Proceedings of the Eighth International Symposium on Biological Control of Weeds
2-7 February 1992, Lincoln University, Canterbury, New Zealand.
Delfosse, E.S. and R.R. Scott (eds). DSIR/CSIRO, Melbourne, pp. 213-9 (1995).

Beyond "Before-and-After:" Experimental Design and Evaluation in Classical Weed Biological Control

A.S. McClay

Alberta Environmental Centre, Bag 4000, Vegreville, Alberta T9C IT4, Canada

To evaluate the effectiveness of weed biological control agents, it is necessary to determine their effects on weed populations in the field. The experimental approach, whose key elements are replication of experimental units and randomization of treatments, is the best way to separate these effects unambiguously from other sources of variation. I surveyed 57 published evaluations of classical weed biological control projects. Only 16 studies used experimental methods, and among these only 6 used measures of weed population. Randomization, formal experimental design, and thorough statistical analysis were rarely applied. Possible reasons for the infrequent use of experiments include the mobility of many biological control agents, the logistic difficulty and long duration of the required experiments, the shortage of agents in the early stages of projects, and the spectacular nature of some historic biological control successes, which appears to make quantitative evaluation unnecessary. A greater emphasis on sound experimental design is necessary to preserve the credibility of classical weed biological control.

Introduction

Classical biological control is often described as an ecological field experiment on the largest scale (Myers 1978). The introduction of weed biological control agents into new regions provides opportunities to study many intriguing questions in population dynamics, genetics, environmental adaptation, and plant-insect interactions. In our fascination with these questions affecting the *mechanism* of biological control, however, we should not lose sight of the *outcome* of the process. In this paper I will, therefore, focus on a single, specific question, which is central to the business of applied biological control. This is: how do we show that a biological control agent is effective in reducing the severity of a weed problem? I do not consider here the separate question of what degree of control is regarded as "sufficient", but only the methods used in assessing the degree of control. I will consider the elements required for evaluating the effectiveness of a weed biological control project, review the methods described in the published literature, and suggest some guidelines for evaluation of future projects. I will also argue that the experimental method is the preferred approach for evaluating the effectiveness of biological control, and that we need to give more attention to experimental design, replication, and statistical analysis than has been done until now.

Elements Needed in Evaluation of Biological Control

The requirements for evaluation of a weed biological control project have been considered by many authors (e.g., Crawley 1990, Harris 1981, McEvoy *et al.* 1991, McLaren and Amor 1979, Wapshere *et al.* 1989). I consider the most important elements to be:

1. Assess the Weed, Not the Agent

Even the most thriving biological control agent will not control its target weed if it does not cause damage of a type to which the weed population is vulnerable. Post-release studies focused only on the performance of the biological control agent–its establishment, population dynamics, dispersal, etc.–can therefore not answer the question posed above.

2. Assess Weed Populations, Not Individuals

Weeds cause problems because they occur as populations. Any evaluation of weed control must, therefore, be based on some measure of weed population, expressed on a per unit area basis. Many frequently-used measures, such as dimensions or biomass of individual plants, are expressed on a per plant basis and thus carry no information on the weed population. Others, such as percent infestation, are really measures of biological control agent population rather than of weed response. Such measures may be valuable in studies of the mechanism of biological control, but they cannot themselves tell us if biological control has been successful.

3. Assess in the Field

The outcome of classical biological control depends on many environmental factors acting over a prolonged time on both the weed and the biological control agent. These factors may be difficult or impossible to reproduce in greenhouse or artificial field plot conditions. The best test of the ability of an agent to provide control is therefore to study it under natural field conditions.

4. Prove Responsibility of the Agent

The method of evaluation used must be able to ensure that effects are properly attributed to their causes; that is, to separate reductions in weed population caused by the biological control agent from spatial and temporal variation due to factors such as weather, soil conditions, competition, allelopathy, management, and other phytophages and diseases.

The essential problem is, therefore, to take some measure of a weed population in the field and demonstrate, against a background of natural variability, that the introduction of a biological control agent causes a reduction in that measure: in other words, to detect a treatment effect and separate it from other sources of variability. We can approach this problem by either the observational or the experimental route. These approaches differ, not in the technical complexity of their methods or of their statistical analyses, but in the level of intervention by the investigator. In observational studies the choice of sites or plants attacked by the biological control agent is left to the agent itself, or to factors external to the study, while in experimental studies the investigator controls the allocation of attack. These 2 categories were described by Hurlbert (1984) respectively as "mensurative experiments" and "manipulative experiments."

The observational method depends on documenting a correlation in space or time between attack by the biological control agent and a reduction in weed populations. Although this may be fairly convincing evidence in clear-cut cases, this approach can never conclusively separate the effects of biological control agents from those of other environmental factors. If we observe that at sites where the agent is abundant, the weed population is lower, this may be because the agent is controlling the weed. However, it may also be because the agent prefers, for example, dry sites, and the weed grows better in moist sites. Correlations in time are subject to similar problems. For example, it has been observed at a number of sites that release of *Ceutorhynchus litura* (F.) (Coleoptera: Curculionidae) has been followed by a decline in Canada thistle populations (Peschken and Wilkinson 1981, Rees 1990). However, it is not yet clear whether the insect is responsible for these declines. The advantage of the experimental approach is that by allocating biological control treatments to experimental units according to an appropriate design, their effects can be separated from those of all other environmental factors affecting the weed population.

Literature Survey

With these principles in mind, I surveyed 57 published papers describing the approaches used in evaluation of specific weed biological control projects against 33 target weeds in 10 countries. Papers were selected from previous *International Symposia on Biological Control of Weeds*, from a computerized literature search, from my own reprint collection, and by following up citations in other papers. The criteria I used in selecting papers were:

- classical biological control of a weed with an arthropod or plant pathogen;
- post-establishment field study in region of introduction;
- original research report, containing field data and information on methods;
- stated or implied intent to evaluate effectiveness of biological control agent(s) in controlling the weed in the field; and
- published in an accessible source in 1971 or later.

I excluded papers describing the establishment, spread, or population dynamics of biological control agents, and containing no, or only casual or anecdotal, data on target weed impact, unless their *stated intent* was to evaluate the effectiveness of the agent. I also excluded review or historical papers, unless they contained details of methods and original field data. Although not a random sample, I believe these papers are typical of the methods used for evaluating the effectiveness of classical weed biological control over the last 20 yrs.

Aspects of the papers analysed included:

- the approach used for evaluation (experimental, observational, or informal/descriptive);
- the measures used to estimate target weed abundance or performance;
- the use of replication, randomization and experimental design; and
- approaches to data analysis and presentation.

A full list of the papers selected and of the categories used in the analysis is available on request.

Results

Approaches to Evaluation

Of the 57 papers analysed, only 16 (28%) used experimental methods as defined above. Among these the most popular technique was the insecticidal (or fungicidal) check (10 papers). In 4 studies, release sites were compared with control sites. Only 2 studies used an exclusion or confinement cage check.

Forty-two papers (74%) used systematic observational methods of evaluation; the most popular approach (29 papers) was a longitudinal study, in which one or several sites were sampled repeatedly over a period ranging from months to years. If the study period begins at or before the time of agent release, this is the traditional "before-and-after" study. Only 9 of the longitudinal studies, however, reported data recorded at or before the time of release, and many began several years after agent establishment. Ten studies reported data from control or unattacked sites or plants, and 13 gave neither pre-release nor control site data. The other frequently-used observational approach (15 papers) was to compare data taken simultaneously from sites or plants attacked by the biological control agent and from sites or plants to which the biological control agent has not spread. These differed from the experimental studies referred to above in that the selection of release and control sites was not made by the investigator as part of the study.

Six papers (11%) used only an informal or descriptive approach to evaluation; these papers may have included some quantitative data, but did not describe a systematic sampling scheme or experimental design. A further 27 papers (47%) contained an informal or descriptive account as well as more systematic observational or experimental data. The use of sequential photographs is often recommended (Goeden and Ricker 1981); such photographs were used in 7 (12%) of the papers.

Types of Measures Used

Forty-five papers (79% of the total) reported measures taken at the level of individual plants. The most commonly used (34 papers) were measures of intensity of attack, such as percentage of stems mined or number of larvae per seed head. Numbers of structures produced (leaves, flowers, stems etc.) were reported in 24 papers, plant dimensions (height, diameter etc.) in 20 papers, reproductive output in 19, mortality, longevity, etc., in 13, and biomass in 10 papers.

Thirty-one papers (54% of the total) reported population-level measures on the weed. The most commonly used was number of plants per unit area (23 papers). The only other frequently used population-level measures were biomass per unit area and percentage cover, both reported in 11 papers. Reproductive output per unit area and seed densities in the soil were reported in only 2 and 1 paper, respectively.

Population-level measures were used less frequently in experimental papers (35%) than in observational papers (55%).

Design, Analysis and Data Presentation

These were difficult to categorize, as plot layouts, experimental designs, and sampling schemes were often described in vague terms. In the observational studies, it was not practicable to draw a sharp distinction between replication and multiple sampling, so all studies where quantitative measures were calculated from multiple samples at any level were considered to be replicated. Experimental studies were considered to be replicated if they involved the systematic application of the same treatments and assessment methods on more than one experimental unit (plant, plot, release site etc.) Studies were identified as being "pseudoreplicated" (Hurlbert 1984) when multiple subsamples taken from unreplicated treated and control plots or sites were explicitly or implicitly used to test for treatment effects.

Of the 42 observational studies, 27 were replicated at some level, 11 were unreplicated, and 5 were pseudoreplicated. Fourteen made reference to some randomization procedure in sampling.

Among the 16 experimental studies, 12 were replicated, 1 was unreplicated, and 4 were pseudoreplicated (1 paper described a factorial experiment which was properly replicated with regard to one factor but pseudoreplicated with regard to another). Only 6 papers referred to randomization of treatments, and 5 mentioned a specific experimental design (e.g. randomized complete block, Latin square). In 1 of these cases a figure showed clearly that the plot layout did not in fact correspond to the design stated in the methods section, and in some others vague descriptions of layout left this point in doubt.

In the whole sample of papers, most results were presented in the form of simple calculation of means from samples, presented graphically or in tables, without statistical analysis (26 papers, 46% of total). In 8 papers (14%) means were compared by *t*-tests or similar elementary significance tests, and 11 papers (19%) stated that analysis of variance was used. In only 1 of these papers, however, was a full analysis of variance table presented, showing sources of variation, degrees of freedom, mean squares and F-values. A variety of other statistical and mathematical analyses, including regression and life-tables, were used in 11 papers (19%).

Discussion

Existing Practice

This survey has shown that the experimental approach has been infrequently used in evaluating the effectiveness of weed biological control agents. In those experimental studies which have been done, a majority focused on effects at the level of individual plants rather than on the weed population. Data on weed populations were reported more frequently in observational studies than in experimental studies. Almost half of the observational studies, however, reported neither pre-release nor control site data, and therefore tell us little about the effectiveness of the biological control agent.

Although most experimental studies were replicated in some way, few mentioned that treatments were randomized and few specified any formal experimental design. Pseudoreplication was common in both observational and experimental studies: for example, in insecticidal check experiments, multiple samples from each of a pair of sprayed and unsprayed plots may be treated as if they were replicates. Clearly, very few of the 57 papers surveyed would be capable of providing a statistically valid, experimental demonstration of the effectiveness of a biological control agent. By contrast, a comparison of these results with any issue of *Weed Science* or *Weed Research* will show that in the evaluation of chemical weed control, the use of replicated, randomized, field plot experiments, with a clearly specified design and a full statistical analysis, is the norm.

There are a number of reasons for the neglect of experimental method in classical weed biological control. Most biological control agents, unlike herbicides, are actively mobile, and can disperse between experimental plots. Thus, either widely separated plots must be used, or agents must be chemically or mechanically excluded from control plots. Wide separation increases between-plot variability and the amount of land required. Chemical and mechanical exclusion methods are labour-intensive and expensive. Experimental plots must be maintained over the months or years required for biological control to take effect, rather than the time-scale of days or weeks for chemical control. This adds to the cost and logistic difficulty of experiments. In the early stages of biological control projects, the agents are often not available in sufficient numbers for replicated tests. At this stage investigators may wish to release their insects in what they feel will be the most suitable sites for establishment, rather than allowing any randomization procedure to influence the selection of release sites. The history of weed biological control suggests another explanation: some successful projects have led to such spectacular control of once abundant weeds that quantitative evaluation might seem superfluous. Not all successes, however, will be so self-evident.

Suggestions for Future Practice

To preserve the credibility of the discipline, standards of proof in biological control of weeds should be as rigorous as in any other area of applied biological research. This may become particularly important as private sector involvement and government regulation increase. Regulation may eventually require documented proof of the efficacy of a biological control agent before it can be commercially distributed, and regulators are used to seeing statistically analysed data from properly designed and planned experiments. Implementation of a successful biological control project sometimes involves committing considerable resources to rearing or collection and redistribution of agents. This money will be better spent, and may be more readily available, if it is focused on agents whose efficacy has been rigorously demonstrated. The 4 elements listed above should be kept in mind when planning the evaluation process.

Because of the scarcity of material in the early stages of projects, I suggest that experimental evaluation should begin when an agent is established in large enough numbers at one or more sites to provide sufficient numbers for replication, and when there is enough damage to the target weed to give some grounds for believing that the agent may be effective. By this time sufficient information should be available on the biology of the agent and its interactions with the target weed to guide the design of experiments.

To design and analyse field experiments correctly, it is important to understand the distinctions among experimental units (*EU*), replicates and subsamples (e.g., Tietjen 1986). An experimental unit is any unit of study which can be assigned a particular treatment independently of the treatments assigned to other *EUs*. A replicate is 1 of a series of *EUs* that get the same treatment. For example, 6 (simultaneous) quadrat samples from within a single field cage are not replicates because they are not independent; if insects are released into the cage, all 6 quadrats will necessarily receive the insect treatment. The cage is the *EU* and the quadrats are subsamples. Replication of this experiment would require the use of multiple cages. Subsampling may increase the accuracy of measurements on individual *EUs*, and thus reduce experimental error, but it does not provide additional degrees of freedom for hypothesis testing.

The essential roles of replication and randomization of treatments in ecological field experiments are clearly described by Hurlbert (1984). Together they ensure that, on the average, all experimental units differ only in what treatments they receive. It is not necessary that replicate plots be identical, although the more uniform they are the fewer replicates will be needed. The crucial point is not identity of plots but random assignment of treatments. Many possible experimental designs and plot layouts are available to control for various types of environmental variability and to allow various hypotheses and combinations of treatments to be tested (e.g., Anderson and McLean 1974, Box *et al.* 1978, Gomez and Gomez 1984). It is important to select an appropriate design at the beginning of a study to ensure that a valid analysis will be possible.

The appropriate experimental methods and design to evaluate a particular agent, or combination of agents, against a particular target weed depend on many factors. One of the most widely varying of these is the agent's dispersal rate. *Aphthona nigriscutis* Foudras (Coleoptera: Chrysomelidae) released against leafy spurge in Canada spreads at a rate of 5-10 m yr^{-1} for the first few years (A.S. McClay and P. Harris, unpublished data); in contrast, *Epiblema strenuana* (Walker) (Lepidoptera: Tortricidae) released against parthenium weed in Queensland spread at a rate of up to 1.6×10^5 m yr^{-1} (McFadyen 1985).

When dispersal is slow, the use of replicated pairs of release and control plots, separated by sufficient distance that the agents will not reach the control plots for 2-3 yrs, may be a suitable experimental method. The release plot should be selected randomly from within each pair of plots--it is not valid to select a release site and then choose a nearby "control" site after the fact. To reduce between-plot error the release and control plots within each pair should be matched as closely as possible with regard to their soil type, topography, vegetation etc. Such error can be further controlled by using measurements of environmental factors at the sites as covariates. This is a randomized complete block design with 2 treatments. Data can be analysed by analysis of variance if the appropriate assumptions are met. A non-parametric method, the Wilcoxon signed-ranks test, is also available if these assumptions are violated (Sokal and Rohlf 1981). This type of experiment could be incorporated at very little additional cost into a standard redistribution and monitoring program.

When dispersal is rapid, experimental evaluation requires some type of exclusion, whether by cages, insecticides, mechanical removal or a combination. A good example of the combined use of these techniques is described by McEvoy *et al.* (1991). Cages are expensive and vulnerable to damage; they also produce alterations in microclimate which must be controlled for by sham cages. A number of different insecticides have given satisfactory results in insecticidal-check trials (e.g. Brown *et al.* 1987, Gordon *et al.* 1985, Louda 1984). Plots used in exclusion experiments can be fairly close together, tending to reduce between-plot error as well as travel times and expense. Blocking and covariates can also be used to reduce between-plot error in this type of experiment.

It is obviously not possible here, nor am I qualified, to make recommendations on the details of experimental design for every possible evaluation program. My purpose is to argue that biological control researchers should be conscious of the experimental nature of their work when evaluating the success of biological control, that field experiments should be planned with sufficient replication and an appropriate design to test a specific hypothesis, and that experimental design and analysis should be described in explicit detail in publications. Most researchers have access to statisticians or biometricians who can advise on experimental design and analysis; their help should be sought when an evaluation is being planned, and not after the data have been collected.

Field experiments can certainly be expensive and time-consuming, and funding for weed biological control is always scarce. However, I believe the onus is on weed biological control researchers, when preparing work plans and proposals, to provide for appropriate experimental tests to determine whether biological control agents actually do the job for which they were released. We need to make clear to all project clients that such experiments are not an optional extra, but an essential part of the program.

Acknowledgements

I thank my colleagues Lloyd Dosdall, Zack Florence, Peter Harris, John O'Donovan, and Paul Sharma for valuable comments and discussions.

References

Anderson, V.L. and R.A. McLean. 1974. *Design of Experiments: A Realistic Approach*. Dekker, New York, 418 p.

Box, G.E.P, W.G. Hunter, and J.S. Hunter. 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. Wiley, New York, 653 p.

Brown, V.K., A.V. Gange, I.M. Evans, and A.L. Storr. 1987. The effect of insect herbivory on the growth and reproduction of two annual *Vicia* species at different stages in plant succession. *Journal of Ecology* **75**:1173-89.

Crawley, M.J. 1990. Plant life-history and the success of weed biological control projects. *Proceedings of the VII International Symposium on Biological Control of Weeds*, 6-11 March 1988, Rome, Italy. Delfosse, E.S. (ed.). Ministero dell' Agricoltura e delle Foreste, Rome/CSIRO, Melbourne, pp. 17-26.

Goeden, R.D. and D.W. Ricker. 1981. Santa Cruz Island - revisited. Sequential photography records the causation, rates of progress, and lasting benefits of successful biological weed control. *Proceedings of the V International Symposium on Biological Control of Weeds*, 22-27 July 1980, Brisbane, Australia. Delfosse, E.S. (ed.). CSIRO, Melbourne. pp. 355-65.

Gomez, K.A. and A.A. Gomez. 1984. *Statistical Procedures for Agricultural Research*. Second edition. Wiley, New York, 680 p.

Gordon, A.J., R.L. Kluge and S. Neser. 1985. Effect of the gall midge, *Zeuxidiplosis giardi* (Diptera: Cecidomyiidae) on seedlings of St. John's wort, *Hypericum perforatum* (Clusiaceae). *Proceedings of the VI International Symposium on Biological Control of Weeds*, 19-25 August 1984, Vancouver, Canada. Delfosse, E.S. (ed.). Agriculture Canada, Ottawa, pp. 743-8.

Harris, P. 1981. Evaluating biocontrol of weeds projects. *Proceedings of the V International Symposium on Biological Control of Weeds*, 22-27 July 1980, Brisbane, Australia. Delfosse, E.S. (ed.). CSIRO, Melbourne. pp. 345-53.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187-211.

Louda, S.M. 1984. Herbivore effect on stature, fruiting and leaf dynamics of a native crucifer. *Ecology* **65**:1379-86.

McEvoy, P., C. Cox and E. Coombs. 1991. Successful biological control of ragwort, *Senecio jacobaea*, by introduced insects in Oregon. *Ecological Applications* **1**:430-42.

McFadyen, R.E. 1985. The biological control programme against *Parthenium hysterophorus* in Queensland. *Proceedings of the VI International Symposium on Biological Control of Weeds*, 19-25 August 1984, Vancouver, Canada. Delfosse, E.S. (ed.). Agriculture Canada, Ottawa, pp. 789-96.

McLaren, I.W. and R.L. Amor. 1979. Assessment of effectiveness. Australian Applied Entomological Research Conference, Queensland Agricultural College, June 1979, pp. 127-40.

Myers, J.H. 1978. Biological control introductions as grandiose field experiments: adaptations of the cinnabar moth to new surroundings. *Proceedings of the IV International Symposium on Biological Control of Weeds*, 30 August-2 September 1976, Gainesville, FL, USA. Freeman, T.E. (ed.). University of Florida, Gainesville, pp. 181-8.

Peschken, D.P. and A.T. Wilkinson. 1981. Biological Control of Canada thistle (*Cirsium arvense*): releases and effectiveness of *Ceutorhynchus litura* (Coleoptera: Curculionidae) in Canada. *Canadian Entomologist* **113**:777-85.

Rees, N.E. 1990. Establishment, dispersal and influence of *Ceutorhynchus litura* on Canada thistle (*Cirsium arvense*) in the Gallatin Valley of Montana. *Weed Science* **38**:198-200.

Sokal, R.R. and F.J. Rohlf. 1981. *Biometry: The Principles and Practice of Statistics in Biological Research*. Second edition. W.H. Freeman, New York, 859 p.

Tietjen, G.L. 1986. *A Topical Dictionary of Statistics*. Chapman and Hall, New York, 171 p.

Wapshere, A.J., E.S. Delfosse and J.M. Cullen. 1989. Recent developments in biological control of weeds. *Crop Protection* **8**:227-50.