Pesticides and Phytoseiid Mites: Strategies for Risk Assessment

FRANK M. BAKKER* AND JOSEP A. JACAS†

*Department of Pure and Applied Ecology, Section Population Biology, University of Amsterdam, Kruislaan 320, NL-1098SM Amsterdam, The Netherlands; and †ETSI Agrónomos, Unidad de Protección de Cultivos, Universidad Politécnica de Madrid, E-28040-Madrid, Spain

Received November 9, 1994

Laboratory toxicity trials may predict effects of chemicals under field conditions, but errors are inevitable. A chemical may be presumed harmless when in fact it has a detrimental effect, or it may appear highly toxic in the laboratory, but not in the field. Error rates depend on experimental setups, evaluation criteria, and ecological attributes, such as dietary range, of the organisms under study. The authors analyze results of standardized toxicity studies of pesticides on four species of predatory mites and assess the feasibility of drawing accurate conclusions from laboratory trials alone. This is by contrasting laboratory and field data, while varying interpretation criteria. At a 5% critical error rate, it was found that correspondence between lab and field experiments is only obtained for products harmless to *Typhlodromus pyri*. For this species these constitute only 30% of the total number of products in our database. Outcomes from lab tests with *Amblyseius andersoni* correspond with field results (for *A. andersoni* and *A. finlandicus*) either for products yielding harmless or for products yielding harmful side effects. The decision rules required to reach either classification are not compatible and hinge on field thresholds that may be unrealistic. For *Phytoseiulus persimilis* only harmful insecticides and harmless fungicides enabled the setting of decision rules that resulted in correspondence between lab and field trials of more than 95%. Why these species require different interpretation criteria is discussed together with suggestions for improvement of existing test protocols and the feasibility of using indicator species. © 1995 Academic Press, Inc.

INTRODUCTION

Due to growing environmental concern the array of minimum safety data requirements for pesticide marketing is expanding. Recently impacts on terrestrial arthropods, particularly natural enemies of pests, have been included in registration protocols. Such studies are not new; integrated pest management has required these data for a long time. The list of available test protocols is certainly not exhaustive (see, e.g., Croft, 1990) and ranges from tests for specific situations to standardized protocols developed by expert groups (e.g., IOBC, 1988; Hassan, 1989, 1992a; Samsøe-Petersen, 1990). These protocols are now being discussed in the frame of harmonized European legislation regarding the marketing and use of pesticides.

For economical reasons there is a limit to the safety data set

industry can provide. Hence test protocols should be quick, inexpensive, and reliable. Both the use of indicator organisms and the use of sequential testing schemes may reduce testing effort to acceptable limits. Here existing protocols and the decision rules to interpret them will be discussed. The focus will be exclusively on four species of predatory mites (Acari: Phytoseiidae). Three of these are important in orchards and vineyards; the fourth is a widely used biocontrol agent in greenhouses. Their different ecology will be considered in the discussion.

In sequential testing protocols the decision to perform field trials depends on the outcome of tests performed at lower levels of complexity; in the case of phytoseiid mites this is usually the laboratory test. The need to perform field trials is determined by preset effect thresholds to which laboratory (lab) results are compared. Depending on these so-called trigger values lab results can be considered conclusive or inconclusive, which implies that they require testing under more realistic conditions (Fig. 1). The objective is to reduce the amount of field trials to the necessary minimum (maximum efficiency). There is however a trade-off because the number of erroneous decisions should also be minimized (maximum accuracy).

Two types of erroneous decisions are possible. First, harmlessness may be concluded at an early stage while the pesticide proves to be harmful under field conditions (type A error). Second, harmfulness may be decided where a compound appears harmless in the field (type B error). Clearly, from the viewpoints of environmental protection and pest management, errors of type A are more dangerous. However, for pesticide manufacturers, the chances of a product initially harmful in lab tests turning out to be of lesser effect in the field may provide ample reason not to accept this classification as conclusive. As the existing protocols were developed for advisory purposes a conflict of interest arises. Quick, efficient, and reliable screening procedures are always desired, but some will emphasize avoidance of type A errors, whereas others will focus on type B ones.

Industry, represented by the Beneficial Arthropod Registration Testing group (BART), has added to the discussion regarding the oncoming European harmonized legislation by proposing a testing protocol. In their approach, harmless is con-

sidered a conclusive outcome regardless of the level of testing in which this result was obtained (Barrett, 1992). According to BART, in regulatory testing, emphasis should be on demonstrating that products will not significantly affect our environment rather than on indicating their suitability for use in IPM schemes (Barrett, 1992). However, the authors feel that their approach is at odds with this objective as it avoids type B errors, but not type A ones. Enigmatically, the International Organization for Biological Control (represented by the Working Group "Pesticides and Beneficial Organisms"; PBO) (Hassan, 1989; Hassan *et al.*, 1985, 1991; Samsøe-Petersen, 1990) likewise only accepts harmfulness at the final stage of the sequential scheme. This is when field effects exceed 75% reduction in the number of predatory mites. On the other hand, in the approach of the European and Mediterranean Plant Pro-

tection Organization (OEPP) and the Council of Europe, low risk can only be assigned based on non-toxicity proven in lab trials, whereas high risk can be assigned at any stage (OEPP/EPPO, 1989, 1990, 1994). Similar strategies are found in Bakker *et al.* (1992), Carter *et al.* (1992), and Hassan (1992b).

The risk assessment approaches mentioned above have in common that results of toxicity trials are expressed as a single figure, E , denoting the effects by the product tested on life history (laboratory) or numbers of beneficials (field). By comparing E with preset threshold values (T), products are classified (Fig. 1). Because these T values may trigger the decision to perform further testing, they have been termed trigger values (e.g., Barrett, 1992). This terminology will be followed. Clearly, the setting of trigger values is crucial both for the efficiency of the test protocol and for the chances of erroneous

FIG. 1. Sequential test protocol validated in this paper. E refers to effect observed in trial and T to reference threshold value. We investigated which combinations of T_1 and T_3 would lead to type A error rates less than 5% and which combinations of T_2 and T_3 to type B error rates less than 5%. Questionmarks indicate uncertainty concerning accuracy of interpretation lab results.

decisions (Table 1). It has been suggested that trigger values should be kept flexible, while interpretation, especially of field trials, should be left to experts (Aldridge and Carter, 1992; Carter *et al.*, 1992). Although the need for a certain level of flexibility is defensible, it makes the requirements for safety data less transparent both for those submitting and for those having to interpret. The objective of this paper was to determine whether trigger values could be found that keep error rates within 5% limits. The results of this validation exercise will also be used to comment on the need of expert judgment and the feasibility of using indicator species.

MATERIALS AND METHODS

The validation exercise was based on comparisons of effects observed in lab and field experiments for the following species of predatory mites (Acari, Phytoseiidae): *Amblyseius andersoni* (Chant) (= *potentillae*), *A. finlandicus* (Oudemans) (only field), *Typhlodromus pyri* Scheuten, and for organophosphate-resistant *Phytoseiulus persimilis* Athias-Henriot in lab and greenhouse experiments. Field tests with *A. andersoni* were performed on four different crops, those with *T. pyri* in two. Lab tests with these species involved a strain with documented resistance to organophosphates and carbamates (Anber and Overmeer, 1988; van de Baan *et al.*, 1985) and a susceptible one. In addition, *A. andersoni* lab data were also compared to *A. finlandicus* field results.

The bulk of the data set used came from publications by PBO (Hassan *et al.*, 1985, 1987, 1988, 1991, 1994; Samsøe-Petersen, 1990). These publications categorize effects obtained at any level of testing in four classes (harmless, slightly harmful, moderately harmful, and harmful) determined by preset threshold values. In addition data from Oomen *et al.* (1991), Duso (1994) and unpublished information from Blümel (1991), Calis (1987), Grove (1993), and Overmeer and van Zon (1981, 1983) were used. These data enabled replacement of most of the results expressed according to the four IOBC categories by actual figures for the effect (*E*) as calculated by the standard procedure described by Overmeer and van Zon (1982).

Lab and field results were categorized in a series of 2×3 matrices. These were obtained by (1) taking two thresholds, T_1 and T_2 , that divided lab results in products being classified as harmless, inconclusive (requiring further testing), or harmful and (2) assuming a threshold, T_3 , dividing field results in either harmful or harmless. Hence, products always appeared in one of the six cells provided in Table 1. Type A error rates were defined as the fraction of products erroneously classified harmless (cell 4/cells 1 + 4), whereas type B errors were similarly computed from cells 3 and 6.

The different ecology of the species involved makes differential responses related to biocidal activity of products more likely in some cases than in others. For this reason, conditional dependence of the frequencies in the six cells on species and biocidal activity of pesticides (insecticidal, fungicidal) was

TABLE 1
Matrix Showing the Effect of Trigger Values (T) on Test Classifications

Lab results	Field results	
	Harmless	Harmful
Harmless	Cell 1 Correct	Cell 4 Type A error
Retest	Cell 2 Inconclusive	Cell 5 Inconclusive
Harmful	Cell 3 Type B error	Cell 6 Correct

Note. In our analyses cell contents were varied by changing trigger values. T_1 ranged from 9 to 59%, T_2 ranged from 59 to 99%, while T_3 had three values of 25, 50 and 75%.

tested in a contingency table. When no statistically significant interactions (i.e., no conditional dependence) were found, results were not treated separately. However, for biological reasons, results were never pooled beyond the species level. Consequently, the performance of the sequential testing protocol was validated using the following comparisons: *A. andersoni*—*Amblyseius* spp. ($n = 110$); *T. pyri*—*T. pyri* ($n = 105$); *P. persimilis*—*P. persimilis* separately for fungicides ($n = 22$) and insecticides ($n = 31$).

The final aim of this analysis was to assess whether there existed combinations of trigger values for which error rates were, or could be expected to be, below certain limits (e.g., a 5% error rate). The effect of varying T_1 and T_3 on type A error rates and of varying T_2 and T_3 on type B ones was analyzed. For this purpose error rates were calculated under all possible combinations of the following trigger values: T_1 from 9 to 59%, T_2 from 59 to 99% (both in steps of 5%), and T_3 as 25, 50, or 75%. Subsequently, the error rates thus obtained were plotted in the corresponding parameter space. Using the distance weighted least-squares technique (McLain, 1974; in Wilkinson 1988) for interpolation, a three-dimensional surface was derived. The lines connecting equal error rates (error isoclines henceforth) were projected on the plane defined by the trigger values.

Systat 3.2 for the Macintosh was used for all statistical analyses (Wilkinson, 1987), as well as for the drawing of contour plots (Wilkinson, 1988).

Although in this paper exclusive reference is made to active ingredients, it should be borne in mind that experiments were always performed with formulated products. When the same active ingredient appeared in different formulations the experiments were treated as independent replicates. This study also treated as independent tests cases where a single field trial was compared with lab experiments involving different strains of

the same species (e.g., when a susceptible and a resistant strain were simultaneously tested).

RESULTS

Using three trigger values leads to six possible classifications of a lab-field comparison. The study refers to each possible position in the 2×3 matrix as a cell (see Table 1). To get a first impression of the frequency distribution of outcomes over cells, lab and field data using the following trigger values were compared: T_1 (upper boundary for classifying lab results as harmless) = 30%, T_2 (lower boundary for classifying lab results as harmful) = 99%, and T_3 (field threshold dividing harmless and harmful results) = 50%. This was done separately for insecticides and fungicides. As Table 2 indicates insecticides are found mainly in cell 6 (harmful-harmful), whereas fungicides occur more in cell 1 (harmless-harmless). To test the null hypothesis that for both product types the distribution over cells is not species dependent, the frequencies were analyzed in a contingency table. This hypothesis had to be rejected because conditional dependence of outcome on species was demonstrated both for insecticides ($G = 70.45$, $df = 15$, $P < 0.0001$) and for fungicides ($G = 33.73$, $df = 15$, $P < 0.04$). Inspection of the data in Table 2 suggests that the frequency distributions of data obtained with the orchard/vineyard inhabiting species (*Amblyseius* spp. and *T. pyri*) are more similar to each other than to the one obtained with the greenhouse bio-control agent, *P. persimilis*. Indeed, analyzing results in a contingency table excluding *P. persimilis* demonstrated independence of outcome on species ($G = 5.04$, $df = 3$, $P = 0.169$).

Independence of outcome on species indicates that species may be substitutable and hence that the use of indicator species is a feasible option. To further test this hypothesis comparison was made of frequency distributions of outcomes over the six cells obtained with interspecific comparisons with frequency distributions obtained with intraspecific comparisons. Both for *T. pyri* and for *Amblyseius* spp. no significant differences were

found ($G = 3.448$, $df = 5$, $P = 0.613$ for *T. pyri* and $G = 1.891$, $df = 4$, $P = 0.756$ for *Amblyseius*). Hence, for the orchard/vineyard-inhabiting predatory mites discussed here, field results can be forecasted with the same level of accuracy regardless of the species tested in the lab. Table 3 illustrates the similarity in error rates obtained with both intra- and interspecific comparisons.

Amblyseius spp.

The data on *Amblyseius* spp. came from comparisons of *A. andersoni* lab results with field results obtained with either *A. andersoni* or with *A. finlandicus*. No lab data for *A. finlandicus* were available. Although the distribution of outcomes over cells was slightly different for both combinations ($G = 14.85$, $df = 5$, $P = 0.011$), error rates were identical. Consequently, in the analysis of the effect of trigger values on error rates pooled data has been used.

The results of varying T_1 and T_3 on type A error rates are provided in Fig. 2A. The 0.05 error isocline is obtained for field thresholds higher than 60%, regardless of the value given to T_1 . Thus, accurately concluding upon harmlessness after the lab test requires that even reductions of 60% or more, in the number of predatory mites under field conditions, are considered harmless side effects of pesticide use. This is incompatible with the parameter settings needed to keep type B errors within 5% limits. Here reductions over 30% must be considered harmful (Fig. 2B).

Consequently, depending on what is considered a realistic setting for the field threshold (T_3), lab tests on *Amblyseius* spp. can be used to safely conclude that a product will have either harmless or harmful side effects. In either case the sequential testing protocol is equally efficient. When the lab test is used to identify harmless products, a maximum of 43% of all products can be sorted out in the laboratory (for $T_1 = 59\%$ and $T_3 = 75\%$). For the widely accepted value of 30% for T_1 , 35% of all products will be definitely classified in the lab. On the other

TABLE 2
Example of Classifications Obtained for a Single Set of Boundaries (Trigger Values) Dividing Lab and Field Results

Cell	<i>Amblyseius andersoni</i> – <i>Amblyseius andersoni</i>		<i>Amblyseius andersoni</i> – <i>Amblyseius finlandicus</i>		<i>Typhlodromus pyri</i> – <i>Typhlodromus pyri</i>		<i>Phytoseiulus persimilis</i> – <i>Phytoseiulus persimilis</i>		Total	
	Ins	Fung	Ins	Fung	Ins	Fung	Ins	Fung	Ins	Fung
1	3	6	0	19	4	18	5	16	12	59
2	9	0	2	2	5	2	1	2	17	6
3	2	0	0	0	7	1	0	1	9	2
4	0	2	0	2	0	0	9	3	9	7
5	0	3	1	1	5	4	4	0	10	8
6	11	0	26	4	41	10	12	0	90	14
Total	25	11	29	28	62	35	31	22	147	96

Note. Two boundaries (30 and 99%, respectively) separate lab results in harmless (cells 1 and 4), inconclusive (cells 2 and 5), and harmful (cells 3 and 6). The 50% boundary was used to distinguish harmful (cells 4, 5, and 6) and harmless (cells 1, 2, and 3) field results. Only data on insecticides (Ins) and fungicides (Fung) are presented. Note the skewed distributions for both.

TABLE 3
Error Rates Obtained for Inter- and Intraspecific Comparisons
Involving Orchard/Vineyard-Inhabiting Mites

Predictor species	Predicted species	Accuracy	
		Type A errors (%)	Type B errors (%)
<i>Amblyseius</i> spp.	<i>Typhlodromus pyri</i>	0	10.9
<i>Typhlodromus pyri</i>	<i>Typhlodromus pyri</i>	0	13.1
<i>Typhlodromus pyri</i>	<i>Amblyseius</i> spp.	13.3	15.1
<i>Amblyseius</i> spp.	<i>Amblyseius</i> spp.	13.6	10.9

Note. Error rates are expressed as the percentage of harmless products (lab classification) being harmful in the field (type A error) or vice versa (type B error).

hand, when the lab test is used to identify harmful products (with $T_2 = 99\%$ and $T_3 = 25\%$), a maximum of 42% of all products can be definitely classified without performing a field trial.

Typhlodromus pyri

The lab trials on this species included a carbamate-resistant (R) strain and a susceptible one. The insecticides carbaryl and ethiofencarb appeared indeed harmless to the R strain in the lab, but they caused harmful effects in the field. Because no other major discrepancies were found between strains, pooled data were used but carbamates were excluded from the analyses.

The type A error isoclines provided in Fig. 3A clearly demonstrate that conclusions on harmlessness obtained in the laboratory are robust. They safely predict field results for any value of T_3 when T_1 is taken between 25 and 55%. In practice this means that between 25 and 29% of all insecticides and fungicides do not require retesting in the field. According to Fig. 3B, type B error rates cannot be brought below the 5% level. However, with T_2 at 89 or 94% and T_3 at 25% type B errors represented 6% of all harmful classifications. If the 6% level is taken for granted, the sequential testing protocol gains tremendously in efficiency because 61% of all insecticides and fungicides can now be sorted out as harmful in the laboratory. Thus, at a field threshold of 25%, almost 90% of the products

FIG. 2. Error isoclines calculated for products classified in the lab either as harmless for *Amblyseius* spp. (A) or as harmful (B). Data were obtained by varying the lab trigger values (T_1 or T_2) and giving T_3 the values 25, 50, and 75%. Hatched area shows which combinations lead to error rates lower than 5%.

FIG. 3. Error isoclines calculated for products classified in the lab either as harmless for *Typhlodromus pyri* (A) or as harmful (B). Data were obtained by varying the lab trigger values (T_1 or T_2) and giving T_3 the values 25, 50, and 75%. Hatched area shows which combinations lead to error rates lower than 5%.

can be finally classified after the lab test. As with *Amblyseius* whether a field threshold of 25% is realistic remains to be seen.

Phytoseiulus persimilis

Using pooled data, no combination of trigger values could be found that brought the type A error rate below the 20% level (Fig. 4A). Because the contingency table analysis indicated that the differential response of this species to insecticides and fungicides was more pronounced than in other species, contribution of these products on error rates was checked. This revealed that for certain values of T_1 and T_3 , type A error rates could be brought below the 5% level for fungicides, whereas it was impossible for insecticides to have error rates lower than 40%!

In the case of type B errors insecticides and fungicides were again evaluated separately. This demonstrated that while for fungicides type B error rates could not be brought below 35%, insecticides were always conclusive (Fig. 4B). Thus, with the lab test method employed (detached leaves), it appears possible to sort out harmful insecticides and harmless fungicides with A

and B error rates below 5%. As illustrated in Table 4, compatible evaluation criteria for both product types do exist. At a 50% field threshold, separating harmless from harmful side effects, 36 to 41% of the products are correctly identified after the lab trial. When T_3 is taken as 75% reduction in the number of beneficials, efficiency can be raised to 58% for various settings of T_1 and T_2 .

DISCUSSION

This study has demonstrated that, for the species discussed here, it is possible to accurately (i.e., with a 5% probability of error) conclude from the lab trial whether a product will have either harmless or harmful side effects under field conditions. The crucial question is how to define these two categories. The IOBC Working Group "Pesticides and Beneficial Organisms" (PBO) recognizes four possible outcomes of field trials. Independent of the species tested, reductions in the number of beneficials over 75% are considered harmful side effects, those less than 25% are classified as harmless, while the rest are intermediate (slightly and moderately harmful). If these values are adopted for the sake of the argument, it will be clear that the testing program carried out so far has performed very poor. When T_3 is set at 75% for harmfulness and at 25% for harmlessness and when the 5% error criterion is adopted, of the 276 trials analyzed by this study, only 29 products harmless for *T. pyri* and 12 insecticides harmful for *P. persimilis* were correctly identified in the lab trial. All other outcomes represent either products that had to be retested in the field or outcomes that are unreliable because error rates exceed the 5% level

TABLE 4

Combinations of Trigger Values That Allow for Simultaneous Evaluation of Insecticides Harmful ($n = 31$) and Fungicides Harmless ($n = 22$) for *P. persimilis* at Error Rates below 5%

T_1 (%)	T_2 (%)	T_3 (%)	Efficiency		
			Insecticides	Fungicides	Total
9	59–89	50	0.43	0.39	0.41
9	94	50	0.37	0.39	0.38
9	99	50	0.34	0.39	0.36
9	94	75	0.37	0.39	0.38
9	99	75	0.34	0.39	0.36
19	94	75	0.37	0.70	0.51
19	99	75	0.34	0.70	0.49
29	94	75	0.37	0.83	0.56
29	99	75	0.34	0.83	0.54
39	94	75	0.37	0.87	0.58
39	99	75	0.34	0.87	0.56
49	94	75	0.37	0.87	0.58
49	99	75	0.34	0.87	0.56
59	94	75	0.37	0.87	0.58
59	99	75	0.34	0.87	0.56

Note. Efficiency is the proportion of these products that will be definitely classified after the lab test. T_1 separates harmless lab outcomes, T_2 separates harmful ones, and T_3 divides field results in harmless or harmful.

FIG. 4. Error isoclines calculated for products classified in the lab either as harmless for *Phytoseiulus persimilis* (A) or insecticides as harmful (B). Data were obtained by varying the lab trigger values (T_1 or T_2) and giving T_3 the values 25, 50, and 75%. Hatched area shows which combinations lead to error rates lower than 5%.

(e.g., the 110 trials with *A. andersoni*). Table 5 summarizes the thresholds required for accurate predictions.

The finding that for all species both type A and type B error rates can be kept within 5% limits by the right choice of trigger values is an unexpected result. If lab tests ensure full exposure to high doses of pesticides under otherwise optimum conditions for the beneficial (worst case situation), type A errors should always occur at a much lower frequency than type B ones. This is the rationale for the testing protocols proposed by Hassan (1989) or Samsøe-Petersen (1990). These authors emphasize that, due to the expected overestimation of effects, lab trials can only sort out harmless products. The present analysis indicates that only trials with *T. pyri* convincingly support this point of view. However, while it is already unclear why lab toxicity tests with *T. pyri* should be more accurate than those with *A. andersoni*, it is even more enigmatic why conclusions from lab tests with *P. persimilis* should consider whether products have insecticidal or fungicidal properties. These discrepancies from what is to be expected according to the worst case approach point both at methodological flaws and at an incomplete understanding of mite ecosystems affected by pesticides.

Methodological Flaws

All lab data for the orchard inhabiting predatory mites, *A. andersoni* and *T. pyri*, used for this analysis were obtained on glass arenas. Those for *P. persimilis* came from trials on sprayed detached leaves. As discussed by Bakker *et al.* (1992) and Blümel *et al.* (1993) these testing methodologies do not ensure maximum exposure. Another, presumably more serious, problem is associated with the choice of test dose. According to recommendations by PBO, so far products were tested at the highest recommended field concentration using a deposit of 1–2 $\mu\text{l}/\text{cm}^2$. This represents low to very low volume rates in

field practice. Orchards often receive much higher volumes (≥ 1000 liters/ha) (Matthews, 1991). Thus, the lab tests underestimated the worst case with a factor of 10 or even more. Although this may explain why unacceptable type A error rates were found in certain cases and acceptable B error rates were found in others, it does not explain why this was not true for all species combinations.

Apart from these deficiencies in testing methodology, the quantitative interpretation of lab results is also open for discussion. The measure of effect (*E*) used in this paper takes into account mortality and reproduction according to a formula first proposed by Overmeer and van Zon (1982). In their approach average survival to adulthood and average oviposition figures are combined linearly. The intrinsic rate of increase r_m (Birch, 1948) might be a more realistic parameter on which to base quantitative estimations of the effects of pesticides and the setting of trigger values. Paying attention to species-specific attributes, particularly age at first reproduction, will indicate more clearly what the potential population dynamical consequences of pesticide use will be. Fast-growing species, such as *P. persimilis*, will be affected disproportionately more by delayed reproduction than, for example, *T. pyri*, which has a lower r_m . Thus, the effects presented in the literature used underestimated, in many cases, the effect for *P. persimilis* relative to the other species. This might partially explain the higher type A error rate found for this species. As a more ecologically based measure of effect and trigger values T_1 and T_2 , the authors propose difference in intrinsic growth rates due to the pesticide, scaled to its innate capacity.

Finally, the reproducibility of the field trials used as a reference for the lab tests depends on, e.g., the timing of application, application methods, environmental conditions, resistance status of the predators, and experimental setup. Hence, their interpretation is not unambiguous. Interpretation of field trials can be facilitated by using a harmful reference product instead of a predetermined threshold value. This has been proposed by OEPP/EPPO (1989, 1990, 1994). With respect to the cases studied in this paper, it should be kept in mind that the field tests with *A. andersoni*, *A. finlandicus*, and *T. pyri* were performed in different countries by different workers. Such systematic differences may also have contributed to the interspecific differences reported here. In these analyses, however, the authors have relied on the expert opinion of the researchers involved in the testing program.

TABLE 5

Possibilities for Drawing Conclusions from Laboratory Trials at Error Rates below 5% for Three Species of Predatory Mites

Species	Harmless	Harmful	T_1	T_2	T_3	Efficiency (%)
<i>Amblyseius</i> spp.	I + F		59		75	43
<i>Amblyseius</i> spp.		I + F		99	25	42
<i>T. pyri</i>	I + F	I + F ^a	25–55	89–94	25	90
<i>T. pyri</i>	I + F		25–55		any	29
<i>P. persimilis</i>	F	I	39–59	94	75	58
<i>P. persimilis</i>	F	I	9	59–89	50	41

Note. Lab results for *T. pyri* and *A. andersoni* were obtained with sprayed petri dishes (Overmeer and van Zon, 1983), those for *P. persimilis* were obtained with sprayed detached bean leaves (Oomen, 1988). The trigger values T_1 and T_2 refer to evaluation criteria (percentage effect) used to distinguish harmless and harmful outcomes, respectively; T_3 is the field boundary value separating harmless from harmful outcomes. Efficiency is the maximum proportion of products that can be sorted out after the lab trial. For this analysis only data on fungicides (F) and insecticides (I) were used. Note that standard criteria for the evaluation of field experiments consider field effects $>75\%$ harmful and effects $<25\%$ harmless.

^a At a type B error rate of 6%.

Potential Ecological Causes for Type A Errors

Although the argument that non- or low toxicity demonstrated under worst case conditions will not translate into more adverse effects in field trials (type A errors) may seem plausible, there are several ecological reasons why this is not necessarily so. For example, slight effects at the individual level may have profound population dynamic consequences. At the population level, Jansen and Sabelis (1992, 1994) demon-

strated tritrophic systems may attain two steady states (bistability). Perturbations in the density of the predator can result in the system shifting from a plant–herbivore–predator steady state to a plant–herbivore limit cycle. At the metapopulation level, regional persistence of predator–prey populations depends on an adequate balance between dispersal rates both of predator and prey (Nachman, 1988). Hence, when a product enhances predator dispersal rates (repellency), this balance can be disturbed and can lead to the disappearance of the predators, even when no direct toxic effects are found. Pesticides have also been found to disrupt spatial coincidence between predatory mites and their prey (Walker and Penman, 1978, in Croft, 1990).

Indirect effects occur when predators are affected through their food supply. For example, pesticides affecting rust mite populations had a severe impact on predatory mites, although they were not toxic to the latter (Croft and McGroarty, 1977, in Croft, 1990). There are indications that such indirect effects underlie at least some of the interspecific differences in type A error rates reported here. For example, whereas with the orchard/vineyard-inhabiting species type A errors were mostly caused by fungicides, in the case of the greenhouse biocontrol agent, *P. persimilis*, unacceptable type A error rates (40%) were caused by insecticides, while fungicides rarely led to erroneous conclusions. A major difference between these predators is their diet breadth. Certainly in the greenhouse, the diet of *P. persimilis* consists almost exclusively of two-spotted spider mites. The other species can feed on a wide range of additional food sources, including rust mites and pollen (Dicke *et al.*, 1988) and presumably spores of powdery mildew. For this reason, fungicides, unless having marked acaricidal properties, are not likely to affect *P. persimilis*, but their indirect effect on the other species may be crucial (Blommers, 1994). Similarly, changes in spider mite abundance caused by insecticides are more likely to be buffered for the outdoor species than for the greenhouse predator.

Implications for Sequential Testing Protocols

The objective of sequential testing protocols is to reduce the amount of field trials to the necessary minimum. The present analyses demonstrate that in this respect existing protocols do not perform satisfactory. From a total of 276 comparisons, only 29 lab trials provided conclusive results (products harmless to *T. pyri*). Emphasizing harmlessness rather than harmfulness reduces efficiency for two reasons. First, for the data base analyzed, the majority of products tested were insecticides. Most of these were finally classified as harmful (see, e.g., Table 2). Second, as discussed earlier, there is a wide range of factors that may lead to type A errors, whereas type B errors will only depend on the test dose. Ideally, the test protocol should be designed such that both harmless and harmful side effects are correctly predicted by the lab trial. This approach was taken by Oomen *et al.* (1991) when evaluating the OEPP

guideline for *P. persimilis* (1990). They reported an efficiency of 83%. Their lab results were compared with extensive practical experience, but because no actual field data were provided, accuracy could not be determined.

When developing test protocols, it should be recognized that pesticides may affect beneficials also when no lethal effects are demonstrated. Especially when the lab trial is used to demonstrate that a pesticide will have no negative side effects, it may be too limited to focus on direct toxicity alone. The risk of making a type A error depends on ecological attributes of the species under study. For example, incorrect assignment of harmlessness is less likely to occur with host-specific parasitoids than with polyphagous predators. For this reason sequential testing schemes and the decision rules to interpret them should depend on specific attributes of the species under study. This may lead to a situation where, depending on the species, different decision rules are required for different product classes (e.g., insecticides, fungicides). This was demonstrated by the analysis of the data for *P. persimilis*. In any case, there is a clear need to validate sequential testing protocols and their interpretation rules with actual field data. The case of *P. persimilis* illustrates how dangerous rigorous application of fixed decision rules can be. An error rate of 40% should not be taken too lightly! At present, predatory mites offer the only possibility for such a validation. In contrast with, e.g., *Trichogramma*, field experiments with phytoseiids were performed regardless of the outcome of the lab test.

The finding that harmful side effects can be accurately predicted from lab trials (cf. Figs. 2B, 3B, and 4B) leads to the suggestion that the worst case approach should not be the only one in pesticide side-effect testing. If the lab trial is designed such that harmful side effects can be forecasted with a type B error risk lower than 5%, the majority of products does not have to be retested in the field. Hence there will be a gain in efficiency.

Forecasting harmful side effects with a lab test implies testing at rates lower than expected field rates. However, the exact choice of dose a priori is not clear. The further the dose is from the expected field rate, the higher the chance of correctly identifying products. However, this is at the expense of efficiency because the number of products thus found will also decrease. Hence, retesting rates will increase. Dose–response studies could pave the way to set discriminatory doses. This has the additional advantage that lab results are made more universally applicable to different field situations (different crops and phenological states, spraying equipment, etc.).

The multitude of ecological factors potentially causing type A errors implies that they cannot be excluded on basis of lab results. Therefore, in the present scenario, the field trial should be used to demonstrate that a product will not be harmful. It is felt that making the field trial a compulsory requirement for demonstrating that a product will not have negative side effects is a more sound way to reduce adverse effects of pesticides on the environment.

CONCLUSIONS

Existing sequential test protocols used to assess side effects of pesticides on phytoseiid mites do not perform satisfactorily. Only 30% of the products tested on *T. pyri* can be sorted out in the lab with a risk of erroneous conclusions lower than 5%. For *P. persimilis* harmful insecticides and harmless fungicides can also be unambiguously identified using current test methods. No combination of trigger values can be derived for sorting out products harmful for *T. pyri*, and insecticides harmless or fungicides harmful for *P. persimilis*. For the remaining cases combinations of lab and field thresholds do exist, but whether the latter are realistic is not clear.

ACKNOWLEDGMENTS

The authors are indebted to Sylvia Blümel and the Mitox Company—University of Amsterdam for their kind permission to use their unpublished results. They also thank Carlo Duso for a draft manuscript of his data. The scrutiny of both Maurice Sabelis and Pieter Oomen tremendously improved the manuscript. J.A.J. was supported through a grant obtained from the Spanish Ministry of Education and Science (Grant PF93 35072960).

REFERENCES

- Aldridge, C., and Carter, N. (1992). The principles of risk assessment for non-target arthropods: A UK registration perspective. *Aspects of Applied Biology: Interpretation of Pesticide Effects on Beneficial Arthropods*, Vol. 31, pp. 149–156. The Association of Applied Biologists, Horticulture Research International, Wellesbourne, Warwick, UK.
- Anber, H. A. I., and Overmeer, W. P. J. (1988). Resistance to organophosphates and carbamates in the predacious mite *Amblyseius potentillae* (Garnan) due to insensitive acetylcholinesterase. *Pestic. Biochem. Physiol.* **31**, 91–98.
- Bakker, F. M., Grove, A. J., Blümel, S., Calis, J. N. M., and Oomen, P. (1992). Side-effect tests for phytoseiids and their rearing methods. *Bull. IOBC/WPRS* **15**(3), 61–81.
- Barrett, K. L. (1992). The BART group approach to regulatory testing. *Aspects of Applied Biology: Interpretation of Pesticide Effects on Beneficial Arthropods*, Vol. 31, pp. 165–170. The Association of Applied Biologists, Horticulture Research International, Wellesbourne, Warwick, UK.
- Birch, L. C. (1948). The intrinsic rate of natural increase of an insect population. *J. Anim. Ecol.* **17**, 15–26.
- Blommers, L. H. M. (1994). Integrated pest management in European apple orchards. *Annu. Rev. Entomol.* **39**, 213–241.
- Blümel, S. (1991). Results of the 7th Joint Testing Programme IOBC Working Group "Pesticides and Beneficials": *Phytoseiulus persimilis*. Unpublished results presented at the 14th meeting of the working group, Southampton, UK.
- Blümel, S., Bakker, F., and Grove, A. (1993). Evaluation of different methods to assess the side-effects of pesticides on *Phytoseiulus persimilis* A-H. *Exp. Appl. Acarol.* **17**, 161–169.
- Calis, J. N. M. (1987). Results of the 6th Joint Testing Programme of the IOBC Working group—*Amblyseius andersoni*. Mitox unpublished internal report.
- Carter, N., Oomen, P. A., Inglesfield, C., Samsøe-Petersen, L., De Clerq, R., and Dedryver, C. A. (1992). A European guideline for testing the effects of plant protection products on arthropod natural enemies. *Aspects of Applied Biology: Interpretation of Pesticide Effects on Beneficial Arthropods*, Vol. 31, pp. 157–163. The Association of Applied Biologists, Horticulture Research International, Wellesbourne, Warwick, UK.
- Croft, B. (1990). *Biological Control Agents and Pesticides*. Wiley, New York.
- Dicke, M., Sabelis, M. W., and De Jong, M. (1988). Analysis of prey preference in phytoseiid mites by using an olfactometer, predation models and electrophoresis. *Exp. Appl. Acarol.* **5**, 225–243.
- Duso, C. (1994). Comparison between field and laboratory testing methods to evaluate the pesticide side-effects on the predatory mites *Amblyseius andersoni* and *Typhlodromus pyri*. *Bull. IOBC/WPRS* **17**(10), 7–19.
- Grove, A. J. (1993). Results of the 6th and 7th Joint Testing Programmes of the IOBC Working Group—*Typhlodromus pyri* and *Phytoseiulus persimilis*. Mitox unpublished internal report.
- Hassan, S. A. (1989). Testing the methodology and the concept of IOBC/WPRS Working Group. In *Pesticides and Non-target Invertebrates* (P. C. Jepson, Ed.), pp. 1–18. Intercept Ltd., Wimborne, Dorset.
- Hassan, S. A. (Ed.) (1992a). Guidelines for testing the effects of pesticides on beneficial organisms: Description of test methods. *Bull. IOBC/WPRS* **15**(3), 186.
- Hassan, S. A. (1992b). Guideline for the evaluation of side-effects of plant protection products on *Trichogramma cacoeciae*. *Bull. IOBC/WPRS* **15**(3), 18–39.
- Hassan, S. A., Bigler, F., Bogenschütz, H., Brown, J. U., Firth, S. I., Huang, P., Ledieu, M. S., Naton, E., Oomen, P. A., Overmeer, W. P. J., Rieckmann, W., Samsøe-Petersen, L., Viggiani, G., and van Zon, A. Q. (1983). Results of the second joint testing programme by the IOBC/WPRS Working Group pesticides and beneficial arthropods. *Z. Angew. Entomol.* **95**, 151–158.
- Hassan, S. A., Bigler, F., Blaisinger, P., Bogenschütz, H., Brun, J., Chiverton, P., Dickler, E., Easterbrook, M. E., Edwards, P. J., Englert, W. D., Firth, S. I., Huang, P., Inglesfield, C., Klingauf, F., Kuehner, C., Ledieu, M. S., Naton, E., Oomen, P. A., Overmeer, W. P. J., Plevoets, P., Reboulet, J. N., Rieckmann, W., Samsøe-Petersen, L., Shires, S. W., Staubli, A., Stevenson, J., Tuset, J. J., Vanwetswinkel, G., and van Zon, A. Q. (1985). Standard methods to test the side-effects of pesticides on natural enemies of insects and mites developed by the IOBC/WPRS Working Group pesticides and beneficial organisms. *Bull. OEPP/EPPO Bull.* **15**, 214–255.
- Hassan, S. A., Albert, R., Bigler, F., Blaisinger, F., Bogenschütz, H., Boller, E., Brun, J., Chiverton, P. A., Edwards, P., Englert, W. D., Huang, P., Inglesfield, P., Naton, E., Oomen, P. A., Overmeer, W. P. J., Rieckmann, W., Samsøe-Petersen, L., Staubli, A., Tuset, J. J., Viggiani, G., and Vanwetswinkel, G. (1987). Results of the third joint testing programme by the IOBC/WPRS Working Group pesticides and beneficial organisms. *Z. Angew. Entomol.* **103**, 92–107.
- Hassan, S. A., Bigler, F., Bogenschütz, H., Boller, E., Brun, J., Chiverton, P. A., Edwards, P., Mansour, F., Naton, E., Oomen, P. A., Overmeer, W. P. J., Polgar, L., Rieckmann, W., Samsøe-Petersen, L., Staubli, A., Sterk, G., Tavares, K., Tuset, J. J., Viggiani, G., and Vivas, A. G. (1988). Results of the fourth joint testing programme carried out by the IOBC/WPRS Working Group pesticides and beneficial organisms. *J. Appl. Entomol.* **105**, 321–329.
- Hassan, S. A., Bigler, F., Bogenschütz, H., Boller, E., Brun, J., Calis, J. N. M., Chiverton, P. A., Coremans-Pelsener, J., Duso, C., Lewis, G. B., Mansour, F., Moreth, L., Oomen, P. A., Overmeer, W. P. J., Polgar, L., Rieckmann, W., Samsøe-Petersen, L., Staubli, A., Sterk, G., Tavares, K., Tuset, J. J., and Viggiani, G. (1991). Results of the fifth joint testing programme carried out by the IOBC/WPRS Working Group pesticides and beneficial organisms. *Entomophaga* **36**, 55–67.
- Hassan, S. A., Bigler, F., Bogenschütz, H., Boller, E., Brun, J., Calis, J. N. M., Duso, C., Grove, A. J., Helyer, N., Heimbach, U., Hokkanen, H., Lewis, G. B., Mansour, F., Moreth, L., W. P. J., Polgar, W., Coremans-Pelsener, J., Samsøe-Petersen, L., Sauphanor, B., Staubli, A., Sterk, G., Tavares, K., Tuset, J. J., and Viggiani, G. (1994). Results of the sixth joint pesticide testing programme carried out by the IOBC/WPRS Working Group Pesticides and beneficial organisms. Manuscript in preparation.
- IOBC (1988). Working Group pesticides and beneficial organisms. *IOBC/WPRS Bull.* **11**(4).
- Jansen, V. A. A., and Sabelis, M. W. (1992). Prey dispersal and predator persistence. *Exp. Appl. Acarol.* **14**, 215–231.

- Jansen, V. A. A., and Sabelis, M. W. (1994). Pest outbreaks in tri-trophic systems: Population dynamical mechanisms and consequences for the evolution of pesticide resistance. In *Theoretical Aspects of Metapopulation Dynamics*, pp. 67–78. Ph.D. Thesis, Univ. of Leiden, The Netherlands.
- Matthews, G. A. (1991). *Pesticide Application Methods*. Longman Group Ltd., New York.
- Nachman, G. (1988). Regional persistence of locally unstable predator/prey populations. *Exp. Appl. Acarol.* **5**, 293–319.
- OEPP/EPPO (1989). Guideline for the evaluation of side-effects of plant protection products, No. 142: *Encarsia formosa*. *Bull. OEPP/EPPO Bull.* **19**, 355–372.
- OEPP/EPPO (1990). Guideline for the evaluation of side-effects of plant protection products, No. 151: *Phytoseiulus persimilis*. *Bull. OEPP/EPPO Bull.* **19**, 355–372.
- OEPP/EPPO (1994). Decision making scheme for the environmental risk assessment of plant protection products: Arthropod natural enemies. *Bull. OEPP/EPPO Bull.* **24**, 17–36.
- Oomen, P. A., Romeijn, G., and Weigers, G. L. (1991). Side-effects of 100 pesticides on the predatory mite *Phytoseiulus persimilis*, collected and evaluated according to the EPPO Guideline. *Bull. OEPP/EPPO Bull.* **21**, 701–712.
- Overmeer, W. P. J., and van Zon, A. Q. (1981). Results of the 3rd Joint Testing Programmes of the IOBC Working Group—*Amblyseius andersoni* and *Typhlodromus pyri*. Univ. of Amsterdam, unpublished internal report.
- Overmeer, W. P. J., and van Zon, A. Q. (1982). A standardized method for testing the side-effects of pesticides on the predacious mite *Amblyseius potentillae* (Acarina: Phytoseiidae). *Entomophaga* **27**, 357–364.
- Overmeer, W. P. J., and van Zon, A. Q. (1983). Results of the 4th Joint Testing Programmes of the IOBC Working Group—*Amblyseius andersoni* and *Typhlodromus pyri*. Univ. of Amsterdam, unpublished internal report.
- Samsøe-Petersen, L. (1990). Sequences of standard methods to test effects of chemicals on terrestrial arthropods. *Ecotoxicol. Environ. Saf.* **19**, 310–319.
- Van de Baan, H. E., Kuypers, L. A. M., Overmeer, W. P. J., and Oppenoorth, F. J. (1985). Organophosphorous and carbamate resistance in the predacious mite *Typhlodromus pyri* due to insensitive acetylcholinesterase. *Exp. Appl. Acarol.* **1**, 3–10.
- Wilkinson, L. (1987). *SYSTAT: The System for Statistics*. Systat, Inc., Evanston, IL.
- Wilkinson, L. (1988). *SYSGRAPH: The System for Graphics for the Macintosh*. Systat, Inc., Evanston, IL.