Bulletin of the World Health Organization, 61 (1): 129-134 (1983)

A note on estimating selection pressures on insecticide-resistance genes*

R. J. WOOD[1] & L. M. COOK[2]

It is useful to be able to measure selection pressures acting on resistance genes in insect vectors of disease, since it is thus possible to predict future changes in frequency and to consider ways to minimize development of resistance. This note describes a method for estimating the selection coefficients, given two or more post-selection phenotype frequencies and knowing the number of generations between them.

The method is applied to published data on Anopheles labranchiae *under selection with DDT. The relative fitness (1-s) of the susceptibles compared with resistants was estimated by this method to be 31–38%. This was an annual estimate, but if the number of generations per year is known, it is also possible to calculate a value per generation. A computer program for making these estimates is given. The calculations depend on the gene being effectively recessive, i.e., on the heterozygote being killed by the dose applied in the field.*

Another approach to estimation of selection is by determining the deviation in gene frequency from the Hardy-Weinberg expectations. By this method, the relative fitness (1-s) of the susceptibles in a population of A. funestris *under dieldrin selection in the north of the United Republic of Cameroon has been estimated to be 40%. There are difficulties with this method, however, because population mixing may result in deviations that mimic the effect of selection. Examples are discussed for* A. gambiae, *where population mixing may occur and heterozygote deficiencies for the dieldrin resistance gene have been observed.*

For both methods of estimation, it is essential to know the real effective dominance of the resistance gene in the wild, i.e., whether the resistance heterozygote is killed or not. This factor is important in the control of resistance.

In a WHO technical document,[a] Muir called attention to the need to measure selection pressures acting on insecticide-resistance genes in malaria vectors. As he pointed out, the estimated selection coefficient would allow future changes in the frequency of the resistance genes to be predicted, and might aid in planning control programmes; even rough estimates of the coefficient could be valuable. The purpose of this note is to point out possible sources of inaccuracy in the calculation, and the ways in which they can be minimized.

Muir suggested that selection coefficients could be estimated either (*a*) by monitoring changes in the proportions of resistant and susceptible phenotypes on an annual or more frequent basis, or (*b*) by the deviation from Hardy-Weinberg equilibrium demonstrated by any one generation after undergoing selection. We shall consider the two possibilities in turn.

PERIODIC MEASUREMENT OF THE FREQUENCY OF RESISTANT INDIVIDUALS

Muir discussed the situation where a dominant susceptible gene at frequency p is selected against; the resistance gene is thus at frequency q, where $q = 1-p$. The gene frequency may be estimated roughly from the frequency of resistant individuals in tested samples assuming the Hardy-Weinberg proportions, and the selection coefficient estimated from the change in frequency between generations or at appropriate time intervals. Although the assumption of Hardy-Weinberg proportions is not justified, Muir suggested that it provides a useful approximation. While this will usually be true when differences in fitness between phenotypes are small, it is likely to lead to large errors as this differential increases.

One example discussed by Muir was the result of tests with DDT on *Anopheles labranchiae* in Algeria,

* From the Department of Zoology, University of Manchester, Manchester M13 9PL, England.

[1] Reader in Genetics.

[2] Reader in Zoology.

[a] MUIR, D. A. *Genetic aspects of developing insecticide resistance of malaria vectors. Part 1. Selection pressure.* WHO unpublished document, WHO/MAL/75.586, 1975.

where the survivors (homozygous for the DDT-resistance gene) had frequencies of 0.32, 0.48, and 0.70 in three successive years (1972, 1973, and 1974). He estimated the selection coefficient (s) of the susceptible genotype, using the Hardy-Weinberg assumption that the frequencies of survivors are estimates of q^2, which gave values of s of 0.56 for 1972–73 and 0.65 for 1973–74.

The validity of any estimate of s depends on sample size. The complete analysis should therefore include both an estimate of the selection coefficients and of the associated standard error. However, starting with only the change in phenotype frequencies, and without considering sample sizes, an improved method of calculation can be suggested. Other methods and problems involved have recently been reviewed by White & White (*10*).

In reality, the observed frequencies of the recessive phenotype are estimates, not of q^2, but of the fraction of survivors, after selection, from a population starting in Hardy-Weinberg equilibrium. Therefore:

$$\frac{q_0^2}{1-s(1-q_0^2)} = b_0 = 0.32$$

$$\frac{q_1^2}{1-s(1-q_1^2)} = b_1 = 0.48$$

$$\frac{q_2^2}{1-s(1-q_2^2)} = b_2 = 0.70$$

where q is the resistance-gene frequency, b is the phenotype frequency of homozygous resistant individuals, and the subscripts 0, 1, and 2 refer to the years 1972, 1973, and 1974.

It will be seen that the equations cannot provide estimates of q because s is unknown. However, the equations may be rearranged to give:

$$q_0^2 = \frac{(1-s)b_0}{1-sb_0}$$

and

$$q_1^2 = \frac{(1-s)b_1}{1-sb_1}$$

If a value of s is chosen arbitrarily, a value for q_0 may be calculated. The expected frequency in the next generation (q_1') is then:

$$q_1' = \frac{q_0^2 + (1-s)p_0q_0}{1-s(1-q_0^2)}$$

This value may be calculated and compared with the value q_1 obtained from the expression $(1-s)b_1/(1-sb_1)$, using the same value of s. The best value of s will be the one that makes q_1 the same as q_1'. A computer program in BASIC which performs this operation is given in Annex 1. The calculated value for the first period is 0.62 and for the second period 0.69, compared with Muir's values of 0.56 and 0.65. The estimated relative fitness $(1-s)$ of the susceptibles is therefore 38% and 31%, compared with Muir's 44% and 35%, a proportionately large reduction which could affect predictions of future changes in gene frequency, although in this particular case the effect on the estimated frequency after one generation is small.

The program calculates the value of s for fixed time intervals. If there are several records, such as the three annual figures in the example, an average is found to give the best fit for all the results. This is the selection coefficient for the time interval involved, in this case one year. In fact, *A. labranchiae* completes about four generations (range 3–7) in a year. The selection coefficient per generation may be calculated by reading in the number of generations between the records, expressed to the nearest whole number. In Annex 1 the average estimate of s is given both on an annual basis and per generation on the assumption of four generations per year. The annual estimate using the data for the three years is 0.66 and the estimate per generation is 0.22. Fig. 1 shows the projected change in frequency of the resistant homozygote over 20 generations.

These calculations depend on the resistance gene being effectively recessive, i.e., on the heterozygote being killed by the dose of insecticide applied in the field. If this is not the case (as was reported by Davidson for dieldrin resistance in *Anopheles gam-*

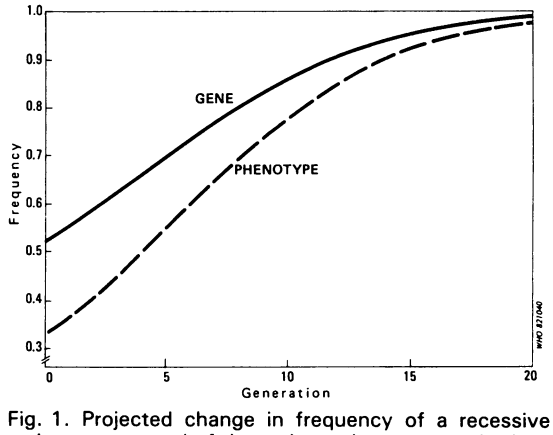

Fig. 1. Projected change in frequency of a recessive resistant gene and of the resistant homozygote in *Anopheles labranchiae* over 20 generations, given a selection coefficient of 0.22, as estimated in Annex 1.

biae (*3*)), then the calculated value of s will be different. Moreover, if the level of effective dominance changes with time as the insecticide deposits age, then meaningful predictions cannot be made unless the rate of change can be estimated. The consequences of differences in effective dominance have been discussed by Wood & Mani (*11*).

DEVIATION FROM HARDY-WEINBERG EQUILIBRIUM

Where all three genotypes can be distinguished, Muir supplies a quick graphical method of estimating the selection coefficient, based on the fact that selection for an effectively recessive gene causes a reduction in the frequency of heterozygotes in relation to the Hardy-Weinberg expectation. However, it is worth pointing out that the same effect may be produced by other causes, perhaps the most likely of which is the mixing of populations that have different resistance-gene frequencies.

Population mixing could be the result of immigration from unsprayed to sprayed areas. A further difficulty arises when mosquitos having domestic and wild subpopulations are sampled near houses, for the two groups will have been subject to different levels of insecticide. In Africa, *Aedes aegypti* has domestic and sylvatic populations that may coexist and hybridize to some extent in the peridomestic situation and the sylvatic form will occasionally enter houses (*7–9*).

A similar problem arises when samples are derived from a mixture of overlapping species that are impossible to distinguish morphologically. Such a situation could well occur in the *Anopheles gambiae* complex (*4*, *5*) where species A and B have been collected together in the same sample. In a particular mixed sample mentioned by Davidson et al. (*5*), species B carried dieldrin resistance while species A did not. However, it was subsequently shown that the dieldrin resistance gene was present in both species in the area of northern Nigeria where the sample had been collected.

Even should a population consist of only one species of the *A. gambiae* complex, this gives no guarantee of an unmixed sample from a resistance point of view. Coluzzi et al. (*2*) have presented evidence for the existence of subpopulations of both species A and B of the *A. gambiae* complex that exhibit different degrees of exophily (outdoor resting).[b]

For the case of k populations mixing in equal proportions with gene frequencies q_i (mean $\bar{q}$), Wahlund's formula (*6*) shows that the frequency of heterozygotes will be $2\bar{p}\,\bar{q} - 2\sigma_q^2$, where:

$$\sigma_q^2 = \frac{{}_1\Sigma^k q_i^2}{k} - \bar{q}^{\,2}$$

Referring to a dieldrin-selected population of *Anopheles funestus* in the north of the United Republic of Cameroon, Muir gives the estimate of s by the graphical method to be 0.6. On the Hardy-Weinberg expectation, q is $0.17 + 0.37/2 = 0.355$, so that the expected frequency of heterozygotes is $2 \times 0.355 \times 0.645 = 0.458$. This is a good deal higher than the observed value, a fact reflected by the large estimated value of s. The shortfall in the observed frequency from that expected is 0.088, which would be equivalent to $2\sigma_q^2$ if it resulted from mixing of populations. If the sample were, in fact, derived from two populations with different gene frequency and their deviation from the mean value $\bar{q}$ is d, then the above equation becomes:

$$\sigma_q^2 = \tfrac{1}{2}[(\bar{q} - d)^2 + (\bar{q} + d)^2] - \bar{q}^{\,2}$$
$$= d^2$$

from which d can be estimated to be 0.209. The observed result would therefore also be obtained in the absence of selection if the sample had been derived from two equally sized populations having gene frequencies of 0.146 and 0.564, respectively. Muir's method can therefore be used to estimate selection only if the possibility of mixing can be ruled out.

As already mentioned, a likely candidate for mixed samples is *A. gambiae*. This species has been investigated extensively in the field for dieldrin resistance, and in a number of cases the frequencies of all three genotypes have been measured. Muir gives an example from El Karo El Ahamda in the Sudan in which there was a deficiency of heterozygotes (23% compared with an expected value of 40.8%). Brown & Pal (*1*) provide data on 12 other populations, in 9 of which there was a deficiency of heterozygotes. These data are listed in Table 1.

A deficiency of heterozygotes could be due to:

(*a*) effective recessiveness of the resistance gene and a changing gene frequency (as indicated by Muir);

(*b*) a mixture of two or more populations carrying the resistance gene at different frequencies (as discussed);

(*c*) a lower fitness of R+ than of RR or ++ genotypes;

(*d*) misscoring of genotypes.

Populations have been observed to be polymorphic for the dieldrin-resistance gene, even when there is no history of exposure to the insecticide, so that heterozygote disadvantage is unlikely. Misscoring cannot be

[b] See also: WHITE, G. B. ET AL. *Review of cytogenetic studies on anopheline vectors of malaria.* WHO unpublished document, WHO/VBC/75.538, 1975.

Table 1. The observed frequencies of dieldrin-resistant phenotypes in *Anopheles gambiae*, based on the results of field tests, compared with the genotype frequencies expected on the assumption of Hardy-Weinberg equilibrium[a]

Strain		RR	R+	++	Frequency of resistance gene (*q*)
Kano 1957 (Nigeria)	Observed	8	22	70	0.19
	Expected	3.6	30.8	65.6	
Kano 1959	Observed	50	42	8	0.71
	Expected	50.4	41.2	8.4	
Kaduna (Nigeria)	Observed	45	29	26	0.59
	Expected	35.4	48.2	16.4	
Didi and Guena (Upper Volta)	Observed	10	16	74	0.18
	Expected	3.3	29.5	67.2	
Dande (Upper Volta)	Observed	17	16	67	0.25
	Expected	6.3	37.5	56.2	
Dougoumato (Upper Volta)	Observed	3	9	88	0.08
	Expected	0.6	13.9	85.5	
Karankasso (Upper Volta)	Observed	1	10	89	0.06
	Expected	0.4	11.3	88.4	
Tangrela (Upper Volta)	Observed	0	5	95	0.03
	Expected	0.1	4.9	95	
Freetown (Sierra Leone)	Observed	12	54	34	0.39
	Expected	15.2	47.6	37.2	
Man (Ivory Coast)	Observed	4	10	86	0.09
	Expected	0.8	16.4	82.8	
Bougoumi (Mali)	Observed	35	17	48	0.44
	Expected	18.9	49.2	31.9	
Kisumu (Kenya)	Observed	33	27	40	0.47
	Expected	21.99	50.2	28.6	
El Karo (Sudan)	Observed	60	23	17	0.72
	Expected	51.1	40.8	8.1	

[a] Data taken from Brown & Pal (*1*) & WHO unpublished document, WHO/MAL/75.586, 1975.

ruled out in view of the practical difficulties involved in testing, but the most important problem is the difficulty of distinguishing change in frequency under selection from the effect of population mixing.

CONCLUSIONS

There is a need to measure selection pressures acting on resistance genes because this may allow prediction of future changes in resistance and may aid in the planning of control programmes.

We have described a method of estimating *s*, the selection coefficient, from observed changes in resistance-gene frequency over fixed regular intervals based on the fact that the frequencies represent the survivors after selection. This gives a better estimate than the method based on the assumption of Hardy-Weinberg proportions.

Although in the example given, our results are very similar to those of Muir,[c] this will not always be the case, particularly with larger values of *s*, i.e., greater differences in fitness between genotypes.

Regarding estimations of *s* based on deviations from Hardy-Weinberg equilibrium, the similarity of the results of selection favouring an effectively recessive resistance gene and population intermixture was noted. This makes estimation of selection by this method difficult. Even when population intermixture can be excluded, this method only works when the three genotypes can be separated and selection favours the recessive genotype, i.e., the heterozygotes are killed. In the case of partial dominance, the heterozygote would not show a reduction in frequency in relation to the Hardy-Weinberg expectation.

[c] See footnote *a*, page 129.

For both methods of estimation it is important to know the effective dominance of the resistance gene under field conditions, i.e., whether or not the heterozygote is killed. The importance of this factor for the evolution of resistance was noted earlier by Davidson (*3*).

RÉSUMÉ

ESTIMATION DES PRESSIONS SÉLECTIVES S'EXERÇANT SUR LES GÈNES DE LA RÉSISTANCE AUX INSECTICIDES (NOTE)

Il est utile de pouvoir mesurer les pressions sélectives s'exerçant sur les gènes de la résistance chez les insectes vecteurs de maladies, car il est ainsi possible de prévoir les modifications futures de la fréquence de ces gènes et d'envisager les moyens d'empêcher dans la mesure du possible l'apparition de la résistance. La présente note décrit une méthode d'estimation des coefficients de sélection, étant donné deux ou plusieurs fréquences phénotypiques après sélection, et connaissant le nombre de générations séparant ces fréquences. Cette méthode est appliquée aux données publiées sur *Anopheles labranchiae* soumis à une pression sélective de DDT.

Par cette méthode, on a évalué à 31-38% l'aptitude relative (1-*s*) des individus sensibles par rapport aux individus résistants. Il s'agit d'une estimation annuelle, mais si on connaît le nombre de générations par an, il est également possible de calculer une valeur par génération. L'article donne le programme informatique permettant de procéder à ces estimations. Les calculs sont basés sur le gène effectivement récessif, c'est-à-dire sur l'hétérozygote qui est tué par la dose appliquée sur le terrain.

On peut aussi estimer la pression sélective en mesurant les écarts par rapport aux prévisions de fréquences géniques selon la loi de Hardy-Weinberg. Par cette méthode, l'aptitude relative (1-*s*) des invididus sensibles chez une population de *Anopheles funestris* soumise à une pression sélective de dieldrine dans le nord de la République-Unie du Cameroun a été estimée à 40%. On se heurte toutefois à des difficultés avec cette méthode, car les mélanges de population peuvent conduire à des écarts qui simulent l'effet de la sélection. Des exemples sont discutés pour *Anopheles gambiae*, pour lequel des mélanges de population peuvent se produire et où l'on observe une diminution du nombre d'hétérozygotes pour le gène de la résistance à la dieldrine.

Pour ces deux méthodes d'estimation, il est indispensable de connaître la dominance effective du gène de la résistance dans la nature, c'est-à-dire de savoir si l'hétérozygote résistant est tué ou non. Ce facteur est en effet important dans la lutte contre l'apparition de la résistance.

REFERENCES

1. BROWN, A. W. A. & PAL, R. *Insecticide resistance in arthropods*, 2nd ed., Geneva, World Health Organization, 1971 (Monograph Series, No. 38).
2. COLUZZI, M. ET AL. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Transactions of the Royal Society for Tropical Medicine and Hygiene*, **73**: 483-497 (1979).
3. DAVIDSON, G. The practical implications of studies on insecticide resistance in Anopheline mosquitoes. *Indian journal of malariology*, **12**: 413-422 (1958).
4. DAVIDSON, G. The five mating-types in the *Anopheles gambiae* complex. *Rivista di malariologia*, **43** (4-6): 167-183 (1964).
5. DAVIDSON, G. ET AL. The *Anopheles gambiae* complex. In: Wright, J. W. & PAL, R., ed., *Genetics of insect vectors of disease*, Amsterdam, Elsevier, 1967, pp. 211-249.
6. LI, C. C. *Population genetics*, Chicago, University of Chicago Press, 1955.
7. PATERSON, H. E. ET AL. *Aedes aegypti* complex in Africa. *Nature (London)*, **259**: 252 (1976).
8. SCOTT, J. A. & MCCLELLAND, G. A. H. Electrophoretic differences between sympatric ecotypes. *Nature (London)*, **256**: 405-406 (1975).
9. TRPIS, M. & HAUSERMANN, W. Demonstration of differential domesticity of *Aedes aegypti* (L) (Diptera, Culicidae) in Africa by mark-release-recapture. *Bulletin of entomological research*, **65**: 199-208 (1975).
10. WHITE, R. J. & WHITE, R. M. Some numerical methods for the study of genetic changes. In: Bishop, J. A. & Cook, L. M., ed., *Genetic consequences of man-made change*. London, Academic Press, 1981, pp. 295-342.
11. WOOD, R. J. & MANI, G. S. The effective dominance of resistance genes in relation to the evolution of resistance. *Pesticide science*, **12**: 573-581 (1981).

Annex 1

COMPUTER PROGRAM

A computer program in BASIC for estimating the selection coefficient of a dominant susceptible gene, using the frequency of resistant homozygotes, is given below. The input is (1) the number of occasions for which data are available, (2) the number of generations between the estimates, and (3) the frequency of the resistant class on each occasion. Examples are given showing the estimation of *s* for Muir's data on *Anopheles labranchiae* assuming (1) one generation between records, and (2) four generations between records. No allowance is made for variation in sample size. The BASIC language is as implemented on Apple II computers.

```
10  PRINT
20  PRINT
30  INPUT "HOW MANY PHENOTYPE
      ESTIMATES?"; J
40  DIM Q (J)
50  DIM B (J)
60  INPUT "HOW MANY GENERATIONS
      BETWEEN ESTIMATES?"; N
70  PRINT "ENTER THE"; J; "RESISTANT
      HOMOZYGOTE FREQUENCIES"
80  FOR I = 0 TO J-1
90  INPUT B(I)
100 NEXT I
110 PRINT
120 S = LOG (B(J-1)/B(0))
130 S = S/(N*(J-1))
140 S = 1-1/EXP (S)
150 GOSUB 320
160 Y1 = Y
170 S1 = S
180 S = S+.1
190 GOSUB 320
200 Y2 = Y
210 S2 = S
220 S = (S1*Y2-S2*Y1)/(Y2-Y1)
230 S1 = S2
240 S2 = S
250 Y1 = Y2
260 GOSUB 320
270 Y2 = Y
280 IF ABS (Y2-Y1)>.00001 THEN 220
290 PRINT
300 PRINT "ESTIMATED VALUE OF S = "; S2
310 GOTO 430
320 Y = 0
330 FOR I = 0 TO J-2
340 Q(I) = (1-S)*B(I)/(1-S*B(I))
350 FOR K = 1 TO N
360 Q(I+1) = Q(I)+ (SQR (Q(I))*(1-SQR
      (Q(I))))*(1-S)
370 Q(I+1) = (Q(I+1)/(1-S*(1-Q(I))))^2
380 Q(I) = Q(I+1)
390 NEXT K
400 Y = Y+(1-S)*B(I+1)/(1-S*B(I+1))-
      Q(I+1)
410 NEXT I
420 RETURN
430 END
```

```
RUN

HOW MANY PHENOTYPE ESTIMATES?  3
HOW MANY GENERATIONS BETWEEN
  ESTIMATES?  1
ENTER THE 3 RESISTANT HOMOZYGOTE
  FREQUENCIES
?.32
?.48
?.7

ESTIMATED VALUE OF S =  .663953324

RUN

HOW MANY PHENOTYPE ESTIMATES?  3
HOW MANY GENERATIONS BETWEEN
  ESTIMATES?  4
ENTER THE 3 RESISTANT HOMOZYGOTE
  FREQUENCIES
?.32
?.48
?.7

ESTIMATED VALUE OF S = .220765001
```