Methods for Estimating Risk of Chemical Injury: Human and Non-human Biota and Ecosystems
Edited by V. B. Vouk, G. C. Butler, D. G., Hoel and D. B. Peakall

Methods for Measuring the Effects of Chemicals on Terrestrial Animals as Indicators of Ecological Hazard

Stuart Dobson

ABSTRACT

The limitations of currently available test methods should be clearly recognized. Acute or subacute toxicity tests are bioassays with death as the endpoint; they were not designed as tests for environmental hazard. Extrapolation and manipulation of data from these tests can be misleading. Test design can radically alter the result obtained. Known sublethal effects which are easily measured can be used to identify how widespread are the results of chemical usage. This is useful but it should be recognized as a monitoring exercise only. The only predictive value here is for very closely related compounds. More elaborate tests are required to give truly predictive information. It is clearly essential that the behaviour and physiology of the test animal in the test situation be as fully understood as possible to allow interpretation of any effects. The way in which environmental variables affect the test situation must also be known. The best way forward in predictive testing is the establishment of such multidisciplinary test systems.

1 INTRODUCTION

> If what a prophet proclaims . . . does not take place or come true, that is a message that God has not spoken. That prophet has spoken presumptuously. Do not be afraid of him.
>
> *Deuteronomy* **18**: 21

Ecotoxicoloy is very much in its infancy. Even the mother science, ecology, cannot be called mature, as we are only beginning to understand its basic principles. Any consideration of methods for assessing ecological damage must begin with this fact clearly in mind. At best, test methods can only reflect current knowledge. Assumptions made in their design often reveal a superficial view of the various scientific disciplines ultimately involved, and often lag behind the

latest discoveries and ideas in relevant fields. Given unlimited resources for testing, unlimited time to conduct those tests and the present state of knowledge of how organisms coexist in natural communities and in relationship to seasonal and other environmental changes, we could seldom with any confidence predict what the overall effects of a new chemical would be. In this situation we must seriously question exactly what our tests are measuring and whether this corresponds with the answers that we really want or need.

We must also distinguish between (1) true predictive tests which have been designed to assess the likely hazard of chemicals to animals before these chemicals have come into widespread agricultural or industrial use and (2) tests which seek to assess the range of effects which have arisen after such chemicals have been used, and proved harmful. The bulk of all reports in the literature deals with the latter type, because predictive tests were either not done, or if they were, they failed to give an indication of hazard which became manifest only after use. Biologists have tended to overestimate the importance of the limited success achieved in this area, with the benefit of hindsight, in assessing the potential success in the more important field of early prediction. In the light of this experience it is difficult to see any justification in the present trend towards trying to obtain more and more information from less and less testing by correlating results from different types of tests.

For a variety of reasons the design and use of tests for assessing hazard to terrestrial animals is the least developed of any of the areas being considered here. The consideration of mammalian toxicity has tended to merge with testing to assess hazard to domestic animals and human health, though the relevant questions to be asked in the two areas are quite different. Regulatory bodies exist to monitor the quality of water. Widespread monitoring of terrestrial habitats is not generally undertaken. Changes in land use and agricultural practice make it difficult to determine what role, if any, chemicals have had on populations of animals because they are often affected by several factors at once. It is difficult to say, therefore, just how many terrestrial 'incidents' there have been and how seriously animals were affected.

Any current explanation of the overall impact of chemicals which have caused environmental problems is not necessarily complete. Test methods have tended to derive eclectically from these current explanations. The inevitable result is that assessments of the value of methods in predicting hazard often rest on circular arguments. Perhaps a more reasonable approach to testing is the consideration of how animals interact with their environment, positive identification of the questions which need to be answered and a dialectical design of test methods. Chemicals which are deliberately applied to the environment or are likely to accidentally contaminate terrestrial habitats are an essential part of modern agriculture and industry. The more information we can obtain from test methods, the less likely we are to condemn chemicals, developed at considerable cost, which could be used safely under controlled conditions.

In this chapter I intend to concentrate on testing in birds. More information is available on the relevant ecology and physiology of birds than any other group, they have been seriously affected by chemical pollutants and they will serve to illustrate the general principles involved. Moreover, internationally recognized tests are regularly conducted on bird species.

2 LETHAL EFFECTS

Results of acute and subacute toxicity tests comprise most of the information available to assess hazard to terrestrial animals. Protocols now used in testing were originally bioassay methods, with death the observed endpoint, to determine the relative potency of different substances. They were never, therefore, specifically designed to provide information on ecological hazard. Whilst many attempts have been made to show that definitive information on environmental hazard can be obtained from such tests, they are really little more than screening exercises which are unlikely to give a reliable guide to hazard in the real world. They are most valuable where the hazard of one substance is well known from wide use and research on its effects, and a second, chemically very similar, substance is being assessed. They are least valuable when used on completely novel chemicals. Their implicit limitations should be clearly in mind when results are interpreted.

If they are to give any useful guide, careful consideration must be given to their design. Two types of short-term toxicity tests have been widely used for birds:

(1) Acute tests: a single dose is given directly into the upper digestive tract.
(2) Short-term dietary tests: contaminated food is fed over a period of several days.

These two tests ask different questions about the toxicity of the material and can give radically different results (Kenaga, 1973). The acute test should only seek to find the effect of short-term contaminants of the environment and, within the limits of species variability, should give good information on the likely hazard of accidental spills of non-persistent materials on local populations. The dosing method is a good way to give a rapid increase in blood levels of material over a short time period, a situation which can arise from the use of certain pesticides. Kenaga (1973) produced a powerful argument showing that they can give very misleading results if taken in practice as indicators of toxicity of contaminants, which are widespread, persistent or both. Using five compounds tested in two bird species, he showed that the acute oral test ordered the chemicals (in decreasing order of toxicity) as Dowco 139 > dimethoate > Dowco 179 > dieldrin > DDT, whereas the dietary testing gave DDT > dieldrin > dimethoate > Dowco 170 > Dowco 139, almost the reverse order. The reason for the difference is this: the range includes materials which are both readily absorbed and readily metabolized and others which are poorly absorbed

and persist in animal tissues. The dietary test is here clearly giving better indications of real effects of pesticides in use. Using information from the literature on food intake of various bird species and on residues in various food types after spraying with pesticides, Kenaga argues that the dosage received by individual birds in terms of mg/kg/day can be readily calculated in situations where direct measurment is not possible. This view has not yet been extensively tested empirically, even for well-known environmental contaminants. Many chemicals have been examined in a variety of bird species in controlled conditions, but have not been investigated in a range of species of different feeding habits in field trials. Such a simple model also fails to consider interactions between chemicals and between chemicals and feeding behaviour of birds. Exposure will be lower than expected if the birds have an aversion to contaminated food and can find an alternative without chemical residue. Kenaga, however, does highlight the relationship between body weight of birds and the food eaten as percentage of body weight. Most of the test species commonly used eat a small percentage of their body weight each day when adult. Kenaga recommends either the use of small birds in dietary toxicity tests or the young of larger birds where a more typical percentage of body weight is consumed. In practice, young birds have tended to be used. It is difficult to see a physiological comparison between the young of large birds and adult small birds. Although different ages of birds are included, no account has been taken of the seasonal cycles of birds. This interesting approach requires more information on the control of feeding before it can be established as a widely useful concept. Even if it proves to be generally useful, establishment of a straightforward method of estimating exposure only answers half the question which needs posing. Widespread species variability in the effects of chemicals is a well-known phenomenon which has plagued hazard assessment. A systematic study of this has not yet been attempted. We simply do not know if a straightforward relationship between a knowledge of the detailed breeding biology, physiology and ecology of a variety of bird species with a good estimation of the variety of exposure experienced by these species will give us a method of more general prediction of effect throughout the group. This should be the next major step forward and involves a multidisciplinary approach to the design of test methods.

The endpoint of a short-term toxicity test is a dose–response curve. The most frequently quoted characteristic of this curve is the LD_{50} (or LC_{50}), the dose which kills half the test animals. This has become a much abused term. It is often taken as a measure of relative toxicity in itself outside the context of the assay from which it was obtained and sometimes, implicitly, as a measure of absolute toxicity. The dose–response relationship is most usually describable by a logistic model. Four parameters are necessary to fully define such a curve. The LD_{50} is the point of least error on the curve and represents a purely arbitrary position on one axis to define the relative positions of two curves on the second axis. To fully describe the curve it needs the definition of upper and lower asymptotes and a

slope value giving the rotation of the line around the LD_{50}. Two LD_{50} values can be compared and give information on the relative potency of test substances only if the other curve parameters are comparable, in practice only if the two lines are parallel in the same assay. If the curves are not parallel, then the comparison at LD_{50} is quite different from comparisons at LD_{10}, LD_{90} or any other points on the curve. It is therefore necessary to run test substances against a standard to check for non-parallelism each time that an assay is performed or at least to give unequivocal evidence that the behaviour of a standard substance is so invariably the same that it does not need repeating (Finney, 1964).

Using a series of LD_{50} values from a given test method for substances of known hazard or safety and determining a cut-off LD_{50} value which will define hazard or safety is not justifiable on any theoretical grounds. The degree of certainty of definition of hazardous or safe can only be expressed in terms of the set of chemicals used to generate the cut-off point. It fails to take into account any other characteristics of the curves. To suggest that it works in practice and is therefore useful is self-deception and ascribes properties to the test method which it can never possess. The cut-off value derived in this way fails with its first mistake; its modification downwards with evidence of new substances does not change its implicitly arbitrary nature. Similarly, the correlation of LD_{50} values from two different assays cannot be used as an argument that a second test is unnecessary, that information from one test can be used to determine what results would come out of the second. Such an exercise can be useful but not to show that one gives no further information than the other. This may be the case, but more evidence is needed to make such a decision. The outliers from such a correlation are often more significant than those points lying close to the line. Such correlations may be useful in identifying unusual substances and suggest further levels of testing for outliers. Caution is also needed in interpreting the slope of the dose–response line. A substance whose toxicity is attenuated by, for example, its fat solubility, will tend to show a lower slope than another with lower fat solubility because some of the material is deposited in fat stores and temporarily detoxified. The conclusion that a flat curve indicates likely chronic effect is not justified. Low slope may indicate attenuation or result from atypical toxic mechanisms. Curves with the same slope do not necessarily indicate similar handling of two materials by the test organism or imply a similar toxic mechanism.

Plotting dose–response curves, or more usually their derived LD_{50} values, against time (a survivorship curve) can give information on the distribution of toxic response within the test population. A clear asymptote can indicate a safe level of exposure. This is only true within the conditions of the test. A wild population may not show the same distribution of response and conditions in the wild may radically alter the shape of the curve. Extrapolations of partial curves potentially increase the error beyond the usefulness of the original concept. An asymptote must be demonstrated, not deduced. Apparently trivial changes in

housing or dosing can radically alter response. Pigeons caged as male/female pairs show clear toxic signs at low doses of organochlorines. Individually caged birds visually isolated from other pigeons show neither lethal nor sublethal effects of the same materials at dose levels at least 20 times higher (Dobson *et al.*, 1976).

We must conclude, therefore, that simple bioassays for lethality will give us relatively little useful data from which to assess environmental hazard.

3 SUBLETHAL EFFECTS

If a chemical is known to have a particular effect which is easily measured, then it becomes fairly straightforward to look at species differences in susceptibility in captive animals. It also becomes easy to monitor its effects in the wild simply by looking at the severity of this one effect in relation to environmental concentration under different conditions. There are many examples of this in practice and this is an area where we have had the most notable success in extrapolation from laboratory to field. We must, however, distinguish this kind of study from generalized predictive tests.

A single example will suffice since all testing is specific to a particular group of chemicals. Organophosphorus compounds (OP) inhibit esterases. The biochemical mechanism is fully understood. The likelihood of materials causing this inhibition is, therefore, completely predictable from chemical structure. Blood esterase activity is easy to measure and can be used as a measure of how widespread the effect of OP spraying has been. The function of blood esterases is not known and their inhibition is not, in itself, a fatal lesion. It is known that blood esterase activity can be reduced to zero, temporarily, with no ill effects. The fatal lesion is presumably on acetylcholinesterase in the central nervous system or at crucial nerve–muscle connections. Blood esterase measurement is thus, like shell-thinning response to DDE, a convenient measure but not indicative of the full range of toxic response. Clearly, neither blood esterase measurement nor eggshell thinning are generally applicable as predictive screens of toxic effect.

Breeding success is of obvious importance in maintaining animal populations. Detailed studies of reproduction in laboratory animals have long been a basic component of toxicological investigation, with the intention of identifying potential problems in domestic animals and humans. Little has been done by way of predictive testing on wild mammals. A test is routinely done to check for reproductive effects of pesticides in birds.

The interaction between environmental factors and breeding performance in birds has received considerable attention in recent years (Murton and Westwood, 1977). It is still not possible to say exactly which environmental variables are responsible for starting and ending breeding in more than a very few species of birds. Even in well-studied species, it is not yet clear how different environmental cues act. The physiological mechanisms underlying the percep-

tion of environmental stimuli and the translation of this into gonadal activity are also imperfectly understood. Whilst light has been studied more extensively in the laboratory than any other variable, other factors such as food supply, temperature, rainfall and nest sites influence gonadal development in different bird species. Field studies have accentuated these other factors rather than photoperiod.

The usual tests on bird reproduction all basically apply a standard photoperiodic stimulus and compare breeding performance in terms of egg production and hatchability over a period similar to a normal breeding season. Food consumption is usually monitored to assess dosage of chemical taken by the birds. Following the experience with DDT metabolites, eggshell characteristics are monitored. These tests can be viewed as taking a simple model of bird breeding, and monitoring the overall effect of a potentially harmful material.

Selection of bird species for reproductive testing presents a major problem. Those species commonly used are of necessity unrepresentative, indeed given the diversity of habitat and breeding cycles within the group it would be difficult to think of any representative species. Species which are easy to breed in captivity are largely domesticated or semidomesticated varieties of groups selected in the past for domestication because they had unusually long breeding seasons and would therefore give greater productivity in captivity. Varieties have been selected further for greater egg or meat production. Almost all species used regularly in testing are precocial. All test species in captivity can have long breeding cycles without a typical refractory phase (during which environmental stimuli which would normally cause reproductive development are inoperative) or with a very abnormal one. The breeding of the birds will thus be maintained so long as the photoperiod is kept long and there is no sudden deterioration in temperature or diet. Egg production continues if the eggs are regularly removed. Moult can be delayed indefinitely in most of the species used (whilst breeding continues), or is atypically continuous throughout breeding in the case of doves. Economic importance of domestic birds is the justification for the almost exclusive use of these species, but we should not assume that data gained from these species can be readily extrapolated to birds in the wild.

A wide variety of photoperiodic stimuli have been used in tests. These seem often to have been derived from light schedules which have been shown to maximize egg production in domestic chickens. If a test is looking for different responses to a light stimulus, then clearly the photoperiod used must be standardized for any test species if results are to be comparable. Most commonly used test species show a steep photoresponse curve, a plot of gonadal growth rate against day length (where this has been determined), that demonstrates a clear threshold requirement for day length and only a relatively small increase over this length of day is required to give maximum rate of gonadal growth. In more typical wild species the photoresponse curve is much flatter though it has not been plotted for many species. The rate at which the gonads grow and at which

the other physiological changes associated with breeding occur can have profound effects on the way the animal handles a chemical. A fat-soluble material shows this most clearly, since in some species loss of stored fat is associated with the development of breeding condition. The handling of all toxic materials is potentially affected since liver metabolism associated with steroid secretion is altered during this phase. Many wild mammals have been shown to depend on photoperiod to determine their time of breeding. Although some of these are easy to keep and breed in the laboratory, they have never been used in the systematic testing of chemicals. Whereas domestic birds are merely atypical in their photoperiodic response, most laboratory mammals have lost the response completely in the course of domestication. A loss of photoperiodic response or a shift in the response curve often occurs in inbred colonies of birds. If the test method is used to predict effects in the wild, the range of photoperiodic response in wild stock must be maintained in the test populations by regular outbreeding. The commonly applied criterion that test birds should be phenotypically similar to wild popultions clearly does not satisfy this requirement.

The role of food in the normal initiation of breeding in birds and mammals has received considerably less attention than that of light, at least in laboratory situations. In domestic birds a reduction in the amount of food available will cause egg laying to stop and moult to start. A very considerable reduction in food consumption is necessary to interfere seriously with the gonadal growth response to a photoperiodic stimulus in galliforms. This does not appear to be the case in more typical bird species where food and light may have equal importance in initiating breeding. Very small reductions in food intake can inhibit gonadal growth in starlings (unpublished results). Under constant conditions the amount of food eaten each day is controlled with extreme accuracy. The weight of food consumed increases with day length (Figure 1). Limiting the length of the feeding day does not alter the amount eaten but can prevent gonadal development completely (Westwood and Dobson, 1980). Timing of food availability or consumption is then clearly important. Considering the importance of feeding behaviour and its control, little has been incorporated into test methods regarding food. Food eaten is monitored in both dietary tests for lethality and reproductive tests largely to establish dose of chemical received. The monitoring is usually done per pen of birds, with no information on individual variation. It is difficult to say whether interference with normal feeding has played any significant role in the effects on chemicals in the wild, more because this has not been seriously examined than for lack of indirect evidence of its involvement. The organochlorines appeared in some experiments to reduce appetite and foraging behaviour. Any fat-soluble material will be at its highest blood level at dawn as overnight fat stores are mobilized. A narcotizing effect at this time of day would itself inhibit reproductive development if the starling results are generally applicable. In both birds and mammals body weight and stored resources are controlled with great precision though the mechanisms by which this is achieved

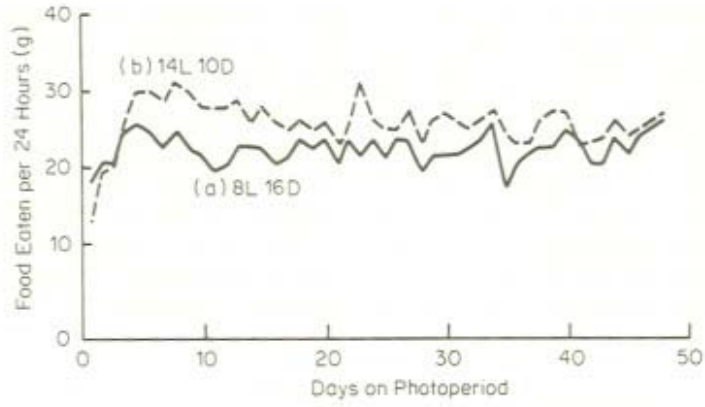

Figure 1 Food eaten per day by starlings held on two different photoperiods: (a) 8 hours of light per day and (b) 14 hours of light per day

are poorly understood (Silverstone, 1976). Any interference with its control could have severe ecological effects and test methods should examine this aspect more closely.

It is important in reproductive studies to distinguish between effects on the adults, which cause a reduced hatching success or survival of the young, and effects directly on the eggs themselves caused by chemical entering the eggs from the female, as pointed out by Peakall and Peakall (1973). Either natural or artificial incubation of eggs or a separate investigation in injected eggs must be carried out (Dunachie and Fletcher, 1969). The implications on population effects in the wild of the two effects are quite different.

Observation of behaviour is superficially an attractive option for testing. It is certainly most important in all testing programmes that overt behavioural changes are noted. Minor behavioural aberration in a test cage may translate into a serious hazard in the wild. Behaviour is very sensitive to any general or specific toxic effect of chemicals in animals. Unfortunately, it is time consuming to perform detailed observation of, say, reproductive behaviour, and almost impossible to extrapolate the results of simple behavioural observation to the real world. The ideal test gives more information than simply implying a deleterious effect. In reproduction, there is considerable interplay between internal physiological change initiating behaviour and external reaction to that behaviour. We understand very little of how hormones stimulate behaviour patterns or of how behavioural response changes hormone secretion. DDE causes a reduction in nest-building behaviour in pigeons (Figure 2(a)). This appears to be a species-specific reaction to the material; birds of prey severely affected by DDE appear to nest-build normally. No observations of field effects

on nest-building have been reported for pigeons. At the same dose, DDE tends to cause a delay in egg laying. Measurement of hormones in the female shows that the lack of stimulus for oviposition normally given by a completed nest delays her laying of the egg (Figure 2(b)). Without measuring all of these factors, a misleading picture of the effects of DDE on birds could emerge from an oversimple test. Caution should therefore be exercised in designing test methods so that results can be interpreted within the context of the test itself, that is that the detailed physiology and behavioural input of the system is sufficiently understood that the information it gives can be interpreted. Extrapolation of imperfectly understood data to the field situation is worthless.

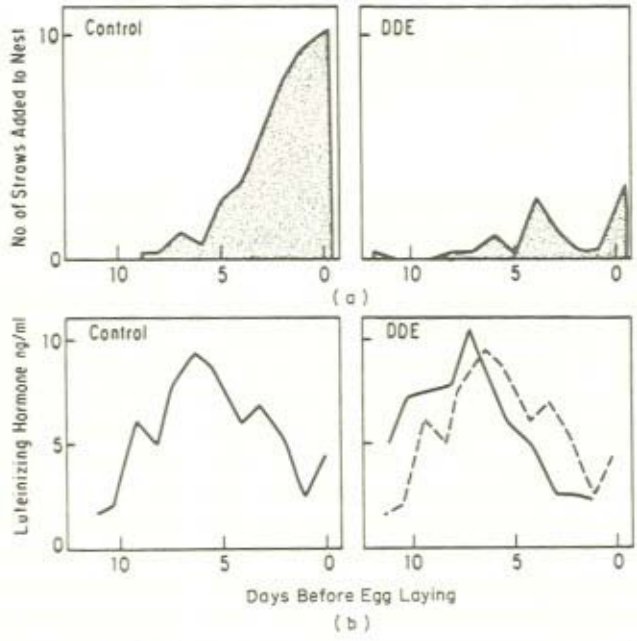

Figure 2 (a) Nest-building activity measured as number of straws added to a nest bowl by feral pigeons; control and DDE treated. (b) Blood levels of luteinizing hormone in the same birds as 2(a) (control curve repeated hatched on the plot of DDE-treated birds)

4 INDIRECT EFFECTS—ECOLOGICAL INTERACTIONS

Many chemicals are associated with deliberate attempts to alter land management. This management obviously has a profound effect on the area's fauna and flora, but this effect is beyond the scope of hazard testing. Where there is a target

organism and it is hoped to specifically control it without seriously disrupting populations of other animals, testing has concentrated on direct toxicity of the control chemical in non-target species, usually vertebrates. Wide-scale testing of indirect effects, the removal of a food source for example, has not been seriously attempted except by fairly superficial direct observation in field trials. We here come into the area of modelling of communities. An extensive literature exists which is impossible to review here. We can consider a general principle.

It has proved impossible over wide areas to eradicate invertebrate species deliberately. The testing of toxicity to invertebrate species is therefore concerned with estimating the severity of temporary reductions in populations and the ecological implications of these temporary effects. Good predictive models exist for fairly simple situations such as the interaction between phytophagous insects, their host plant and a variety of environmental variables. Abundance of food and suitable weather conditions for its collection determine the time of onset of breeding in birds, size of clutch and ultimately breeding success. Two possible means exist for this interaction. Either the bird responds to direct experience of the food situation, it delays breeding and adjusts the commitment it makes to the production of young, or it is coadapted to respond to the same environmental variables which are used by its food organism. In the latter case its commitment is based on a prediction of what the food situation will be at some stage in the future. A chemical which reduces food availability clearly affects the overall breeding success and probable adult survival much more if the second method predominates in any particular bird species because its prediction will be seriously wrong. The conclusion to be drawn is that it is impossible to design test methods which will adequately predict hazard unless we have a much clearer picture of how the physiology and ecology of animal species interact.

5 GENERAL CONCLUSIONS

Tests for lethality over a short period of dosing are an essential first step to define very toxic and very non-toxic chemicals. Their design should be carefully considered to reflect likely hazard in the wild. In the wide grey area between lethal and harmless they are less useful. Their usefulness in this area cannot be improved by simple statistical manipulation of data.

Sublethal effects are difficult to evaluate in any situation which is not understood in considerable physiological detail. It would be more useful to look at a wide range of physiological and behavioural responses in a short-term test which puts a system through a full range of its responses than to attempt to simulate the process as it occurs in the wild. Clear effects could be tested afterwards for species specificity simply and cheaply and a good indication of what to look for in field trials established. Tests should take into account the full range of current knowledge on physiological mechanisms of control.

It is difficult to envisage a system of tests to assess all conceivable hazards to

communities. Better models of how communities of animals and plants interact should be established for a broader understanding of the wider prediction of effects.

6 REFERENCES

Dobson, S., Dobson, B. C., Murton, R. K., and Westwood, N. J. (1976). Physiological effects of organochlorine pollutants. In *Annual Report of the Institute of Terrestrial Ecology*, pp. 62–63. The Institute of Terrestrial Ecology, Abbots Ripton, Huntingdon, UK.

Dunachie, J. F., and Fletcher, W. W. (1969). An investigation of the toxicity of insecticides to bird's eggs using the egg injection technique. *Ann. appl. Biol.*, **64,** 409–413.

Finney, D. J. (1964). *Statistical Method in Biological Assay*. Charles Griffin and Co. Ltd., London: 668 pages.

Kenaga, E. E. (1973). Factors to be considered in the evaluation of the toxicity of pesticides to birds in their environment. *Envir. Qual. Saf.*, **2,** 166–181.

Murton, R. K., and Westwood, N. J. (1977). *Avian Breeding Cycles*. Clarendon Press, Oxford, UK: 594 pages.

Peakall, D. B., and Peakall, L. M. (1973). Effect of a polychlorinated biphenyl on the reproduction of artificially and naturally incubated dove eggs. *J. appl. Ecol.*, **10,** 863–868.

Silverstone, T. (Ed.) (1976). *Appetite and Food Intake: Report of the Dahlem Workshop on Appetite and Food Intake, Berlin, 1975, December 8–12*. Abakon-Verlagsgesellschaft, Berlin: 497 pages.

Westwood, N. J., and Dobson, S. (1980). Environmental variables and onset of breeding in the starling. In *Annual Report of the Institute of Terrestrial Ecology*, pp. 15–17. Institute of Terrestrial Ecology, Abbots Ripton, Huntingdon, UK.