Arab Gulf J. scient Res. **3** (1), pp.237-247 (1985)

The Distribution of Sample Egg-Count and Its Effect on the Sensitivity of Schistosomiasis Tests

A.G. A.G. Babiker*

International Centre for Theoretical Physics, Trieste, Italy

ABSTRACT. Let D be the total number of Schistosomiasis eggs produced by an individual in a day, counting only those eggs which leave the body, and let S denote the number of eggs observed on a slide under the microscope and p the probability that a given egg ends up in a given slide. Assuming that D follows a negative binomial distribution, the distribution of S and that of D given S are studied.

Explicit expressions for the conditional mean and variance of D given S and for the probability of a false negative slide are obtained. Two different cases are considered. In the first case, p is assumed to be constant, and in the second, more general case p is assumed to follow a beta distribution. In either case, the probability of a false negative slide, which is a measure of the insensitivity of the laboratory test, is shown to ultimately decrease with the degree of clumping of the daily egg excretion D.

I. Introduction

Schistosomiasis, one of the most important helminthic infections, is endemic in many parts of the world and is estimated to afflict nearly 5% of the world population with more than twice as many at risk (Jordan and Webbe 1969, Iarotski and Davis 1981). The disease is caused by flatworms (Schistosomes) the main types of which are *S. haematobium, S. mansoni* and *S. japonicum*.

Unlike microparasitic infections, where categorization of individuals into susceptible, infected, etc., is quite an adequate frame for describing the epidemiology of the disease, for helminthic infections, models must take into account the dis-

* *Permanent address:* School of Mathematical Sciences, University of Khartoum, Khartoum, Sudan.

tribution of the number of worms harboured by an individual. Thus, for Schistosomiasis, the problem of relating sample egg-count (an observable quantity) to worm load (an unobservable quantity) becomes quite important. For studies of this relation, we refer to Cheever (1981) and Cheever *et al.* (1977).

Published prevalence data is mainly based on egg counts in small samples of urine or faeces, depending on the type, and thus tend to underestimate the proportion of individuals infected. The method of testing for *S. mansoni* used in the Sudan and many other places uses 75 mg of faeces, and the average daily output of faeces is in the range of 150 g, so that the proportion tested is at best of the order of 10^{-3}. It is widely accepted that schistosomes are not randomly distributed among individuals, but tend to be distributed in an over-dispersed or clumped fashion, *i.e.* a small proportion of the individuals carry most of the worms in the community, and there is empirical evidence to support this (*see* Bradley and May 1978, Anderson and May 1982). It is, thus, likely that the total number of eggs excreted daily by an individual is also aggregated or clumped. A simple theoretical distribution which describes this clumping quantitatively *via* a single parameter k is the negative binomial distribution. The smaller the value of k, the more clumped or over-dispersed is the distribution. As k tends to infinity, the distribution tends to the Poisson distribution, which is the case when the eggs are randomly distributed.

In this note we, therefore, assume that the number of eggs excreted daily by an individual follows a negative binomial distribution. We derive the distribution of the number of eggs seen on a slide and obtain expressions for the probability of a false negative as a measure of the insensitivity of the test first under a simplifying homogeneity condition and then under a more realistic hetrogeneity assumption. It turns out that beyond a certain degree of clumping of the daily egg excretion, depending on the intensity of the disease in the community, the sensitivity of the test increases with clumping.

II. Notation

Let D denote the total daily egg excretion of an individual in the community. We assume that each egg has a probability p of appearing on a given slide, and we denote by S the total number of eggs on the slide. The quantity p will depend, among other things, on the proportion of the amount of faeces tested. Typically, $E(p)$ is of the order of 10^{-3}. We assume that D has a negative binomial (N.B.) distribution with mean μ and clumping parameter k. For the distribution of p, we consider two cases: in the first of which we make the simplifying assumption that p is a constant, and in the second case we assume that p has a beta distribution. The reason for this is that the family of the beta distribution is relatively simple mathematically, yet, being a two-parameter family, it includes a vast number of distributions over the interval (0, 1).

III. Case I

The probability p is assumed to be a constant throughout this section. The probability law of D is given by:

$$P\{D = d\} = (1-\gamma)^k (k)_d \frac{\gamma^d}{d!} \; ; \; d = 0, 1, 2, \ldots$$

where

$$(k)_d = \frac{\Gamma(k+d)}{\Gamma(k)} \quad ; \quad \gamma = \frac{\mu}{k+\mu} .$$

We denote such a distribution by N.B. (γ,k). For given D, S has the binomial distribution ***Bi(D, p)***, so that,

$$P\{S = x | D = d\} = \binom{d}{x} p^x q^{d-x} \; ; \; x = 0, 1, \ldots d \; ; \; q = 1 - p.$$

Hence, the joint probability law of S and D is given by

$$P\{S = x; D = d\} = \binom{d}{x} p^x q^{d-x} (1-\gamma)^k (k)_d \frac{\gamma^d}{d!} \; ; \; d \geqslant x = 0, 1, 2, \ldots$$

$$= \frac{p^x}{x!} \frac{q^{d-x}}{(d-x)!} (1-\gamma)^k (k)_d \gamma^d$$

$$= \left(\frac{1-\gamma}{1-\gamma q}\right)^k (k)_x \left(\frac{\gamma p}{1-\gamma q}\right)^x \frac{1}{x!} \left\{(1-\gamma q)^{k+x} (k+x)_{d-x} \frac{(\gamma q)^{d-x}}{(d-x)!}\right\}$$

so,

$$P\{S = x; D = d\} = \left(1 - \frac{\gamma p}{1-\gamma q}\right)^k \frac{1}{x!} (k)_x \left(\frac{\gamma p}{1-\gamma q}\right)^x \times$$

$$\left\{(1-\gamma q)^{k+x} (k+x)_{d-x} \frac{(\gamma q)^{d-x}}{(d-x)!}\right\} ;$$

$$d \geqslant x = 0, 1, 2, \ldots \tag{1}$$

It follows from (1) that

$$P\{S = x\} = \left(1 - \frac{\gamma p}{1-\gamma q}\right)^k \frac{1}{x!} (k)_x \left(\frac{\gamma p}{1-\gamma q}\right)^x ; x = 0, 1, \ldots \tag{2}$$

a negative binomial distribution with the same clumping parameter k as D. We also have

$$P\{D = d | S = x\} = (1-\gamma q)^{k+x} (k+x)_{d-x} \frac{\gamma q}{(d-x)!} \; ; \; d = x, x+1, \ldots \tag{3}$$

so that, given $S = x$, $D - x \sim$ N.B. $(\gamma q, k + x)$.

The conditional mean and variance of D are given by:

$$E(D|S=x) = x + \frac{(k+x)\gamma q}{1-\gamma q} = \left(1 + \frac{\gamma q}{1-\gamma q}\right)x + \frac{k\gamma q}{1-\gamma q}$$

$$= \left(1 + \frac{q\mu}{k+p\mu}\right)x + \frac{kq\mu}{k+p\mu} \tag{4}$$

$$V(D|S=x) = \frac{\gamma q(k+x)}{(1-\gamma q)^2} = \frac{q\mu(k+x)\,(k+\mu)}{(k+p\mu)^2} \tag{5}$$

Barring density dependence of the fertility of egg-layers, and ignoring other sources of variability in the number of eggs layed daily by each paired female schistosome, the conditional mean and variance of the number of paired schistosomes W are easily obtained from (4) and (5) observing that

$$E(W|S=x) = \frac{1}{\lambda} E(D|S=x) \quad , \quad \text{and}$$

$$V(W|S=x) = \frac{1}{\lambda^2} V(D|S=x) \quad ;$$

where λ is the number of eggs produced daily by each paired female schistosome.

3.1. Sensitivity of the Test for Infection

It follows from (3) that, among all individuals with egg-free slides, $D \sim N.B.(\gamma q, k)$, and so, on the average only a proportion $(1-\gamma q)^k$ of them would actually be worm free. A measure for the insensitivity of the test is provided by what is called the probability of a false negative. This is the probability that a slide contains no eggs when the individual is in fact infected. This is given by:

$$P\{S=0|D>0\} = \frac{P\{D>0|S=0\}P\{S=0\}}{P\{D>0\}}$$

$$= \frac{(1-(1-\gamma q)^k)[(1-\gamma)/(1-\gamma q)]^k}{1-(1-\gamma)^k}$$

i.e.

$$P\{S=0|D>0\} = \frac{(1-\gamma q)^{-k}-1}{(1-\gamma)^{-k}-1} = \frac{[1-q\mu/(k+\mu)]^{-k}-1}{[1-\mu/(k+\mu)]^{-k}-1} \tag{6}$$

To investigate the way in which the probability of a false negative changes with the degree of clumping we rewrite (6) as

$$P\{S=0|D>0\} = Q(\alpha;p,\mu) = \frac{f(\alpha;p,\mu)-1}{g(\alpha;\mu)-1} \tag{6a}$$

where

$$f = \frac{(1+\alpha)^{\mu/\alpha}}{(1+p\alpha)^{\mu/\alpha}} \quad , \quad g = (1+\alpha)^{\mu/\alpha} \text{ and } \alpha = \frac{\mu}{k}$$

represents the degree of clumping of the distribution of D.

For large α, we have

$$(1+\alpha)^{\mu/\alpha} = \alpha^{\mu/\alpha}(1+1/\alpha)^{\mu/\alpha}$$

$$= \alpha^{\mu/\alpha}(1+\mu/\alpha^2 + 0(1/\alpha^3)); \text{ and}$$

$$(1+p\alpha)^{-\mu/\alpha} = (p\alpha)^{-\mu/\alpha}(1-\mu/p\alpha^2 + 0(1/\alpha^3)$$

so that

$$f = p^{-\mu/\alpha}(1+\mu/\alpha^2 + 0(1/\alpha^3)(1-\mu/p\alpha^2 + 0(1/\alpha^3))$$

$$= (1-(\mu/\alpha)\log p + 0(1/\alpha^2))(1+\mu/\alpha^2 + 0(1/\alpha^3))(1-\mu/p\alpha^2 + 0(1/\alpha^3))$$

and

$$g = (1+(\mu/\alpha)\log\alpha + 0(\log\alpha/\alpha))(1+\mu/\alpha^2 + 0(1/\alpha^3))$$

$$\therefore Q = \frac{f-1}{g-1} = \frac{-\log p + 0(1/\alpha)}{\log\alpha + 0(1/\alpha)} \simeq -\frac{\log p}{\log\alpha}$$

Hence, beyond a certain value of α depending on μ and p, Q decreases with α.

For small α we consider the derivative of Q w.r.t. α. Routine computations show that

$$\frac{\partial Q}{\partial \alpha} = A(\alpha; p, \mu)\left[D(\alpha; p) - \frac{B(\alpha)}{E(\alpha; p)}\right] \tag{7}$$

where

$$A = \mu fg/\{\alpha(1+\alpha)^{\mu/\alpha}(g-1)^2[(1+p\alpha)^{\mu/\alpha} - 1]E\} \geq 0$$

$$D = [(1+\alpha)^{\mu/\alpha} - 1]/[(1+p\alpha)^{\mu/\alpha} - 1]$$

$$B = (1/\alpha)\log(1+\alpha) - 1/(1+\alpha) \quad \text{and}$$

$$E = (1/\alpha)\log(1+p\alpha) - p/(1+p\alpha).$$

Now,

$$\lim_{\alpha \to 0} D = (e^{\mu} - 1)/(e^{p\mu} - 1)$$

and it can be easily shown that

$$\lim_{\alpha \to 0} \frac{B}{E} = \frac{1}{p^2}$$

$$\therefore \lim_{\alpha \to 0} (D - B/E) = (e^{\mu} - 1)/(e^{p\mu} - 1) - 1/p^2$$

$$= \frac{1}{p^2(e^{p\mu} - 1)} (p^2 e^{\mu} - e^{p\mu} - p^2 + 1) > 0$$

$\therefore$ for sufficiently small α, $\dfrac{\partial Q}{\partial \alpha} > 0$ and Q increases with α.

We conclude that for given p and μ, there exist two values α_1 and α_2 such that Q increases with α for $0 < \alpha < \alpha_1$, and decreases with α for $\alpha > \alpha_2$.

3.2. Prediction of D

To predict the value of D for an individual with $S = x$, the best predictor $\hat{d}$ in terms of minimizing $E[(D - d)^2 | S = x]$ is obviously the conditional mean $[1/(p\mu + k)][(k + \mu)x + kq\mu]$, provided that k and μ are known. Otherwise, from large cross-sectional data on S, we can obtain consistent estimators $\hat{k}$ and $\hat{\mu}$ and use $\hat{D}(x) = \dfrac{1}{(p\hat{\mu} + \hat{k})} [\hat{k} + \hat{\mu})x + \hat{k}\hat{\mu}q]$ which is a consistent estimator of $E(D|S = x)$.

The method used now to convert $S = x$ to eggs per gram of faeces, is to multiply x by a constant factor. This has obviously a constant mean square error as an estimate of eggs/gram as compared with the mean square error of $\hat{D}(x)$ as an estimate of $E[D|S = x)$. The latter error tends to zero as the size of the sample used to estimate μ and k, tends to infinity.

3.3. A Limiting Situation

The case when D has a Poisson distribution with mean μ can be treated by letting $k \to \infty$ keeping μ fixed. In this case, S has a Poisson distribution with mean μp and the conditional distribution of $D - S$ given $S = x$ is Poisson with mean $\mu(1 - p)$ so that

$$E(D|S = x) = x + \mu q, \quad \text{and}$$

$$V(D|S = x) = \mu q.$$

The corresponding probability of a false negative is then

$$P\{S = 0|D > 0\} = (e^{-\mu p} - e^{-\mu})/(1 - e^{-\mu}).$$

IV. Case II

We now assume that p has a beta distribution with parameters r and u so that p has the probability density function

$$f(p) = \frac{1}{B(r, u)} p^{r-1} q^{u-1}, q = 1 - p, 0 < p < 1;$$

where $B(r, u)$ is the beta function $\Gamma(r)\Gamma(u)/\Gamma(r + u)$. The probability generating function $g_{S|D}$ of S given D is easily seen to be

$$g_{S|D}(t) = \frac{1}{B(r, u)} \int_0^1 p^{r-1}(1 - p)^{u-1}(1 + p(t - 1))^D dp$$

$$= F(-D, r; r + u; (1 - t)) \tag{8}$$

where F is the hypergeometric function;

$$F(a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} x^n;$$

c.f. Abramowitz and Stegun (1965).

The probability generating function of S is then

$$g(t) = \frac{1}{B(r, u)} \int_0^1 p^{r-1}(1 - p)^{u-1}(1 - \gamma)^k \sum_{n=0}^{\infty} \frac{(1 + p(t - 1))^n \gamma^n}{n!} (k)_n \, dp$$

$$= \frac{(1 - \gamma)^k}{B(r, u)} \int_0^1 \frac{p^{r-1}(1 - p)^{u-1}}{[1 - (1 + p(t - 1))\gamma]^k} dp$$

$$= \frac{1}{B(r, u)} \int_0^1 p^{r-1}(1 - p)^{u-1} \left[1 - \frac{\gamma}{1 - \gamma}(t - 1)p\right]^{-k} dp$$

Thus,

$$g_S(t) = F\left(k, r; r + u; \frac{\gamma}{1 - \gamma}(t - 1)\right) \tag{9}$$

The x*th* derivative of g_S at $t = 0$ is

$$g_S^{(x)}(0) = \left(\frac{\gamma}{1 - \gamma}\right)^x \frac{(k)_x (r)_x}{(r + u)_x} F\left(k + x, r + x; r + u + x; \frac{-\gamma}{1 - \gamma}\right)$$

$$= \frac{B(r + x, u)}{B(r, u)} (1 - \gamma)^k (k)_x \gamma^x F(k + x, u; r + u + x; \gamma)$$

and so,

$$\frac{1}{x!} g_S^{(x)}(0) = P\{S = x\}$$

$$= \frac{B(r + x, u)}{B(r, u)} (1 - \gamma)^k F(k + x, u; r + u + x; \gamma) (k)_x \frac{\gamma^x}{x!}; x = 0, 1, \ldots \tag{10}$$

We also have,

$$P\{S = x; D = d\} = \frac{(1-\gamma)^k \gamma^d (k)_d}{B(r,u)\, x!\, (d-x)!} \int_0^1 p^{r+x-1}(1-p)^{u+d-x-1} dp$$

$$= \frac{(1-\gamma)^k \gamma^d (k)_d}{B(r,u)\, x!\, (d-x)!} B(r+x, u+d-x);\ d \geqslant x = 0, 1, \ldots \tag{11}$$

and thus,

$$P\{D = d | S = x\} = \frac{B(r+x, u+d-x)}{B(r+x,u)\, F(k+x, u; r+u+x; \gamma)} \times$$

$$\times (k+x)_{d-x} \frac{\gamma^{d-x}}{(d-x)!};\ d = x, x+1, \ldots \tag{12}$$

The conditional mean of D is directly obtained from (12) as:

$$E(D|S=x) = x + \frac{u\gamma(k+x)\, F(k+x+1, u+1; r+u+x+1; \gamma)}{(r+u+x)\, F(k+x, u; r+u+x; \gamma)} \tag{13}$$

We observe that the conditional mean of D is no longer linear in x. This may very well partially explain the reported apparant density dependent regulation of egg production as exhibited by comparing data on S and W the number of paired female schistosomes harboured by an individual (Anderson and May 1982, Cheever 1968).

We conclude by obtaining an expression for the probability of a false negative. Using (10) and (12) we get

$$P\{S = 0 | D > 0\} = [F(k, u; r+u; \gamma) - 1]/[(1-\gamma)^{-k} - 1] \tag{14}$$

We show that this probability tends to zero as $k \to 0$, so again, the sensitivity of tests increases with clumping beyond a certain degree of overdispersion.

Using the integral representation of the hypergeometric function we write (13) in the form

$$P\{S = 0 | D > 0\} = \frac{1}{B(r,u)} \int_0^1 \{[(1-\gamma t)^{-k} - 1]/[(1-\gamma)^{-k} - 1]\}\, t^{u-1}(1-t)^{r-1} dt$$

$$= \frac{1}{B(r,u)} \int_0^1 \left\{ \frac{[1 - \mu t/(\mu+k)]^{-k} - 1}{[1 - \mu/(\mu+k)]^{-k} - 1} \right\} t^{u-1}(1-t)^{r-1} dt. \tag{13a}$$

Using L'Hospital's rule, one easily shows that the integrand in (13a) tends to zero as k tends to zero. Since the integrand is clearly dominated by $t^{u-1}(1-t)^{r-1}$, the dominated convergence theorem gives the desired result that $P\{S = 0 | D > 0\}$ tends to zero as k tends to zero.

On the other hand, as $k \to \infty$

$$F(k, u; r + u; \gamma) \to \frac{1}{B(r, u)} \int_0^1 e^{t\mu} t^{\mu - 1}(1 - t)^{r - 1} dt \ ;$$

which is Kummer's function $M(u, r + u; \mu)$, (Abramowitz and Stegun 1965, Ch. 13); and is also the moment generating function of $1 - p$. In this limiting case, D has a Poisson distribution with mean μ and

$$P\{S = 0 | D > 0\} = \frac{M(u, r + u; \mu) - 1}{e^{\mu} - 1} \ .$$

V. Discussion and Conclusions

Under the assumption that D, the number of eggs produced daily by an individual (counting only those eggs which leave the body) follows a negative binomial distribution, and that each egg has probability p of actually appearing in a given slide, it follows that the total number of eggs S on a slide has a negative binomial distribution with the same clumping factor as D. Given that x eggs were observed on a slide, the conditional mean and variance of D are linear functions of x (*c.f.* equations (2), (4) and (5)).

The keto method of stool examination for *S. mansoni* uses 75 mg of faeces divided into three slides, and the total number of eggs observed is then multiplied by $^{40}/_{3}$ to obtain an estimate of the number of eggs in one gram of faeces. This implicity assumes that the eggs are uniformly distributed within the faeces, and, thus, effectively estimates D by multiplying x by a constant factor. The results above show that a better estimate of D (in the sense of minmum mean square error) is

$$\hat{D}(x) = \frac{1}{p\hat{\mu} + \hat{k}} [(\hat{k} + \hat{\mu})x + \hat{k}\hat{\mu}(1 - p)]$$

where $\hat{\mu}$ and $\hat{k}$ are any consistent estimates of μ and k the mean and clumping factor of D. The estimates $\hat{\mu}$ and $\hat{k}$ may be obtained from large cross-sectional data on S.

An expression for the probability of a false negative slide is given in equation (6). This probability approximates the proportion of infected individuals who would be falsely classified as uninfected. Published prevalence data which is usually based on egg counts is, thus, an underestimate of the prevalence. The probability of a false negative slide depends on μ, k and p. It clearly decreases with p and our results show that beyond a certain degree of clumping as measured by $\alpha = \mu/k$, this probability decreases with α.

Looking at the more general case, when d is assumed to be a random variable with a beta distribution, it turns out that the conditional mean of D, given x observed eggs on a slide, is no longer linear in x. This could well give an alternative

explanation, other than density dependent regulations, of reported data comparing S and W, the number of eggs observed and the number of paired female schistosomes, respectively.

As in the earlier case, when p is constant, the probability of a false negative in this more general situation ultimately decreases with overdispersion.

References

Abramowitz, M. and **Stegun, I.A.** (1965) *Handbook of Mathematical Functions*, Dover Publications, N.Y.

Anderson, R.M. and **May, R.M.** (1982) Population dynamics of human helminth infections: Control by Chemotherapy, *Nature, Lond.* **297**: 557-563.

Bradley, D.J. and **May, R.M.** (1978) Consequences of helminth aggregation for the dynamics of schistosomiasis, *Trans. R. Soc. trop. Med. Hyg.* **72**: 262-273.

Cheever, A.W. (1968) A quantitative post-mortem study of schistosomiasis mansoni in man, *Am. J. trop. Med. Hyg.* **17**: 38-64.

Cheever, A.W., Kamel, I.A., Elwi, A.M., Mosimann, J.E. and **Danner, R.** (1977) *Schistosoma mansoni* and *S. haematobium* infections in Egypt. II. Qualitative parasitological findings at necropsy, *Am. J. trop. Med. Hyg.* **26**: 702-716.

Iarotski, L.S. and **Davis, A.** (1982) The schistosomiasis problem in the world: Results of a WHO questionnaire survey, *Bull. WHO* **58**: 115-127.

Jordan, P. and **Webbe, G.** (1969) *Human Schistosomiasis*, Heinemann Medical Books, London.

(Received 15/10/1983;
in revised form 13/10/1984)

توزيع عدد بويضات البلهارسيا فى العينة المعملية وأثر ذلك على دقة الاختبار المعملى

عبد القيوم عبد الغنى بابكر

مدرسة العلوم الرياضية - جامعة الخرطوم - ص.ب ٣٢١ - الخرطوم السودان

ليكن D عدد بويضات البلهارسيا التى يفرزها الشخص يوميا وليكن S عدد البويضات التى تظهر فى عينة المختبر و p احتمال ظهور أية بويضة فى عينة مختبر معينة. هذا البحث يُعنى بدراسة توزيع D والتوزيع الشرطى للمتغير D بمعلومية S.

بافتراض أن D يتبع توزيع ذات الحدين السالب تستخلص هذه الدراسة تعابيراً جبرية للوسط الشرطى والتباين الشرطى للمتغير D بمعلومية S وكذلك احتمال عدم ظهور أية بويضة بالعينة إذا كان الشخص مصاباً بالمرض وذلك فى حالتين. فى الحالة الأولى نفترض أن p مقدار ثابت وفى الحالة الثانية، وهى الأعم، نفترض أن p متغير يتبع توزيع بيتا.

وفى كلتا الحالتين، نبين أن احتمال عدم ظهور أية بويضة فى عينة المصاب يتناقص فى النهاية مع زيادة درجة تشتت المتغيرِّ D.