribution. *Risk Anal*, 22, 277-90.
- Mwangi, T. W., Fegan, G., Williams, T. N., Kinyanjui, S. M., Snow, R. W. & Marsh, K. (2008) Evidence for over-dispersion in the distribution of clinical malaria episodes in children. *PLoS ONE*, 3, e2196.
- Noe, D. A., Bailer, A. J. & Noble, R. B. (2010) Comparing methods for analyzing overdispersed count data in aquatic toxicology. *Environ Toxicol Chem*, 29, 212-9.
- O'Hara, R. B. & Kotze, D. J. (2010) Do not log-transform count data. *Methods in Ecology & Evolution*, 1, 118-122.
- Olayinka, O. S. & Abdullahi, S. A. (2009) An overview of industrial employees' exposure to noise in sundry processing and manufacturing industries in Ilorin metropolis, Nigeria. *Ind Health*, 47, 123-33.
- Remme, J., Ba, O., Dadzie, K. Y. & Karam, M. (1986) A force-of-infection model for onchocerciasis and its applications in the epidemiological evaluation of the Onchocerciasis Control Programme in the Volta River basin area. *Bulletin of the World Health Organization*, 64, 667-681.
- Seto, E. Y., Lee, Y. J., Liang, S. & Zhong, B. (2007) Individual and village-level study of water contact patterns and *Schistosoma japonicum* infection in mountainous rural China. *Trop Med Int Health*, 12, 1199-209.
- Shmueli, G., Minka, T., Kadane, J. B., Borle, S. & Boatwright, P. B. (2005) A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 127–142.
- Shoukri, M. M., Asyali, M. H., VanDorp, R. & Kelton, D. (2004) The Poisson inverse Gaussian regression model in the analysis of clustered counts data. *Journal of Data Science*, 2, 17-32.
- Sileshi, G. (2006) Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bull Entomol Res*, 96, 479-88.
- Smith, T., Armstrong Schellenberg, J. & Hayes, R. (1994) Attributable fraction estimates and case definitions for malaria in endemic areas. *Statistics in Medicine*, 13, 2345-2358.
- Stonehouse, J. M. & Forrester, G. J. (1998) Robustness of the *t* and *U* tests under combined assumption violations. *Journal of Applied Statistics*, 25, 63-74.

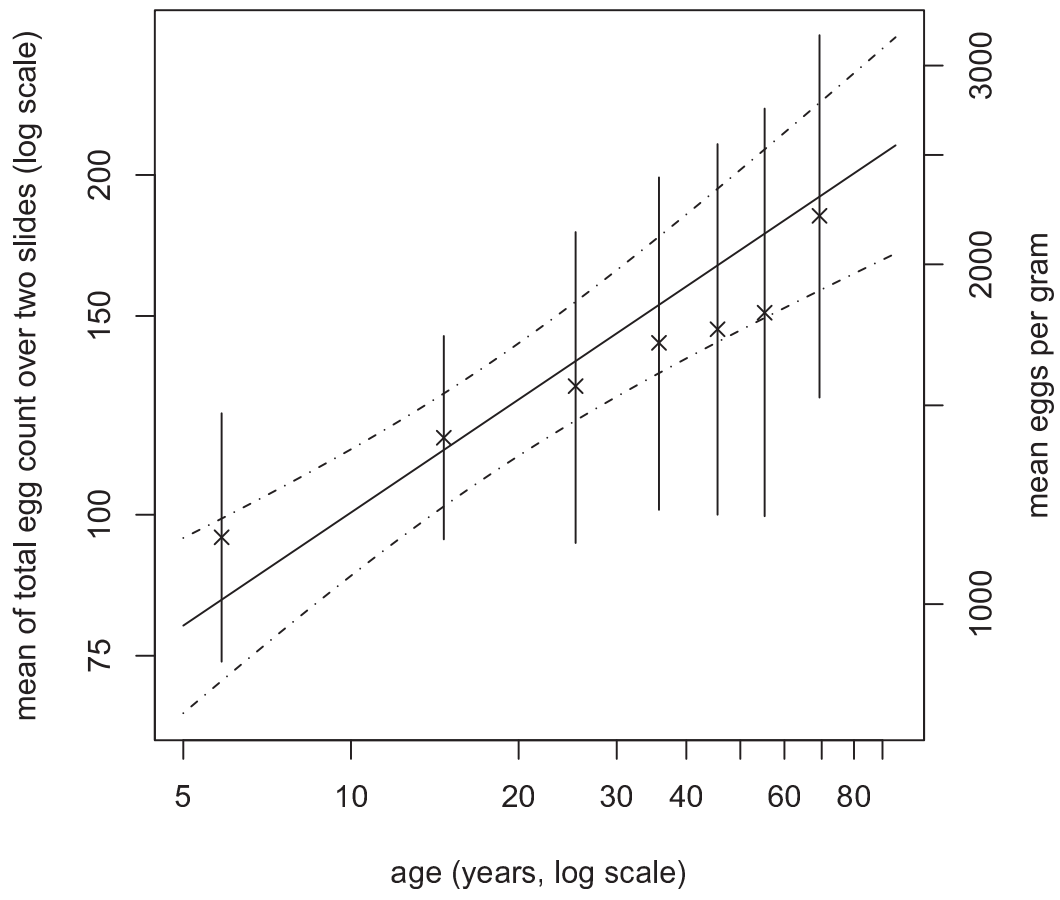
- Walker, M., Hall, A., Anderson, R. M. & Basanez, M. G. (2009) Density-dependent effects on the weight of female *Ascaris lumbricoides* infections of humans and its impact on patterns of egg production. *Parasit Vectors*, 2, 11.
- Warrell, D. M. & Gilles, H. M. (Eds.) (2002) *Essential Malariology*, London, Arnold.
- White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.
- Williams, C. B. (1937) The use of logarithms in the interpretation of certain entomological problems. *Annals of Applied Biology*, 24, 404-414.
- Wilson, K. & Grenfell, B. T. (1997) Generalized linear modelling for parasitologists. *Parasitology Today*, 13, 33-38.
- Wojtczak, M. & Viemeister, N. F. (2008) Perception of suprathreshold amplitude modulation and intensity increments: Weber's law revisited. *J Acoust Soc Am*, 123, 2220-36.
- Yu, K., Lu, Z. & Stander, J. (2003) Quantile regression: applications and current research areas. *Statistician*, 52, 331-350.

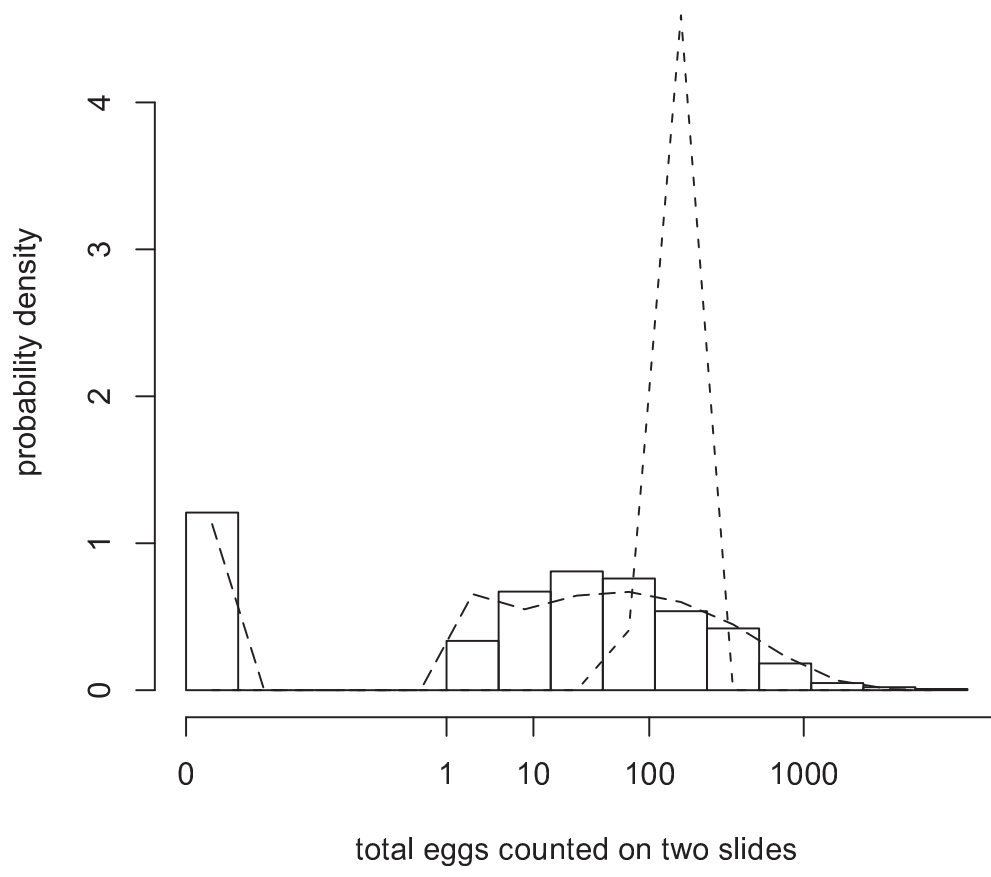












## Papers Included in the Literature Review

- Alker, A. P., Kazadi, W. M., Kutelemani, A. K., Bloland, P. B., Tshetu, A. K. & Meshnick, S. R. (2008) *dhfr* and *dhps* genotype and sulfadoxine-pyrimethamine treatment failure in children with falciparum malaria in the Democratic Republic of Congo. *Trop Med Int Health*, **13**, 1384-91.
- Allam, A. F., Kader, O., Zaki, A., Shehab, A. Y. & Farag, H. F. (2009) Assessing the marginal error in diagnosis and cure of *Schistosoma mansoni* in areas of low endemicity using Percoll and PCR techniques. *Trop Med Int Health*, **14**, 316-21.
- Appawu, M. A., Dadzie, S. K., Baffoe-Wilmot, A. & Wilson, M. D. (2001) Lymphatic filariasis in Ghana: entomological investigation of transmission dynamics and intensity in communities served by irrigation systems in the Upper East Region of Ghana. *Trop Med Int Health*, **6**, 511-6.
- Audibert, M., Mathonnat, J. & Henry, M. C. (2003) Malaria and property accumulation in rice production systems in the savannah zone of Cote d'Ivoire. *Trop Med Int Health*, **8**, 471-83.
- Avila, J. C., Villaroel, R., Marquino, W., Zegarra, J., Mollinedo, R. & Ruebush, T. K. (2004) Efficacy of mefloquine and mefloquine-artesunate for the treatment of uncomplicated *Plasmodium falciparum* malaria in the Amazon region of Bolivia. *Trop Med Int Health*, **9**, 217-21.
- Berhe, N., Geitung, J. T., Medhin, G. & Gundersen, S. G. (2006) Large scale evaluation of WHO's ultrasonographic staging system of schistosomal periportal fibrosis in Ethiopia. *Trop Med Int Health*, **11**, 1286-94.
- Bethony, J., Williams, J. T., Kloos, H., *et al.* (2001) Exposure to *Schistosoma mansoni* infection in a rural area in Brazil. II: household risk factors. *Trop Med Int Health*, **6**, 136-45.
- Black, C. L., Steinauer, M. L., Mwinzi, P. N., Evan Secor, W., Karanja, D. M. & Colley, D. G. (2009) Impact of intense, longitudinal retreatment with praziquantel on cure rates of schistosomiasis mansoni in a cohort of occupationally exposed adults in western Kenya. *Trop Med Int Health*, **14**, 450-7.

- Boisier, P., Ramarokoto, C. E., Ravoniarimbina, P., Rabarijaona, L. & Ravaoalimalala, V. E. (2001) Geographic differences in hepatosplenic complications of schistosomiasis mansoni and explanatory factors of morbidity. *Trop Med Int Health*, **6**, 699-706.
- Bonnet, S., Paul, R. E., Gouagna, C., *et al.* (2002) Level and dynamics of malaria transmission and morbidity in an equatorial area of South Cameroon. *Trop Med Int Health*, **7**, 249-56.
- Bosompem, K. M., Bentum, I. A., Otchere, J., *et al.* (2004) Infant schistosomiasis in Ghana: a survey in an irrigation community. *Trop Med Int Health*, **9**, 917-22.
- Bousema, J. T., Gouagna, L. C., Meutstege, A. M., *et al.* (2003) Treatment failure of pyrimethamine-sulphadoxine and induction of *Plasmodium falciparum* gametocytaemia in children in western Kenya. *Trop Med Int Health*, **8**, 427-30.
- Cox, S. E., Staalsoe, T., Arthur, P., *et al.* (2005) Maternal vitamin A supplementation and immunity to malaria in pregnancy in Ghanaian primigravids. *Trop Med Int Health*, **10**, 1286-97.
- Cramer, J. P., Nussler, A. K., Ehrhardt, S., *et al.* (2005) Age-dependent effect of plasma nitric oxide on parasite density in Ghanaian children with severe malaria. *Trop Med Int Health*, **10**, 672-80.
- Critchley, J., Addiss, D., Ejere, H., Gamble, C., Garner, P. & Gelband, H. (2005) Albendazole for the control and elimination of lymphatic filariasis: systematic review. *Trop Med Int Health*, **10**, 818-25.
- Cunin, P., Tchuem Tchuente, L. A., Poste, B., Djibrilla, K. & Martin, P. M. (2003) Interactions between *Schistosoma haematobium* and *Schistosoma mansoni* in humans in north Cameroon. *Trop Med Int Health*, **8**, 1110-7.
- Drakeley, C. J., Jawara, M., Targett, G. A., *et al.* (2004) Addition of artesunate to chloroquine for treatment of *Plasmodium falciparum* malaria in Gambian children causes a significant but short-lived reduction in infectiousness for mosquitoes. *Trop Med Int Health*, **9**, 53-61.
- Durrani, N., Leslie, T., Rahim, S., Graham, K., Ahmad, F. & Rowland, M. (2005) Efficacy of combination therapy with artesunate plus amodiaquine compared to monotherapy with chloroquine, amodiaquine or sulfadoxine-pyrimethamine for

- treatment of uncomplicated *Plasmodium falciparum* in Afghanistan. *Trop Med Int Health*, **10**, 521-9.
- Ensink, J. H., Mahmood, T. & Dalsgaard, A. (2007) Wastewater-irrigated vegetables: market handling versus irrigation water quality. *Trop Med Int Health*, **12 Suppl 2**, 2-7.
- Esteban, J. G., Gonzalez, C., Bargues, M. D., *et al.* (2002) High fascioliasis infection in children linked to a man-made irrigation zone in Peru. *Trop Med Int Health*, **7**, 339-48.
- Falade, C., Mokuolu, O., Okafor, H., *et al.* (2007) Epidemiology of congenital malaria in Nigeria: a multi-centre study. *Trop Med Int Health*, **12**, 1279-87.
- Fanello, C. I., Karema, C., van Doren, W., Rwagacondo, C. E. & D'Alessandro, U. (2006) Tolerability of amodiaquine and sulphadoxine-pyrimethamine, alone or in combination for the treatment of uncomplicated *Plasmodium falciparum* malaria in Rwandan adults. *Trop Med Int Health*, **11**, 589-96.
- Feldmeier, H., Helling-Giese, G. & Poggensee, G. (2001) Unreliability of PAP smears to diagnose female genital schistosomiasis. *Trop Med Int Health*, **6**, 31-3.
- Fleming, F. M., Brooker, S., Geiger, S. M., *et al.* (2006) Synergistic associations between hookworm and other helminth species in a rural community in Brazil. *Trop Med Int Health*, **11**, 56-64.
- Gilgen, D. D., Mascie-Taylor, C. G. & Rosetta, L. L. (2001) Intestinal helminth infections, anaemia and labour productivity of female tea pluckers in Bangladesh. *Trop Med Int Health*, **6**, 449-57.
- Gryseels, B., Mbaye, A., De Vlas, S. J., *et al.* (2001) Are poor responses to praziquantel for the treatment of *Schistosoma mansoni* infections in Senegal due to resistance? An overview of the evidence. *Trop Med Int Health*, **6**, 864-73.
- Hanelt, B., Steinauer, M. L., Mwangi, I. N., *et al.* (2009) A new approach to characterize populations of *Schistosoma mansoni* from humans: development and assessment of microsatellite analysis of pooled miracidia. *Trop Med Int Health*, **14**, 322-31.
- Henry, M. C., Rogier, C., Nzeyimana, I., *et al.* (2003) Inland valley rice production systems and malaria infection and disease in the savannah of Cote d'Ivoire. *Trop Med Int Health*, **8**, 449-58.

- Idro, R., Aloyo, J., Mayende, L., Bitarakwate, E., John, C. C. & Kivumbi, G. W. (2006) Severe malaria in children in areas with low, moderate and high transmission intensity in Uganda. *Trop Med Int Health*, **11**, 115-24.
- Jardim-Botelho, A., Raff, S., Rodrigues Rde, A., *et al.* (2008) Hookworm, *Ascaris lumbricoides* infection and polyparasitism associated with poor cognitive performance in Brazilian schoolchildren. *Trop Med Int Health*, **13**, 994-1004.
- Jukes, M. C., Nokes, C. A., Alcock, K. J., *et al.* (2002) Heavy schistosomiasis associated with poor short-term memory and slower reaction times in Tanzanian schoolchildren. *Trop Med Int Health*, **7**, 104-17.
- Kabatereine, N. B., Brooker, S., Tukahebwa, E. M., Kazibwe, F. & Onapa, A. W. (2004) Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. *Trop Med Int Health*, **9**, 372-80.
- Keraita, B., Konradsen, F., Drechsel, P. & Abaidoo, R. C. (2007) Effect of low-cost irrigation methods on microbial contamination of lettuce irrigated with untreated wastewater. *Trop Med Int Health*, **12 Suppl 2**, 15-22.
- Keraita, B., Konradsen, F., Drechsel, P. & Abaidoo, R. C. (2007) Reducing microbial contamination on wastewater-irrigated lettuce by cessation of irrigation before harvesting. *Trop Med Int Health*, **12 Suppl 2**, 8-14.
- Kongs, A., Marks, G., Verle, P. & Van der Stuyft, P. (2001) The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections. *Trop Med Int Health*, **6**, 163-9.
- Larocque, R., Casapia, M., Gotuzzo, E., *et al.* (2006) A double-blind randomized controlled trial of antenatal mebendazole to reduce low birthweight in a hookworm-endemic area of Peru. *Trop Med Int Health*, **11**, 1485-95.
- Luoba, A. I., Wenzel Geissler, P., Estambale, B., *et al.* (2005) Earth-eating and reinfection with intestinal helminths among pregnant and lactating women in western Kenya. *Trop Med Int Health*, **10**, 220-7.
- Maguire, J. D., Tuti, S., Sismadi, P., *et al.* (2005) Endemic coastal malaria in the Thousand Islands District, near Jakarta, Indonesia. *Trop Med Int Health*, **10**, 489-96.



- Mani, T. R., Rajendran, R., Munirathinam, A., *et al.* (2002) Efficacy of co-administration of albendazole and diethylcarbamazine against geohelminthiasis: a study from South India. *Trop Med Int Health*, **7**, 541-8.
- Mani, T. R., Rajendran, R., Sunish, I. P., *et al.* (2004) Effectiveness of two annual, single-dose mass drug administrations of diethylcarbamazine alone or in combination with albendazole on soil-transmitted helminthiasis in filariasis elimination programme. *Trop Med Int Health*, **9**, 1030-5.
- Marchant, T., Schellenberg, J. A., Edgar, T., *et al.* (2002) Socially marketed insecticide-treated nets improve malaria and anaemia in pregnancy in southern Tanzania. *Trop Med Int Health*, **7**, 149-58.
- Mas, J., Ascaso, C., Escaramis, G., *et al.* (2006) Reduction in the prevalence and intensity of infection in *Onchocerca volvulus* microfilariae according to ethnicity and community after 8 years of ivermectin treatment on the island of Bioko, Equatorial Guinea. *Trop Med Int Health*, **11**, 1082-91.
- Mayor, A., Saute, F., Aponte, J. J., *et al.* (2003) *Plasmodium falciparum* multiple infections in Mozambique, its relation to other malariological indices and to prospective risk of malaria morbidity. *Trop Med Int Health*, **8**, 3-11.
- Mens, P., Spieker, N., Omar, S., Heijnen, M., Schallig, H. & Kager, P. A. (2007) Is molecular biology the best alternative for diagnosis of malaria to microscopy? A comparison between microscopy, antigen detection and molecular tests in rural Kenya and urban Tanzania. *Trop Med Int Health*, **12**, 238-44.
- Meyrowitsch, D. W. & Simonsen, P. E. (2001) Efficacy of DEC against *Ascaris* and hookworm infections in schoolchildren. *Trop Med Int Health*, **6**, 739-42.
- Mockenhaupt, F. P., Eggelte, T. A., Till, H. & Bienzle, U. (2001) *Plasmodium falciparum* *pfprt* and *pfmdr1* polymorphisms are associated with the *pfdhfr* N108 pyrimethamine-resistance mutation in isolates from Ghana. *Trop Med Int Health*, **6**, 749-55.
- Mockenhaupt, F. P., Ehrhardt, S., Dzisi, S. Y., *et al.* (2005) A randomized, placebo-controlled, double-blind trial on sulfadoxine-pyrimethamine alone or combined with artesunate or amodiaquine in uncomplicated malaria. *Trop Med Int Health*, **10**, 512-20.

- Montresor, A., Ramsan, M., Chwaya, H. M., *et al.* (2001) School enrollment in Zanzibar linked to children's age and helminth infections. *Trop Med Int Health*, **6**, 227-31.
- Montresor, A., Zin, T. T., Padmasiri, E., Allen, H. & Savioli, L. (2004) Soil-transmitted helminthiasis in Myanmar and approximate costs for countrywide control. *Trop Med Int Health*, **9**, 1012-5.
- Mugittu, K., Adjuik, M., Snounou, G., *et al.* (2006) Molecular genotyping to distinguish between recrudescents and new infections in treatment trials of Plasmodium falciparum malaria conducted in Sub-Saharan Africa: adjustment of parasitological outcomes and assessment of genotyping effectiveness. *Trop Med Int Health*, **11**, 1350-9.
- Muhumuza, S., Kitimbo, G., Oryema-Lalobo, M. & Nuwaha, F. (2009) Association between socio economic status and schistosomiasis infection in Jinja District, Uganda. *Trop Med Int Health*, **14**, 612-9.
- Naus, C. W., Booth, M., Jones, F. M., *et al.* (2003) The relationship between age, sex, egg-count and specific antibody responses against Schistosoma mansoni antigens in a Ugandan fishing community. *Trop Med Int Health*, **8**, 561-8.
- Nebie, I., Tiono, A. B., Diallo, D. A., *et al.* (2008) Do antibody responses to malaria vaccine candidates influenced by the level of malaria transmission protect from malaria? *Trop Med Int Health*, **13**, 229-37.
- N'Goran, E. K., Utzinger, J., N'Guessan, A. N., *et al.* (2001) Reinfection with Schistosoma haematobium following school-based chemotherapy with praziquantel in four highly endemic villages in Cote d'Ivoire. *Trop Med Int Health*, **6**, 817-25.
- Nguansangiam, S., Day, N. P., Hien, T. T., *et al.* (2007) A quantitative ultrastructural study of renal pathology in fatal Plasmodium falciparum malaria. *Trop Med Int Health*, **12**, 1037-50.
- Nguyen, T. H., Nguyen, V. D., Murrell, D. & Dalsgaard, A. (2007) Occurrence and species distribution of fishborne zoonotic trematodes in wastewater-fed aquaculture in northern Vietnam. *Trop Med Int Health*, **12 Suppl 2**, 66-72.

- Njama-Meya, D., Kanya, M. R. & Dorsey, G. (2004) Asymptomatic parasitaemia as a risk factor for symptomatic malaria in a cohort of Ugandan children. *Trop Med Int Health*, **9**, 862-8.
- Oladejo, S. O. & Ofoezie, I. E. (2006) Unabated schistosomiasis transmission in Erinle River Dam, Osun State, Nigeria: evidence of neglect of environmental effects of development projects. *Trop Med Int Health*, **11**, 843-50.
- Ouedraogo, A., Tiono, A. B., Diarra, A., Nebie, I. O., Konate, A. T. & Sirima, S. B. (2010) The effects of a pre-season treatment with effective antimalarials on subsequent malaria morbidity in under five-year-old children living in high and seasonal malaria transmission area of Burkina Faso. *Trop Med Int Health*, **15**, 1315-21.
- Owusu-Agyei, S., Binka, F., Koram, K., *et al.* (2002) Does radical cure of asymptomatic *Plasmodium falciparum* place adults in endemic areas at increased risk of recurrent symptomatic malaria? *Trop Med Int Health*, **7**, 599-603.
- Owusu-Agyei, S., Smith, T., Beck, H. P., Amenga-Etego, L. & Felger, I. (2002) Molecular epidemiology of *Plasmodium falciparum* infections among asymptomatic inhabitants of a holoendemic malarious area in northern Ghana. *Trop Med Int Health*, **7**, 421-8.
- Peyerl-Hoffmann, G., Jelinek, T., Kilian, A., Kabagambe, G., Metzger, W. G. & von Sonnenburg, F. (2001) Genetic diversity of *Plasmodium falciparum* and its relationship to parasite density in an area with different malaria endemicities in West Uganda. *Trop Med Int Health*, **6**, 607-13.
- Pfeiffer, K., Some, F., Muller, O., *et al.* (2008) Clinical diagnosis of malaria and the risk of chloroquine self-medication in rural health centres in Burkina Faso. *Trop Med Int Health*, **13**, 418-26.
- Pischke, S., Buttner, D. W., Liebau, E. & Fischer, P. (2002) An internal control for the detection of *Onchocerca volvulus* DNA by PCR-ELISA and rapid detection of specific PCR products by DNA Detection Test Strips. *Trop Med Int Health*, **7**, 526-31.

- Polman, K., Stelma, F. F., De Vlas, S. J., *et al.* (2001) Dynamics of egg counts and circulating antigen levels in a recent *Schistosoma mansoni* focus in northern Senegal. *Trop Med Int Health*, **6**, 538-44.
- Rajendran, R., Sunish, I. P., Mani, T. R., *et al.* (2006) Community-based study to assess the efficacy of DEC plus ALB against DEC alone on bancroftian filarial infection in endemic areas in Tamil Nadu, south India. *Trop Med Int Health*, **11**, 851-61.
- Ramaiah, K. D., Das, P. K., Vanamail, P. & Pani, S. P. (2003) The impact of six rounds of single-dose mass administration of diethylcarbamazine or ivermectin on the transmission of *Wuchereria bancrofti* by *Culex quinquefasciatus* and its implications for lymphatic filariasis elimination programmes. *Trop Med Int Health*, **8**, 1082-92.
- Ramaiah, K. D., Vanamail, P., Pani, S. P., Yuvaraj, J. & Das, P. K. (2002) The effect of six rounds of single dose mass treatment with diethylcarbamazine or ivermectin on *Wuchereria bancrofti* infection and its implications for lymphatic filariasis elimination. *Trop Med Int Health*, **7**, 767-74.
- Ranjit, M. R., Das, A., Chhotray, G. P., Das, B. P., Das, B. N. & Acharya, A. S. (2004) The PfCRT (K76T) point mutation favours clone multiplicity and disease severity in *Plasmodium falciparum* infection. *Trop Med Int Health*, **9**, 857-61.
- Raso, G., Utzinger, J., Silue, K. D., *et al.* (2005) Disparities in parasitic infections, perceived ill health and access to health care among poorer and less poor schoolchildren of rural Cote d'Ivoire. *Trop Med Int Health*, **10**, 42-57.
- Saathoff, E., Olsen, A., Kvalsvig, J. D., Appleton, C. C., Sharp, B. & Kleinschmidt, I. (2005) Ecological covariates of *Ascaris lumbricoides* infection in schoolchildren from rural KwaZulu-Natal, South Africa. *Trop Med Int Health*, **10**, 412-22.
- Saute, F., Menendez, C., Mayor, A., *et al.* (2002) Malaria in pregnancy in rural Mozambique: the role of parity, submicroscopic and multiple *Plasmodium falciparum* infections. *Trop Med Int Health*, **7**, 19-28.
- Scott, J. T., Diakhate, M., Vereecken, K., *et al.* (2003) Human water contacts patterns in *Schistosoma mansoni* epidemic foci in northern Senegal change according to age, sex and place of residence, but are not related to intensity of infection. *Trop Med Int Health*, **8**, 100-8.

- Seto, E. Y., Lee, Y. J., Liang, S. & Zhong, B. (2007) Individual and village-level study of water contact patterns and *Schistosoma japonicum* infection in mountainous rural China. *Trop Med Int Health*, **12**, 1199-209.
- Shekalaghe, S. A., Bousema, J. T., Kunei, K. K., *et al.* (2007) Submicroscopic *Plasmodium falciparum* gametocyte carriage is common in an area of low and seasonal transmission in Tanzania. *Trop Med Int Health*, **12**, 547-53.
- Sowunmi, A. & Fateye, B. A. (2003) *Plasmodium falciparum* gametocytaemia in Nigerian children: before, during and after treatment with antimalarial drugs. *Trop Med Int Health*, **8**, 783-92.
- Sunish, I. P., Rajendran, R., Mani, T. R., *et al.* (2003) Transmission intensity index to monitor filariasis infection pressure in vectors for the evaluation of filariasis elimination programmes. *Trop Med Int Health*, **8**, 812-9.
- Sunish, I. P., Rajendran, R., Mani, T. R., *et al.* (2002) Resurgence in filarial transmission after withdrawal of mass drug administration and the relationship between antigenaemia and microfilaraemia--a longitudinal study. *Trop Med Int Health*, **7**, 59-69.
- Supali, T., Ismid, I. S., Ruckert, P. & Fischer, P. (2002) Treatment of *Brugia timori* and *Wuchereria bancrofti* infections in Indonesia using DEC or a combination of DEC and albendazole: adverse reactions and short-term effects on microfilariae. *Trop Med Int Health*, **7**, 894-901.
- Tchuem Tchuente, L. A., Behnke, J. M., Gilbert, F. S., Southgate, V. R. & Vercruyssen, J. (2003) Polyparasitism with *Schistosoma haematobium* and soil-transmitted helminth infections among school children in Loum, Cameroon. *Trop Med Int Health*, **8**, 975-86.
- Traub, R. J., Robertson, I. D., Irwin, P., Mencke, N. & Andrew Thompson, R. C. (2004) The prevalence, intensities and risk factors associated with geohelminth infection in tea-growing communities of Assam, India. *Trop Med Int Health*, **9**, 688-701.
- Van den Bossche, P., Ky-Zerbo, A., Brandt, J., Marcotty, T., Geerts, S. & De Deken, R. (2005) Transmissibility of *Trypanosoma brucei* during its development in cattle. *Trop Med Int Health*, **10**, 833-9.

- Vennervald, B. J., Kenty, L., Butterworth, A. E., *et al.* (2004) Detailed clinical and ultrasound examination of children and adolescents in a *Schistosoma mansoni* endemic area in Kenya: hepatosplenic disease in the absence of portal fibrosis. *Trop Med Int Health*, **9**, 461-70.
- Vereecken, K., Naus, C. W., Polman, K., *et al.* (2007) Associations between specific antibody responses and resistance to reinfection in a Senegalese population recently exposed to *Schistosoma mansoni*. *Trop Med Int Health*, **12**, 431-44.
- Villamor, E., Fataki, M. R., Mbise, R. L. & Fawzi, W. W. (2003) Malaria parasitaemia in relation to HIV status and vitamin A supplementation among pre-school children. *Trop Med Int Health*, **8**, 1051-61.
- Vuong, T. A., Nguyen, T. T., Klank, L. T., Phung, D. C. & Dalsgaard, A. (2007) Faecal and protozoan parasite contamination of water spinach (*Ipomoea aquatica*) cultivated in urban wastewater in Phnom Penh, Cambodia. *Trop Med Int Health*, **12 Suppl 2**, 73-81.
- Wanji, S., Tendongfor, N., Esum, M. E. & Enyong, P. (2002) *Chrysops silacea* biting densities and transmission potential in an endemic area of human loiasis in south-west Cameroon. *Trop Med Int Health*, **7**, 371-7.
- Wilson, S., Vennervald, B. J., Kadzo, H., *et al.* (2007) Hepatosplenomegaly in Kenyan schoolchildren: exacerbation by concurrent chronic exposure to malaria and *Schistosoma mansoni* infection. *Trop Med Int Health*, **12**, 1442-9.
- Yanze, M. F., Duru, C., Jacob, M., Bastide, J. M. & Lankeuh, M. (2001) Rapid therapeutic response onset of a new pharmaceutical form of chloroquine phosphate 300 mg: effervescent tablets. *Trop Med Int Health*, **6**, 196-201.
- Ziem, J. B., Magnussen, P., Olsen, A., Horton, J., Asigri, V. L. & Polderman, A. M. (2006) Impact of repeated mass treatment on human *Oesophagostomum* and hookworm infections in northern Ghana. *Trop Med Int Health*, **11**, 1764-72.