DISCUSSION PANEL: FIELD TEST DESIGN & DATA ANALYSIS

Panel Members: Daniel M. Schmitt, Chairman; William E. Waters; B. Leo Cadogan; Gerald S. Walton

Panel Chairman:

I think it is time to open the panel. On my left respectively are Bill Waters, Professor of Entomology and Forestry, University of California, Berkeley; Gerald S. Walton, Biometrician, Northeastern Forest Experiment Station, Hamden; and Leo Cadogan, Forest Pest Management Institute, Canadian Forestry Service, Sault Ste. Marie, Ontario. My name is Dan Schmitt, Program Manager, Northeastern Forest Experiment Station, U.S.F.S., Broomall, Pa.

I am going to ask each of the panelists to discuss field test design and data analysis for pesticide applications in the context of three general areas: (1) field and pilot tests, (2) current approaches to experimental tests, and (3) suggestions and possibilities for improving experimental tests. Bill?

Bill Waters:

There seems to be a general understanding as to the role of field testing, but not as to its purpose. The purpose should be to find the most efficient combination of treatment variables, i.e. formulation, concentration, application rate, atomization, and timing, that will consistently provide a specified degree of control of the target insect and/or foliage protection under operational conditions. The question is not which treatment or treatment combinations perform best, but which, if any, will achieve the desired results. Comparisons are of course part of the process of evaluation, but they are not the purpose of field testing itself.

In order to achieve the purpose of field testing we need to have performance criteria, which should be stated in advance. These criteria are the explicit bases for evaluation. The best way to state criteria is in terms of what remains following treatment, not in how much a population is reduced nor how much defoliation has been reduced. For the insect this means number of survivors, not percentage killed; for foliage, the amount remaining not the amount lost. If we are truly managing pests, we must to be able to make decisions on alternative treatments in terms of probable outcomes over a time span meaningful to the Forest Manager. The models that are being developed for this purpose therefore deal with insect survival densities and amounts of foliage produced over time with and without treatments.

Once we are clear about the purpose of field testing, we can consider the distinction between field experiments and pilot or operational tests. Both come under the umbrella of field testing.

Field experiments simply involve experimental combinations of the treatment variables applied under field conditions, with the understanding that the insect is present at measurable densities. Practically, this means a range of population levels causing sufficient defoliation to be considered for control. Field experiments also imply that the equipment employed (unless the manner in which it is employed is itself a treatment variable) is representative of conventional operations.

Everyone is eager to go from the laboratory to the field. However a basic rule is that field experiments should be initiated only when the treatment variables have been reduced to a number that can be reasonably encompassed in an efficient experimental design. Generally, these variables are best combined in factorial designs, but a relatively large number of units are required for even a simple complete experiment. For example, a single replication of a factorial experiment with two levels each of four variables requires 32 plots. In many cases, incomplete designs are necessary. To acquire all of the information needed a programmed sequence of experiments must be planned.

Pilot tests are operational field tests of the treatment found to be optimal by field experiments. The purpose is to determine how well the treatment performs under fully operational conditions. When field experiments fail to distinguish an optimum treatment, there may be justification in subjecting several to a pilot scale operation. In such cases, special care should be taken to acquire the information necessary for objective evaluation, with adequate data taken to assure that comparisons can be made in terms of specified performance criteria. One need only consider the tremendous cost of pilot scale tests to appreciate the importance of sound progression through all developmental steps to the pilot scale test, and the need for adequate data from the test itself.

Currently we seem to design, often simplistically, both field experiments and pilot tests and then compromise them because of 'realities'. I have had much experience with such negotiations and compromise and, in my opinion, they have in many cases killed us, i.e. time and again we have obtained inconclusive results. It behooves all of us to start with basic concepts, objectives, appropriate experimental designs, and adequate sampling, and stick to them.

A related problem affecting the current difficulties in interpreting the results of field tests with B.t. is differences in expectations regarding outcomes--expectations of what constitutes satisfactory performance with regard to suppression of the insect and foliage protection-- and, in fact, of opinion as to differences in expectation of which objective is the most important, or whether there is an optimal balance. If there is not agreement on what we expect, how the devil can we design an experiment or pilot test to get specific information to evaluate performance? Because we don't have a clear understanding on what constitutes satisfactory performance nor when it is optimal to move from the laboratory to the field, we have a continued history of what I would call premature field experiments.

In listening to the previous speakers, I was struck by the fact that people essentially referring to the same range of data, were supplying quite different interpretations. Obviously, there can be little developmental progress if the same data mean different things to different people.

For the spruce budworm there is a scientific basis from population studies for specifying minimal performance or even better, what the outcome will be. There is a basis in spruce budworm research for saying whether 5% survival is or is not satisfactory, for example. Better, it is possible to state, given x number of insects on a 45 cm branch after a control operation, what the number is likely to be the next season. It is the neglect of population information, and unrealistic faith in efficacy couched in percent population reduction, percent foliage reduction, etc. that has caused public agencies to lose credibility and, in the case of the US Forest Service, to find itself in court. Managers simply do not understand, after being told that an operation on their land resulted in a certain percent population reduction or foliage protection classified as successful -- why, the following season they must spray again.

The only other comment I want to make on future improvement applies to sampling. Our basis for evaluating field tests is the data we gather. The small data sets usually planned for, and subsequent compromises, have left us with a very small data base on which to make evaluations. Our current procedures in this respect are not only scientifically unsound, but irresponsible in view of the terrible price we have paid in credibility. I will make one suggestion regarding population sampling. Since the objective is to determine whether or not the surviving population is at a level specified for a satisfactory control operation, sequential sampling is the most efficient procedure.

Chairman:

Thank you, Bill. Jerry?

Jerry Walton:

I think the main concern in a field experiment is that the objectives be very clear and well defined. Certainly it would not be useful to restrict research objectives nor should the same data on populations, foliation, etc. be collected for every objective. I see the field experiment as a method of answering very specific questions and not as an occasion to collect miscellaneous information.

Nor do I believe that field experiments should be necessarily duplicate operations, e.g. with respect to gallonage and dosage. It may be the purpose of the experiment to establish an optimum, in which case it is efficient to determine the end points for the parameter in question. Whether or not these end points are realistic from an operational standpoint is irrelevant.

Once specific questions, i.e. objectives, have been formulated the place to start is with an adequate design. Constraints on time, space, and funding may make compromises unavoidable. But one must constantly ask if a compromise is really necessary. The availability of suitable test areas is often a problem-- a limiting problem. In such situations some adjustments in ideal design may be necessary. One possibility is to use smaller spray blocks. But if the blocks are too small, the results will not be transferable to operating conditions. So there has to be some happy medium. Knowing costs and other constraints, one can proceed toward the optimal design.

Having mentioned costs, it should be clearly understood that information costs money; you are going to pay for what you get. Also, if you don't get the information, you lose everything. But we often do things that are costly and do not provide useful information. For example, placing deposit collectors next to trees with the idea that the droplet record on the collector can be correlated, by tree, with population reduction or foliage saved. The difficulty here is that population mortality on a single tree is so largely unknown that you cannot establish a correlation that is practically significant. And yet a great deal has been spent on this sort of thing.

There is a tendency also to expand objectives, a feeling that collecting additional information for another objective is a more useful way to allocate resources than intensifying the sampling to collect data relevant to the specific objectives of the field experiment. The result is a waste of sampling effort. I think the Core test[1] is a prime example. You go out with very marginal sampling efforts and many sub-objectives on the assumption that things are going to go right -- and they never do. Any time anybody says, 'well, as long as we are out there we may as well measure ...', don't. It is a waste of time better spent on the collection of relevant data.

Conforming to conventional approaches to field experiments is another insidious means of wasting resources. For example, we tend to use the same post-spray sampling time sequence, irrespective of what insecticide, microbial or chemical, is used, or what species the target insect is. I think a lot of our problems stem from traditional approaches with little thought to the problem and experimental conditions at hand.

There is another kind of mind set that affects data analyses and interpretation of field tests. I call it the DDT syndrome, the feeling that at least one of the treatments must produce results equivalent to what DDT would have. Saying it, or on paper, this is nonsense. We don't do research this way. But why then is there wide spread resistance to adjusting post-spray population data for natural mortality? It is also traditional to consider a field experiment a success if, using sophisticated statistical testing methods, you can detect whether or not you sprayed anything. That is all "statistical significance" means, and it is hardly a worthy criterion for success.

Do we need to measure foliage protection? I don't think that is the way to ask the question. Better, given the objectives of the experiment, does foliage protection tell us anything? If, for example, the size of the surviving population after treatment compared to pre-treatment size is the issue, foliage information is not relevant. If the rate of mortality is the issue (i.e. how soon the insects die), then the level of foliage protection is an indication of how effective the treatment was. In brief we should clearly think through the purpose of a field experiment, frame the questions to be answered specifically, design efficiently, and collect data, and enough data, to answer the questions, and collect no other data.

Before discussing improvements, I'd like to comment on performance as a general objective of field experiments. My general attitude towards a field test is that it is to derive information. It is my feeling that trying to reach goals that will make the user happy has interfered with doing just that. Interim specfic objectives should be interposed as part of a planned program of testing leading to optimal performance, rather than performance as an objective per se.

Another point, which struck me especially at this meeting, is that operations people normally make incremental adjustments in spray and insecticide parameters to improve efficacy. They may well be ahead of research in appraising the effects of such changes, though without the analysis. So maybe the area in which research can help most would be evaluation of the radical and innovative things that people in the field can't justify.

---

[1]Refers to a microbial field test reported in "Spruce budworm core B.t. test -- 1980: combined summary. USDA Forest Service RES. PA NE-506." The chairman of this panel suggested the general approach and imposed funding contraints. Experimental design and sampling methods were the product of an ad hoc committee.

I do not like the current use of controls -- check plots. There is something suspect about their selection, about the way control populations behave. Perhaps it would be better to establish a standard treatment as a basis for evaluating experimental ones.

Again, depending on the objectives, you may not need controls. If the objective is to compare two treatments, the control in the traditional concept does not provide relevant information. Finally, although new statistical methods and approaches to experimental design and data analysis are coming on line all the time, there is nothing in statistical theory per se that is going to make the entomologist particularly confortable in his field experimental work.

Chairman:

Thank you Jerry. Leo.

Leo Cadagon:

There are misunderstandings about field tests, especially field tests involving the use of microbials. Microbials must satisfy biological or natural laws; however when encompassed as part of a spray, their application is governed by certain physical laws. Predicting the behavior of biological entities is very much more hazardous, because of their biological properties, than of abiotic entities which conform to chemistry and physics.

The Canadian Forestry Service has an experimental protocol intended to lead to the registration of the pesticide and all that that implies. The protocol starts in the laboratory. Laboratory investigations treat toxicology, LD-50 assays, half life of the material, formulation tests, analyses of stability, determination of physical and chemical parameters affecting delivery as a spray, etc. In general laboratory investigations are intended to identify problems at an early stage and resolve them. There is no reason, for example, to go to the field and find out 'it didn't stick'.

The next step is single tree application experiments. Here information is gathered on deposit characteristics, more information on stability under more realistic conditions. We then proceed to small scale trials. These are thorough experimental trials. The intention is not just to evaluate potency but to acquire information leading to improvements in the material or its application. We then proceed to semi-operational trials on about 300-400 hectares. These are often laid out or supervised by forest or forest pest managers but data collection and the rigor of analysis approximate the level of small scale field tests.

If all goes well we will have identified an insecticide, its appropriate formulation for the delivery systems which are likely to be used, and performance information over a range of probable conditions. We are now ready to conduct large plot operational trials.

I mentioned earlier that microbials, particularly B.t., have unique properties some of which are very desirable. Consequently there is a temptation to go into the field early to see how it works, bypassing much of the protocol described. If the product fails, one can argue that you should not have gone to the field. But a field season has been wasted and often field tests provide data which suggest that a single factor caused the failure and the manager can fix it. So the product does not go back to the laboratory; the manager "fixes" the problem and the product is field tested again the next year. In this way an error-prone, insensitive, cyclical adjustment process is introduced. In fact microbials are one of the few products where, years after registration, exploratory research that should probably have been conducted prior to registration is still being conducted.

Regarding current methodology I presume Dan (the chairman) was asking for opinions on the B.t. Core tests. Now, conducting research in a natural forest is a difficult task. Conducting research with microbials is doubly difficult. I presume the core tests were initiated to deal with inconsistencies. Although the concept (multi-location field tests, each with the same experimental design) is a good one, I am not sure it can deal with inconsistencies in microbials.

Classical agricultural designs cannot always be adapted to natural forests, although they may be suitable for plantation R&D. Designs for forest research must consider the many constraints that are evident in natural forests. One of the most restrictive for formulation and application R&D is the availability of adequate sites -- those with suitable insect populations, accessibility, and which are conducive to safe aircraft operations. The design must take into account logistical requirements, be realistic, and not carved in stone.

However the original design should still be optimal. If absolutely necessary, compromise! For example I do not think you can find enough areas in Ontario to conduct a truly randomized design involving more than two variables. Last, but not least, you have to recognize the "Chancellor of the Exchequer". We can always design the experiments we like, but the money might not be there.

Finally, after listening to the discussions of the past two days, I feel that perceived difficulties with products, formulations, and delivery systems, might be really design problems. It should be recognized that design includes specifications for implementation, including sampling. We need to honestly look at the procedures we employ in relation to design.

Regarding improvements in current methods, I think first of all we have to tackle attitude. Researchers in microbials are perceived to be promoting, not testing, microbials.

As mentioned there is room for improvement in design. Whatever the outcome of the core test, to the extent it was a cooperative effort, it was a start in the right direction. I think all the actors need to get together very early and settle design problems. Positive controls in the sense of a standard treatment are worth looking at, but they add expense. Actually I have no problem with check plots provided they are representative of the area and the population. Sampling is a problem, but it is one we can deal with.

Analyses and interpretation of microbial field tests are difficult. Anyone who has seen Ed Kettela's presentation ("Review of problems with efficacy of B.T.K. in spruce budworm control operations in the Maritime Provinces of Canada" in this Proceedings) will appreciate what I am talking about. For example if A, with 98% defoliation is significantly different than B, with 96%, practically what does this mean? Probably nothing. We have an obligation to carefully and objectively interpret data and not to take refuge behind a barricade of statistics and meaningless statements of "significance". Thank you.

Chairman:

Alright, we will now entertain questions and comments. Wladimir?

Dr. Wladimir Smirnoff. Dr. Waters commented that field tests should be representative of operational conditions. But if the product or formulation provided for experimentation is under or over strength interpretation of the results is going to be very very difficult.

I'd also like to comment on Leo's remarks on registration. Because a product is registered people spray big territories with it without paying much attention to the label and also there can be a problem of the type I just mentioned. Sometimes believing in the registration without regard to the label or professional research advice can lead to big mistakes.

People also don't appreciate true costs compared to out of pocket costs. For example in the name of economy some applicators prefer to increase the dosage, and volume rather than perform calibration tests, and so artificially increase the cost by 50%. Others try products which worked on 6th instar larvae which is ridiculous since defoliation is already completed.

I agree with Dr. Waters that we must have standard criteria for evaluating pesticide applications. Currently criteria fluctuate from operation to operation, and it is impossible to fairly evaluate performance.

Next, combining B.t. with some chemical is unacceptable since it has never been studied.

The dye Erioacid red, for example, has been added at the rate of two kilo/100 gal. regardless of the effect on the formulation. No further comment.

Chairman:

Thank you, Wladimir. I believe you were endorsing some of the panels' comments. I think the panel tangentially addressed the matter of information expected from field vs pilot tests. Does anyone from the audience care to comment. Jim Bean?

James Bean:

Listening to the discussion, I am sure Bill (Waters) went back to probably 1962 because he is saying some of the things he said then. And it is just as apropos today. There is a definite need to start at the beginning and get good basic information. I agree we have often gone to field tests too soon.

John Cunningham:

I never really thought of two particular categories of tests. Viruses are rather a peculiar type of biological insecticide and it is probably necessary to take them to the field at a fairly early stage to determine if you will get a virus epizootic. In other words you want to see if there is a cummulative effect.

You can get great results in the laboratory, but in the field there may be a success or a total failure. The forest includes a range of factors that cannot be evaluated in the laboratory. John Podgewaite may want to comment on this as he is working with Gypsy Moth virus. So I feel for potential virus candidates, the only thing to do is to take them out and try them in the field at a fairly early stage. The alternative is to use viruses as B.t. -- in which case they become fairly expensive stomach poisons. We really can't go this route with viruses.

Another factor, not mentioned so far, affecting field tests is the health of the target population. Now older populations of budworm become heavily infected with microsporidia. There is extremely little work done on the differences in population quality and how such differences affect insecticide applications, although Wladimir has done some research on this subject in relation to B.t.

I conducted some small laboratory tests and find high mortality following NPV infection, if the larvae are previously infected with microsporidia or B.t. On the negative side virus yield from such insects is very much reduced.

Another factor that crops up in the field is fungus epizootics. They may affect your check plots or treated plots. The only way to check this possibility is to examine your population samples microscopically in the laboratory.

Chairman:

Thank you John. You did raise an issue -- population quality -- that needs to be more thoroughly explored. And, of course, there is another, host quality. The panelists and the comments from the audience have stressed the inadequacy of information collected. And here I will raise a question. It is customary to collect information on current defoliation. In the case of microbials, in a research sense, is this the information we want? Anybody care to comment?

Paul Fast:

I will. If you plot some population measure, larvae per bud, or buds per larvae, against defoliation, you will get the following kind of relation (illustrates). The point is you can measure the variance about the regression or correlation and get a darn good idea of what effect your measured population is having on defoliation. We find that larvae/bud instead of larvae per branch provides a narrower variance.

Now once your experiment is finished you can plot the location of your test plot on the graph. You can then see by inspection whether it falls within the confidence limits of the graph and determine whether you got the desired effect.

This is more expensive because it means doing three or four or five checks at different population levels in order to get the curve, but it is much higher resolution than using one or two or several check plots to compare differences in a number of parameters between treated and check plots. In brief the resolution of the evaluation technique we currently use is not good enough to detect the kinds of differences we are interested in.

Chairman:

Thank you Paul. Wladimir, you wanted to add something?

Wladimir Smirnoff:

Yes. Based on my figures of foliage protection obtained over many years in permanent plots with careful measurement of larvae/bud, larvae/branch, dead larvae, new buds, etc. before and after treatment, I can say that it is possible to project foliage protection in terms of tons/hectare/year.

Chairman:

Thank you Wladimir. Now I'd like to add a thought regarding lab tests, field test, pilot tests. The idea, which has been roundly trashed everytime I mentioned it, was that the lab tests should provide all the information on toxicology, stability, resistance to water, etc. and the purpose of the field test then was to essentially raise the question of why the field test did not come out like the laboratory test.

It sounds like a fairly ridiculous idea, but it occurred to me that if such a research protocol had been in place, the need for the kind of research that Charles Wiesner, Paul Fast, and others are doing now, would have become evident much, much sooner.

Dr. Waters:

I don't think that is a legitimate or relevant question. Admittedly, laboratory research is the foundation for the field experiments, but the questions to be raised and the hypotheses to be tested in the field experiments are of a different sort, having to do with operating characteristics of the material. The information base for the two kinds of research is not the same. The lab tests lead to the field tests, but one doesn't test the results of one against the other.

Chairman:

I won't pursue this matter any more, but will turn to a subject first mentioned by Dr. Waters -- the need for programmed experiments. He pointed out that experiments could be conducted over time. That is a missing ingredient in field testing -- a total design in which time is one of the variables and in which one set of factorials would be put in one year and the total experiment might require three years.

Imants Millers:

The problem is that, in B.t. in particular, the product changes faster than you conduct the research. For example of what value were the core[1] tests? Current products in use are completely different than the tested ones.

Oswald Morris:

I have heard a lot of rather wild statements this afternoon and so will make one of my own. First, I do not think there is any such thing as a failure or a negative result in research. Research is a continuum. Negative results are merely steps in a positive direction.

Carol S. Glenister:

I am a private contract researcher working with Gypsy Moth. People concentrate on evaluating levels of defoliation when they do not know what the volume of leaf tissue is in the field. Assume you have a population, one population, and it will consume 100 kg. of leaves. If the volume available is different in the plots -- and you don't know it -- the percent defoliation will be different. Add different treatments and different populations on the plots, it is easy to see that the current expression of the defoliation is meaningless.

Dr. Waters:

This is just one of the problems in expressing results in terms of defoliation. There are other problems with using defoliation as a measure of treatment effectiveness. The use of check plots for this purpose is particularly questionable. These problems can be avoided by getting sample counts of known precision and using them to relate to populations in treated plots.

Carol Glenister:

In the lab you have known mortalities and known population. In the field you essentially don't know mortalities and you think you can go from percent defoliation to residual population. But you can't unless you know what the volume is that is being consumed.

Now there are models available in which foliage weight can be estimated from tree diameter. You can still use visual percent defoliation estimates but you would have to convert it back to volume consumed in the stand. You can then better estimate what happened following treatment to your population and then compare percent mortality in the field with lab mortality.

John Cunningham:

One thing we need to remember in expressing results, like "foliage protected," is the client. In gypsy moth the client is usually the home owner; in budworm, the forest manager, etc. Each insect affects a different client, so you need to express results that are meaningful in the context of the client's problem.

Leo Cadogan:

Yes, but the researcher's clients also differ at the experimental stage and the field stage. In research I think our clients are our colleagues or peers, and it is very difficult to say to them, "it didn't work" or, "it was fairly good." That is fine for the homeowner, but not for persons who expect a more definitive statement of results.

Dan Kucera:

Getting back to the availibility of virus suitable for field and pilot testing, cooperators who provide such areas are increasingly concerned about environmental impacts. I think in the future we are going to have to be very careful in telling the land manager what we plan to do, how we are going to do it, what we hope for in our results; and we must assure the land manager that we are going to take all the precautions we can. Without this we are going to be very restricted in what we can actually test.

(At this point the panel ended)