FOLIA PARASITOLOGICA (PRAHA) 20: 227—234, 1973.

ON THE NATURE OF RELATION BETWEEN THE INCIDENCE OF INFECTION IN HOST POPULATION AND THE MEAN DENSITY OF PARASITE

K. A. BREYEV

Zoological Institute, Academy of Sciences of the USSR, Leningrad

Abstract. The density of parasite in the host population is presented as a probabilist problem and a mathematical approach to this phenomenon is given.

The studies on population ecology of parasites are very important because they help understand factors and relations regulating their numbers. Such studies, however, are often hindered by difficulties in getting sufficient material on numbers of any species under varied conditions. The difficulties are especially great as regards endoparasites since killing and dissection of the host are needed to obtain their complete count. However the method to obtain the data on the proportion of specimens in the host population infected with a certain species of parasite (incidence of infection) is much more simple and often does not entail any dissection. Due to this fact the above problem becomes very interesting. In addition, its investigation may lead to other interesting conclusions.

Now we can formulate the essence of the problem. Let the proportion of infected specimens in the host population be E; the mean density of parasite —M; the total number of specimens in the host population or in the representative sample from it —N; number of parasites in one host —$x_0, x_1, x_2, x_3, \ldots, x_n$; frequencies of each value of x — $f_0, f_1, f_2, f_3, \ldots, f_n$.

The density (or abundance) of any species usually means the number of specimens present per unit of inhabited area. As regards parasites, one host is such a unit of area. Hence

$$M = \frac{1}{N} \sum_{i=0}^{n} x_i f_i \tag{1}$$

$$E = \frac{N - f_0}{N} = 1 - \frac{f_0}{N} \tag{2}$$

f_0 — frequency of x_0 (non-infected host specimen).

From the mathematical point of view the number of any species can be considered as a discrete random variable, because due to the great variety of factors and complexity of relations determining it we cannot forecast its precise value at any particular moment of time per any unit of inhabited area.

Let this random variable be ξ_x and then, according to the accepted terminology in the theory of probability, its value will be:

$$\xi_x = \begin{pmatrix} x_0, x_1, x_2, x_3, \dots x_n \\ P_0, P_1, P_2, P_3, \dots P_n \end{pmatrix}$$

x — designates the possible values of random variable (in our case the number of parasites per one host) and P—the probability of each value of x. The set of ordered pairs, x and P, is the probability function or the distribution of random variable.

In this case the mean density of parasite is determined by mathematical expectation of discrete random variable

$$M\xi_x = \sum_{i=0}^{n} x_i P_i$$

It is not hard to notice that the equation (I) is the particular case of this function, since we can determine approximately each value of p_i by ratio $\frac{f_i}{N}$. Hence, $M = \sum x_i \frac{f_i}{N} = \frac{1}{N} \sum x_i f_i$. And as $\frac{f_0}{N} = P_0$, from equation (2): $E = 1 - P_0$.

Consequently the solution of the problem consists mainly in determining the dependence between the value of the zero member of the distribution of random variable and its mathematical expectation.

Thus our problem can be solved as a probabilist one only, proceeding from the idea of the species number as random variable, and on the basis of a statistical model of its distribution. Let us illustrate the solution using the examples of the most usual types of distribution applied in ecology: that of Poisson, binomial and negative binomial.

The dependence is most easily determined in the case of Poisson distribution, that has only one parameter. The formula of zero member of this distribution is:

$$P_0 = \frac{M^\circ}{O!} e^{-M} = \frac{1}{e^M}$$

As $E = 1 - P_0$, we have $1 - E = \frac{1}{e^M}$, $e^M = \frac{1}{1 - E}$

$$M = \ln\left(\frac{1}{1 - E}\right) \qquad (3)$$

which is the very formula of dependence between the incidence of infection in terms of unit parts —E, and the mean density of the parasite —M.

The following dependences are characteristic of the binomial distribution $(p + q)^n$: $p + q = 1$, $M = np$. Zero member of distribution $P_0 = q^{\frac{M}{n}}$, or $1 - E = q^{\frac{M}{n}}$.

By logarithmic transformation of this equation we obtain:

$$M = p\,\frac{\lg(1 - E)}{\lg q}; \qquad M = p\,\frac{\lg(1 - E)}{\lg(1 - p)} \qquad (4)$$

As binomial distribution has two parameters, our formula should also have, besides the incidence of infection —E, the second parameter —p i.e. the probability of infection of one host with one parasite. Evidently this probability does not remain constant under conditions causing different incidence of infection and a new question arises regarding the characteristic of its alteration. We shall consider it below.

The negative binomial distribution $p^k(1-q)^{-k}$ has two parameters: the mathematical expectation —M and the distribution exponent —k, characterizing the extent of aggregation of randome variable; p and q are auxiliary parameters: $p = \frac{k}{k+M}$; $q = 1 - p$.

Zero member of distribution $P_0 = p^k$, but $P_0 = 1 - E$, then $1 - E = p^k$; $1 - E = \left(\frac{k}{k+M}\right)^k$ and with inverse values:

$$\frac{1}{1-E} = \left(\frac{k+M}{k}\right)^k; \qquad \frac{k+M}{k} = \sqrt[k]{\frac{1}{1-E}}; \qquad 1 + \frac{M}{k} = \sqrt[k]{\frac{1}{1-E}};$$

$$M = k\left(\sqrt[k]{\frac{1}{1-E}} - 1\right) \tag{5}$$

In order to determine the mean density of the parasite by means of the value of incidence of infection one should know also the value of the second distribution parameter—k and there is the question about its constancy or variability under conditions of different incidence of infection, i.e. under different ratios of densities of host and parasite populations. In order to solve this problem a study of parameter k should be done in representative samples from host populations in the whole spectrum of existing densities of parasite population, from the smallest to the largest. However, before doing this let us return to a more simple Poisson distribution and check up the dependence obtained (3) by the examples of distribution of pleurocercoids of *Digramma interrupta* (Rud.) in the population of the crucian carp *Carassius auratus* (Bloch), based on the data kindly made available by Dr. M. N. Dubinina.

The infection of fishes with *D. interrupta* takes place when they feed on *Cyclops* sp. containing invasive procercoids, out of which pleurocercoid larvae develop in the fish organism later on. It is well known (Dubinina 1966) that even in case of artificial infection of *Cyclops* and other crustaceans—the first intermediate hosts of *Digramma*, with a large number of oncospheres, no more than 1—3 parasites per one crustacean can survive till the stage of invasive procercoid, the mean density of procercoids in crustaceans being very low. It is practically impossible to find infected crustaceans even in a sample of more than a thousand specimens taken in heavily infested localities. Apparently the repeated infection of fishes with several procercoids does not affect their development. There were cases when 6—7 pleurorcoids were found in one fish.

Now let us imagine a population of some fish species in which pleurocercoids are able to develop. If such a population inhabits a water reservoir rich in crustaceans infected with procercoids, the distribution of pleurocercoids in the same age group of young fishes in this population must agree with the Poisson law. In fact, the distribution of such a type arises under the following conditions: 1. small probability of random event studied; 2. its identity within the limits of a definite set of objects; 3. complete independence of each event from one another. In our example, if the occurrence of one pleurocercoid in one fish is regarded as an event: 1. there will be a small probability of the fish eating an infected crustacean as an average number of crustaceans with pleurocercoids is small. 2. the probability will be identical for all fishes of the same age; 3. in case of repeated consumption of infested crustaceans the new infection will not affect the development of pleurocercoids.

In fact the counts of pleurocercoids of *D. interrupta* in samples of alevins (*O*+) and yearlings (1+) of crucian carp agree quite well with the Poisson distribution: $P > 0.50$ and $P > 0.80$, respectively (Table 1).

Table 1. The distribution of pleurocercoids of *Digramma interrupta* (Rud.) in the crucian carp *Carassius auratus* (Bloch.)

Number of pleurocercoids per fish X	Number of fishes observed f_e	Expected frequencies by Poisson distribution f_p
	Alevins (0+)	
0	38	37.58
1	9	9.97
2	2	1.45
	49	49 $\chi^2 = 0.3077$ $\nu = 1$ $P > 0.50$
	Yearlings (1+)	
0	33	32.92
1	7	7.23
2	1	0.85
	41	41 $\chi^2 = 0.0340$ $\nu = 1$ $P > 0.80$

Let us check now the correctness of the formula (3) by the considered examples. Alevins (0+) $N = 49; \sum x_i f_i = 13; f_0 = 38$

$$M = \frac{13}{49} = \underline{\underline{0.2653}}; \qquad E = \frac{49 - 38}{49} = \underline{\underline{0.2245}}$$

$$M = \ln\left(\frac{1}{1 - E}\right) = \ln\left(\frac{1}{1 - 0.2245}\right) = \ln 1.2895 = \underline{0.2542}$$

Yearlings (+1) $N = 41; \sum x_i f_i = 9; f_0 = 33$

$$M = \frac{9}{41} = \underline{\underline{0.2195}}; \qquad E = \frac{41 - 33}{41} = \underline{\underline{0.1951}}$$

$$M = \ln\left(\frac{1}{1 - E}\right) = \ln\left(\frac{1}{1 - 0.1951}\right) = \ln 1.2424 = \underline{0.2170}$$

In both cases the expected values of M agree quite well with the observed ones, the errors being 4 per cent and 1 per cent respectively

However, if we use the same formula (3) for infection of 3-year-old crucian carps with pleurocercoids, the result will be different:

3-year-olds (3+) $N = 102; \sum x_i f_i = 23; f_0 = 90.$

$$M = \frac{23}{102} = \underline{0.2255}; \qquad E = \frac{102 - 90}{102} = \underline{0.1176};$$

$$M = \ln\left(\frac{1}{1-E}\right) = \ln\left(\frac{1}{1-0.1176}\right) = \ln 1.1333 = \underline{0.1251}$$

The error is 44 per cent, i.e. there is no agreement here. As we can see from Table 2, the distribution of pleurocercoids of *D. interrupta* in 3-year-old fishes does not fit the Poisson distribution. The lack of fit can be easily explained: the fishes of different age groups eat different numbers of crustaceans per unit of time and therefore the infection with pleurocercoids in adult fishes will increase with each season, the probability of infection in each of them being different. Consequently there will be no identity of probability of the event studied within the population of adult fishes. And that is the reason for considerable deviation from the Poisson distribution.

However, the variability of probability of the studied event is characteristic of the negative binomial distribution, which is the most convenient model in such cases. Having obtained the necessary parameters for this type of distribution we can compute the expected frequencies of 3-year-old crucian carps infected with any number of pleurocercoids (Table 2).* We can now see that these frequencies fit well with the observed data: $P > 0.50$.

Table 2. The distribution of pleurocercoides of *Digramma interrupta* (Rud.) in 3-year-old crucian carp *Carassius auratus* (Bloch)

Number of pleurocercoids per fish X	Number of fishes observed f_e	Expected frequencies by the Poisson distribution f_p	Expected frequencies by the negative binomial distributions f_{NB}
0	90	81.41	89.910
1	6	18.36	6.973
2	3	2.07	2.558
3	2	0.15	1.184
4	0	0.01	0.606
5	1	0.00	0.327 + 0.44
	102	102.00 $\chi^2 = 15.5707$ $P < 0.0005$ $\nu = 1$	102.00 $\chi^2 = 0.2882$ $P > 0.50$ $\nu = 1$

The appropriate formula (5) should be used in order to determine the dependence between the average number of parasite per fish and the incidence of infection. Let us do this calculation where the second parameter of the negative binomial distribution $k = 0.1182$

*) We omit here these calculations as they have no relation to the object of our consideration. The calculations of parameter k were made by the maximum likelihood method (Bliss and Fisher 1953).

3-year-old crucian carps $N = 102$; $\sum x_i f_i = 23$; $f_0 = 90$

$$M = \frac{23}{102} = 0.2255; \qquad E = \frac{102 - 90}{102} = 0.1176.$$

$$M = k\left(\sqrt[k]{\frac{1}{1-E}} - 1\right) = 0.1182\left(\sqrt[0.1182]{\frac{1}{1-0.1176}} - 1\right) =$$

$$= 0.1182\left(1.1333^{\frac{1}{0.1182}} - 1\right) = 0.1182(2.8825 - 1) = 0.2225.$$

There is a good agreement between expected and observed values of N (the error is 1 per cent), which confirms the correctness of relations we have determined.

However, we can use them and determine the mean density of parasite—M by means of the value of incidence of infection—E, only in those cases where the parasite distribution in host population agrees with the Poisson distribution. Such cases occur rarely. Much more frequent are the cases, where the probability of one parasite's occurrence in one host varies within the population. If this variability of probability is random, we have to deal with the negative binomial distribution, where the parameter k is also a component of the dependence formula.

Then it is necessary to find out whether this parameter remains constant, and if it does not, to what extent it can vary in different densities of parasite and host populations.

Naturally, before carrying out such investigations, we should determine the limits within which the type of distribution remains to be constant in different densities of parasite population and in various sex and age groups of the host and the cases when it varies, as we could observe on the example of crucian carp infection with pleurocercoids of *D. interrupta*.

Some variations of the parameters studied are inevitable because their complete coincidence even in two samples of equal volume and taken simultaneously from the same host population would be an exception rather than a rule. In such a case however, the variations are induced by random causes and the differences will remain within the limits of random error.

The situation is very different, if we deal with distribution parameters in samples with different density of the parasite. Basically, there are two possibilities.

1. In spite of different densities of the parasite (intensity of infection), the differences in parameter values remain within the limits of random error, i.e. the parameter is practically constant and then its common value should be determined. The parameters considered are: p—in case of the binomial distribution, i.e. probability of infection of one host with one parasite and survival of the latter to the moment of our counting; k—in case of the negative binomial distribution, i.e. a certain measure of aggregation capability or otherwise a certain measure of possible accumulation of parasites in one host. Such cases are possible only if nature and action of each factor determining the studied parameter remains to be constant irrespective of the density of the parasite population. If we presume here the complete absence of any reaction of the organism and the population of the host to the increasing intensity of infection with the parasite, such cases will be exceptions rather than a rule.

2. During an increase or decrease of the mean density of parasite a variation of its distribution parameters takes place. This variation may be associated with changes of host-parasite relations or of other factors affecting the parasite density. Such cases would be more common, because the probability of limitless parasite reproduction

in the host-parasite system is small and because there is usually some maximum of intensity of infection. If such a maximum exists (and it means some limit of aggregation capability), then, as the parasite density approaches this maximum the probability of survival of individual parasites decreases. Hence, a regular variation of distributional parameters will take place, depending on the incidence and intensity of infection. The function formula showing this regularity will be at the same time a statistical expression of the variation in action of a set of factors affecting the parasite number within the host-parasite system at different densities of parasite. As the maximum limit in the intensity of infection will play an important part in determination of the function of parameter variation and the maximum value is inevitably specific for each pair of host and parasite, the same specificity can be expected for this function.

We can illustrate the above by the following example. The statistical analysis of the distribution of cattle warble fly (*Hypoderma bovis* De Geer) in herds of cattle has shown that the negative binomial distribution fits the 2nd and 3rd instar larvae distribution at all levels of infestation intensity with enough reliability (Breyev 1968a and b). From this work it is evident, that the values of parameter k gradually increase with increasing incidence of infestation. A further analysis has allowed this regularity to be formulated as an equation of linear negative regression:

$$Y = 4.517 - 0.0436X$$

Where Y—value of $\frac{1}{k}$ for incidence of infestation X in per cent.

One can easily compute the value of k for any incidence of infestation. The mean density of parasite can then be calculated by means of formula (5).

However, it is far from all the useful information which can be obtained from the latter equation. Having determined the confidence limits for k in regression for any value of incidence of infestation, we can compute not only the density, but also the limits of its possible deviations towards maximum and minimum values. Thus we obtain the complete characteristics of the studied random variable, i.e. of number of the parasite population in the host population. Using these data we can estimate possible values of the random variable, i.e. the number of parasite larvae per one host, and their frequencies in each separate case of its mean value (density of parasite population) or in case of any incidence of infestation in samples of different volumes and also the limits of variability of the values obtained. For example, we can easily calculate the possible maximum number of larvae in one animal, if we know the incidence of infestation in the population or in the sample of any volume. If during practical work in some sample (e.g. herd of cattle) we come across animals which have the number of parasites greatly exceeding such maxima, we can assume that the sample is not homogeneous and attempt to find out the cause of its heterogenity.

In other words we get the possibility to build up a mathematical model of parasite distribution in the host population under different incidence and intensity of infestation. The example of such a model for *Hypoderma bovis* De Geer has been given in the mentioned paper (Breyev 1968b).

But this is not the whole mass of the information. An equation showing the function of variation of parameter k under different levels of infestation contains also the statistic characteristics of the degree to which the intensity of action of the whole set of factors regulating the parasite number and determining the limit of mean host infestation varies. We may make sure of this by extrapolation of k values, which are typical for a case of low incidence of infestation to higher one. Then we can compute by formula (5) the mean density of parasite and compare the computed data with observed values. Estimations of this kind made for *H. bovis*, enable us to know that the action of factors limiting the number of warble fly larvae becomes twice as large if the latter increases from 0.23 to 1.75 larvae per one host and becomes 31 times as large during further increase up to 4.12 larvae per host (Breyev 1968b).

Thus the study of the nature of relations between the incidence of infection in host population and the density of parasite has made us understand this problem as a probabilist one, understand the species number as a random variable and has further led to some interesting conclusions on possibilities of mathematic modelling of the distribution of the parasite number in host populations and statistical estimation of factors limiting the parasite number. The abundance of new information obtained by such methods of research makes clear their usefullness in studying the population ecology of parasites.

О ХАРАКТЕРЕ СВЯЗИ МЕЖДУ ДОЛЕЙ ЗАРАЖЕННЫХ ОСОБЕЙ В ПОПУЛЯЦИИ ХОЗЯИНА И УРОВНЕМ ЧИСЛЕННОСТИ ПАРАЗИТА

К. А. Бреев

Резюме. Уровень численности паразита в популяции хозяина представлен как задача решаемая только лишь на вероятностной основе, с математическим подходом к этому явлению.

REFERENCES

BLISS C. J., FISHER R. A., Fitting the negative binomial distributions to biological data and note on the efficient fitting of the negative binomial. Biometrics 9: 176—200, 1953.

BREYEV K. A., On the distribution of warble fly larvae in herds of cattle I. Negative binomial distribution as a model of distribution of warble fly larvae. Parazitologiya 2: 322 to 333, 1968a. (In Russian.)

—, On the distribution of warble fly larvae in herds of cattle. II. Exponent k in negative binomial distribution as a measure of dispersal of animals infestation with warble flies. Parazitologiya 2: 381—394, 1968b. (In Russian.)

DUBININA M. N., Cestoda: Ligulidae of the USSR fauna. Izd. "Nauka" L.: 3—259, 1966. (In Russian.)

Received 16 February 1972.

K. A. B., Zoological Institute, Academy of Sciences of the USSR, Lenigrad, USSR