Zedaker, S.M.; Gregoire, T.G.; Miller, J.H. 1994. Sample-size needs for forestry herbicide trials. Canadian Journal of Forest Research, 23(10): 2153-2158.

Sample-size needs for forestry herbicide trials[1]

S.M. ZEDAKER[2] AND T.G. GREGOIRE

Department of Forestry, College of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0324, U.S.A.

AND

J.H. MILLER

George W. Andrews Forestry Sciences Laboratory, Southern Forest Experiment Station, USDA Forest Service, Auburn, AL 36849, U.S.A.

ZEDAKER, S.M., GREGOIRE, T.G., and MILLER, J.H. 1993. Sample-size needs for forestry herbicide trials. Can. J. For. Res. 23: 2153–2158.

Forest herbicide experiments are increasingly being designed to evaluate smaller treatment differences when comparing existing effective treatments, tank mix ratios, surfactants, and new low-rate products. The ability to detect small differences in efficacy is dependent upon the relationship among sample size, type I and II error probabilities, and the coefficients of variation of the efficacy data. The common sources of variation in efficacy measurements and design considerations for controlling variation are reviewed, while current shortcomings are clarified. A summary of selected trials estimates that coefficients of variation often range between 25 and 100%, making the number of observations necessary to detect small differences very large, especially when the power of the test $(1 - \beta)$ is considered. Very often the power of the test has been ignored when designing experiments because of the difficulty in calculating β. An available program for microcomputers is introduced that allows researchers to examine relationships among sample size, effect size, and coefficients of variation for specified designs, α and β. This program should aid investigators in planning studies that optimize experimental power to detect anticipated effect sizes within resource constraints.

ZEDAKER, S.M., GREGOIRE, T.G., et MILLER, J.H. 1993. Sample-size needs for forestry herbicide trials. Can. J. For. Res. 23 : 2153–2158.

Les expériences sur les herbicides forestiers, sont de plus en plus structurées pour évaluer de petites différences entre les traitements lorsqu'on compare les traitements actuellement utilisés, les rapports des mélanges dans les réservoirs, les surfactants et les nouveaux produits utilisés à faible dose. La capacité de détecter de petites différences d'efficacité dépend de la relation entre la dimension des échantillons, des erreurs en probabilité de types I et II et des coefficients de variation des données d'efficacité. Les sources communes de variation dans les mesures d'efficacité, et les considérations de structure pour contrôler la variation, sont revisées pendant que les défauts courants sont clarifiés. À l'aide d'un sommaire d'essais sélectionnées, on estime que les coefficients de variation se situent souvent entre 25 et 100%, ce qui nécessite un très grand nombre d'observations pour détecter de petites différences surtout lorsque la puissance du test $(1 - \beta)$ est considérée. Très souvent la puissance du test a été ignorée dans la structure des expériences à cause de la difficulté à calculer β. Nous introduisons un programme disponible sur micro-ordinateur qui permet aux chercheurs d'examiner les relations entre la dimension des échantillons, les effets de la dimension et des coefficients de variation pour des structures données, α et β. Ce programme devrait aider les chercheurs à planifier des expériences qui optimisent la capacité de détecter expérimentalement les effets de dimension anticipés avec des ressources limitées.

[Traduit par la rédaction]

Introduction

One of the most critical steps in designing forestry herbicide trials is determining the number of experimental units (trees, rootstocks, plots) for replication. The land area available for experimentation, the money, time, and personnel resources, and the number and type of chemical applications to be tested are usually considered, with the most limiting variable determining sample size (n). Little attention is directed toward the resolution with which differences in efficacy (measured as δ_i (the "effect size" of treatment i, or $\mu_i - \mu$) Cohen 1988) among chemicals, rates, and (or) application techniques might be detected, and the probability that differences might exist but would not be found. This is unfortunate, since many decisions as to which herbicide or herbicide mix to use, at what rate, and using which equipment hinge on the detection of marginal differences in the control and cost of treatment. The utility of different measures of efficacy for woody plant herbicide trials has been analyzed and reviewed in numerous reports (Knowe 1991; Zedaker and Miller 1991; Knowe et al. 1990; Zedaker and Freyman 1988; Zedaker and Lewis 1983; Neill et al. 1982). However, none of these studies addressed the variation in efficacy measures relative to their impact on sample-size adequacy. Nor do the current manuals of methods research for forestry and agriculture weed science address the procedures for the determination of an appropriate sample size (cf. Camper 1986; Miller and Glover 1991).

Statistical procedures necessary to determine the adequacy of a particular sample size for simple experimental designs have been well known for some time (Steel and Torrie 1960; Cochran and Cox 1957). The adequacy, or precision, of an experiment refers to its ability to detect treatment effects. In general, the more precise an experiment is, the smaller the difference it is capable of detecting between treatments (Little and Hills 1978). Sample size and precision are related by the probabilities associated with type I (α) and type II (β) errors for hypotheses testing, the variability or dispersion of the variable of interest (variance, coefficient of variation (CV), etc.), and the experimental design (completely randomized,

[1]This paper was presented at the International Conference on Forest Vegetation Management: Ecology, Practice and Policy held April 27 – May 1, 1992, Auburn University, Auburn, Alabama, U.S.A., and has undergone the Journal's usual peer review.

[2]Author to whom all correspondence should be addressed.

Printed in Canada / Imprimé au Canada

randomized block, etc.). Unfortunately, the variation associated with efficacy measures in forestry herbicide trials is rarely published in a useful form, and agricultural statistics texts generally give only the most basic methods of determining sample size for simplistic scenarios. Calculating sample sizes for some statistical tests can be very difficult (Mendenhall and Scheaffer 1973), and although complex graphs and computer programs are available for this task (Odeh and Fox 1975; Dallal 1986), none that the authors have found allow direct comparison of the relationship between sample size and precision without multiple iterations of use. In this paper we discuss the sources of variation in herbicide trials, present typical CVs of efficacy measures observed in herbicide trials with woody plants, and provide graphs from a publicly available computer program that allow direct comparison of sample size – precision trade-offs for stipulated levels of variability, α and β.

Variation in efficacy measures

Variation in efficacy measures is influenced by the experimental unit, the sample size, properties inherent to the efficacy measures themselves, and even the range of efficacy found as a consequence of the herbicide trial design. For woody plant herbicide trials, the experimental units are either individual rootstocks or plots (Glover 1991). Intuitively, the use of individual rootstocks as an experimental unit seems appropriate for treatments applied to individuals such as basal sprays, injection, soil spot-around, and directed foliar spray applications. With sufficient care, the application to each individual will be relatively uniform and the sources of variability associated with different chemical treatments will be those inherent to the rootstocks, the treatments, and their interaction. By convention, variation not associated with treatments is considered error (ε), and whatever the source of experimental errors, replication steadily decreases the nonsystematic error associated with the difference between the average response for different treatments (Cochran and Cox 1957). A number of authors suggest that 30 to 50 individuals is an appropriate number of woody rootstocks to sample for each treatment (Kline et al. 1985; Glover 1991), without specific justification. Although no data have been published on the effect of sample size on variance for woody plant herbicide trials, the variance and CV observed in samples from populations of normally distributed individuals begins to stabilize (change very slowly) as n increases. Figure 1 depicts the change in CV with increasing sample size for two populations of 100 randomly selected observations. Population A simulates observations from a herbicide trial with widely ranging efficacy results (20–100%) and a mean of 56% control; population B simulates observations from a trial with a narrow range (80–100%) and mean of 89%. In both cases CVs begin high and decrease with decreasing rates of change as n increases. CVs seemingly become stable at similar n-values, even with these very different populations.

For broadcast applications of herbicides (both liquid and dry formulations), the use of plots of land as experimental units may be more appropriate. Most often, observations or measurements of efficacy are still made on individual rootstocks, which are treated as subsamples. Individual rootstock data may be averaged or summed to provide a plot-wide efficacy estimate. Other sources of error, in addition to those mentioned for individual rootstocks, are introduced as a result of herbicide application and (or) are inherent in the properties of an individual plot. Even with the utmost care, broadcast applications result in swath skips, overlaps, variation in application rate across the swath, and cases where individual rootstocks receive less chemical as a result of shielding by other plants. Double applying at one-half rates, use of only one swath, and (or) measuring only rootstocks that "appear" to have received a full dose, as is practiced for many experiments, still cannot completely eliminate these additional sources of variation. Variability in soil physical properties between plots may be a particularly large source of variation in soil-active herbicide trials. Lacking specific variance data, three to five plots in which 10 or more individual rootstocks per species have been measured (summing to the 30 to 50 individuals) have been recommended (Glover 1991).

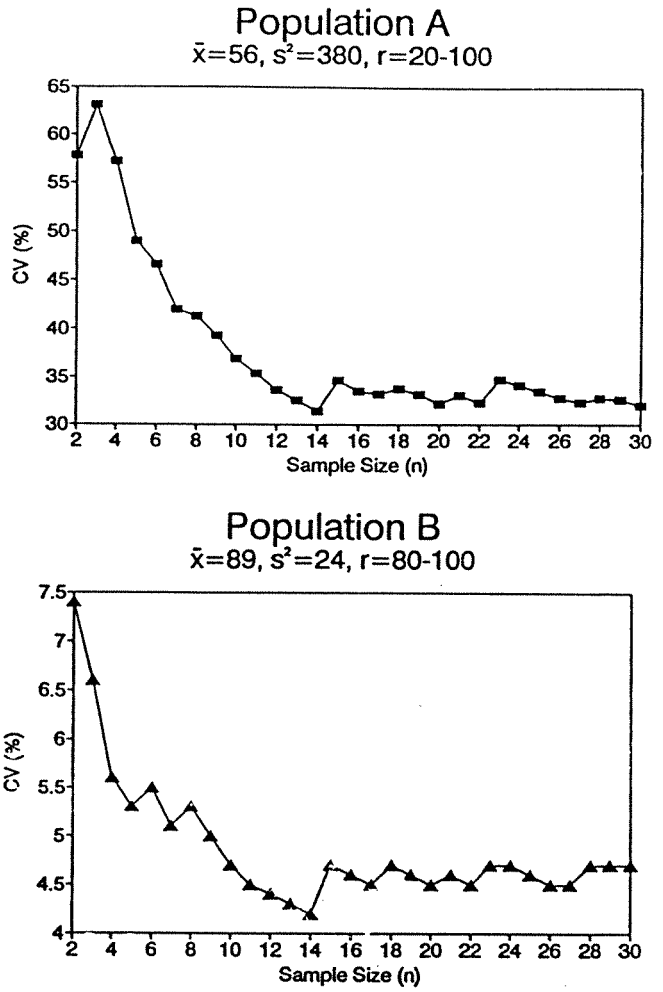

FIG. 1. Coefficients of variation for the first 30 observations drawn at random from two simulated (100 random, normally distributed numbers) populations of herbicide efficacy observations with ranges of 20–100% control (population A) and 80–100% control (population B).

The size of each plot can influence the variation associated with estimates of the per-unit-area efficacy responses. In mensurational studies, small sample plots usually exhibit larger CVs (relative variability) than large plots (Avery and Burkhart 1983). Larger plots tend to average out local variation in the spatial distribution of efficacy effects. Within-plot

TABLE 1. Example, experiment-wide coefficents of variation (%) for woody plant herbicide trials, with individual plants as the experimental unit

Study type	Application	Treatment	No. of levels	Efficacy variable	Species[a]											
					BC	BG	BL	CO	DW	HI	LP	SG	RM	RO	WO	YP
				Visually estimated efficacy												
Release	Backpack, basal directed	Rate	3	Top kill	45	92	25		73	52	16	71	46	27		96
Site preparation	Backpack, foliar broadcast	Mixes and rates	14	Top kill					24		95	26				
Right-of-way	Backpack, basal directed	Mixes	9	Crown volume									5		76	
Site preparation	Backpack, soil broadcast	Rates	9	Crown volume							52	70	58	89		
Site preparation	Backpack, soil broadcast	Formulation	2	Top kill								90		83		
				Measured efficacy												
Site preparation	Backpack, foliar broadcast	Mixes and rates	13	Crown volume		29		37					36	41	36	
Right-of-way	Tractor, foliar broadcast	Rates and adjuvants	11	Height growth			77						64		92	
Release	Backpack, foliar broadcast	Rates	4	Crown volume							17		48	55	115	22
Release	Backpack, foliar directed	Mixes	5	Crown volume	56	43				14	41	18	18	15	57	
Site preparation	Backpack, foliar broadcast	Mixes and rates	13	Height growth		37		39					46	53	43	

[a]BC, black cherry (*Prunus serotina* Ehrh.); BG, black gum (*Nyssa sylvatica* Marsh.); BL, black locust (*Robinia pseudoacacia* L.); CO, chestnut oak (*Quercus prinus* L.); DW, dogwood (*Cornus florida* L.); HI, hickory (*Carya* spp.); LP, loblolly pine (*Pinus taeda* L.); SG, sweetgum (*Liquidambar styraciflua* L.); RM, red maple (*Acer rubrum* L.); RO, northern red oak (*Quercus rubra* L.); WO, white oak (*Quercus alba* L.); YP, yellow-poplar (*Liriodendron tulipifera* L.).

variation may be particularly important, for example, when determining the effects of treatment on survival or mortality of target or crop species in heterogeneous stands. As plot size increases, it may be difficult to identify and fit an adequate number of plots in a study location.

Efficacy can be estimated visually or measured (Zedaker and Miller 1991). Visual assessments entail the estimation of percent of rootstock's top that is damaged or killed by treatment (percent defoliation, crown reduction, etc.; see Miller and Glover 1991). With practice, observers can become reasonably consistent (minimizing repeat sampling error) at categorizing or estimating the condition of crowns of even very large trees (Zedaker and Nicholas 1991; Nicholas et al. 1991). However, a source of variation that has not been addressed is observer bias. Bias, in this case, is the systematic difference between the mean of a visually estimated or observed effect and the mean of that effect when actually measured. For example, if, after looking at a large number of trees, an observer estimates that treated trees had 50% smaller (aerial volume) crowns after treatment, when the actual change in crown volume of those trees measured was 60%, the observer has underestimated the efficacy of the treatment by 10% (bias is −10%). If that bias was consistent across the full range of treatments and the observer had no vested interest in any of the treatments or never knew which treatment he or she was observing, this bias would not impact the experimental outcome. Any violation of these assumptions would invalidate the experiment. Even in the absence of a violation, comparability of experimental results is jeopardized unless observer bias is constant; to the authors' knowledge, no estimates of observer bias (checks of observations against measured standards) have been published in any study using visually estimated woody plant herbicide efficacy assessments.

Measured attributes (height, diameter, crown volume index (diameter^2 × height), etc.) used for efficacy assessment do not suffer from observer bias but are not necessarily inherently less variable. A survey of 10 herbicide trials did not indicate consistent patterns between the CVs observed when efficacy was visually estimated and when it was measured (Table 1). Observed CVs (on the basis of individual rootstocks as replicates) within species, pooled across treatments, varied mostly between 24 and 96, with few less than 20 or exceeding 100. Although the survey is not exhaustive, it provides a suitable range of CVs for different types of studies, application techniques, efficacy variables, and woody plant species that should enable researchers to design experiments better suited to account for the variability of efficacy data. For both measured and visually estimated assessments, any efforts to reduce the error associated with the observation of individual rootstocks (consistently holding a height pole straight, measuring DBH at exactly 4.5 ft (1 ft = 0.30 m) each time, practicing visual estimation on plants with measured attributes, etc.) will be rewarded with lower CVs and increased experimental accuracy.

Finally, the design of the herbicide trial itself, or more specifically the concentration of treatments relative to the expected efficacy range, affects the variability of the results. For example, an experiment designed (using high herbicide rates, very effective herbicides, easy to control species, etc.) so that the range of treatment efficacy varied only between

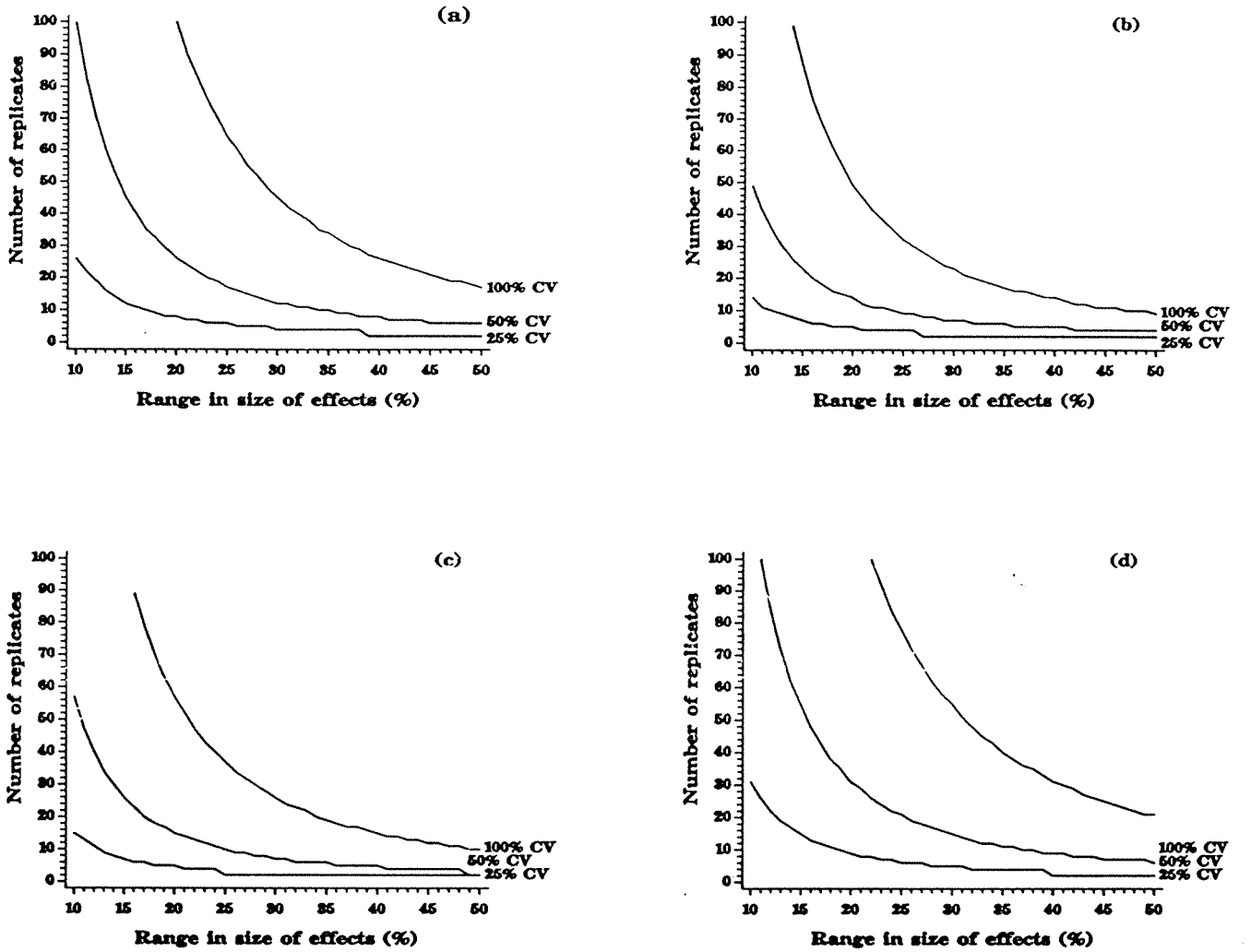

FIG. 2. Sample-size curves for a one-factor completely randomized design. (*a*) Base case sample-size curves when $\alpha = 0.05$ and $\beta = 0.20$, with two treatment levels. (*b*) Sample-size curves when $\alpha = 0.05$ and $\beta = 0.50$, with two treatment levels. (*c*) Sample-size curves when $\alpha = 0.20$ and $\beta = 0.20$, with two treatment levels. (*d*) Sample-size curves when $\alpha = 0.05$ and $\beta = 0.20$, with five treatment levels.

80 and 100% control will have smaller experiment-wide CVs than one in which the expected range of control varied between 20 and 100%.

Obtaining an adequate sample

Forestry herbicide researchers have many different objectives for obtaining efficacy data (Glover 1991); however, the unifying goal is the detection of differences in efficacy between treatments. Although experiments are conducted that produce widely varying results (50 to 80% differences in efficacy), tightly controlled, statistically adequate experiments are needed for the detection of small differences in treatment efficacy. Well-replicated and carefully measured experiments are not necessary for the detection of large treatment differences (see Fig. 2). It is for experiments attempting to detect the effect of treatments within 30% of the mean that CVs, hypotheses testing error probabilities, and sample size are so important. Such "fine-tuning" experiments (known as secondary field evaluations, Glover 1991) are most common once gross differences in chemicals or rates have been discovered (in primary field evaluations). Woody plant herbicide efficacy experiments in which the major objective is to detect small (≤25%) differences in effect due to differences in active ingredient rates, herbicide mixes of two chemicals in various ratios, different surfactant types and rates, and different formulations of the same active ingredient are commonplace. Such experiments often have inadequate power to detect statistically significant differences between treatments due to inadequate sample size and the inherent variability in the observations.

In efficacy experiments, researchers usually test the null hypothesis that herbicide treatments are not different (H_0: $\mu_1 = \mu_2 = \mu_3 \ldots \mu_i$). Researchers focus on the probability of committing a type I error (α, i.e., the probability of rejecting a true H_0) with little attention to the probability of committing a type II error (β, the probability of accepting a false H_0) (Gregoire and Driver 1987). However, the objectives of many studies dictate that protection from a type II error, failure to find a difference of some meaningful magnitude, is as important, if not more so, than false claims of a disparity between treatments. Small differences in efficacy could result in large differences in profitability as a result of changes in growth and yield. For example, the volume yield of loblolly pine is reduced disproportionately by small changes in the

TABLE 2. Means and coefficients of variation (CV) for percent control of crown volume on eight Virginia tracts 1 year after aerial application for release with glyphosate

	Individual trees[a]		Tracts[b]	
	Mean[c] (%)	CV (%)	Mean (%)	CV (%)
Loblolly pine	14	300	15	193
Red maple	48	110	53	55
Yellow-poplar	47	119	59	41
White oak spp.	65	83	80	26
Sweetgum	71	55	92	13

NOTE: Data are from D'Anieri and Zedaker (1990).
[a]All trees on all tracts combined within a species.
[b]Mean for trees within each tract was used as the tract observation for a species, and then the means and CVs for all tracts were computed.
[c]Means differ for individual trees and for tracts because tract means are unweighted; i.e., the number of individuals measured on each tract varied between 20 and 40.

percent of total basal area made up by hardwoods in plantations (Burkhart et al. 1987). As a result, both landowners and chemical companies would be eager to detect small differences between herbicides or herbicide treatments in the efficacy of hardwood control. The ability to sell a chemical depends on a company's ability to differentiate it from those of its competitors. In herbicide efficacy studies, the conventional definitions of α and β errors as representing producers' and consumers' risks, respectively, may be inappropriate (Sokal and Rohlf 1981).

Statistical texts regard the precision, or ability to detect a difference of some stipulated magnitude between treatments, as the power of a test to detect that difference (see Steel and Torrie (1960), Cochran and Cox (1957), Sokal and Rohlf (1981), and Cohen (1988) for detailed discussions). By definition, power is equal to $1 - \beta$. Obviously, the higher the power of a test (the smaller the β), the more likely we are to detect a difference between treatments. Since we are often locked into a particular level of α by convention (in many refereed journals, 0.05), the only way to improve the power of a given test (when CVs are also set) is to increase the sample size. Conversely, if the power or β were fixed, increasing the sample size would decrease the difference between means of treatments at which a statistically significant difference would be found.

Unfortunately, for experiments with complex designs (factorial, randomized blocks, nested, split plots) and more than two treatment levels, the relationship between α, β, δ, and n is complex and dependent on the calculation of different noncentrality factors for different tests (Cohen 1988). The available methods allow the determination of the power of a test for a specific sample size, α and δ, or allow the determination of a sample size for a given power, but the relationship between n and δ for acceptable error rates (fixed α and β) cannot be determined without tedious iterative solutions of complex equations, iterative table or graph use, or multiple runs of existing computer programs (Odeh and Fox 1975; Sokal and Rohlf 1981; Han 1985; Cohen 1988; Dallal 1991). Since it is likely that researchers may want to fix both α and β, the appropriate n will be determined by compromising δ. That is, for a given cost of obtaining larger sample sizes, what is the difference between treatments that could be detected? Or conversely, if a specific δ is meaningful, we would want to know how many individual rootstocks or plots should be measured. Say, for example, that to make users switch from herbicide A to herbicide B, we must be able to detect a 10% difference in their efficacies, if A costs 10% more than B. Therefore, given that we are willing to set α and β probabilities, and that we know generally what CV to expect, why would we want to conduct an experiment that had little chance of detecting a 10% difference in efficacy between the two herbicides?

A computer program to allow direct comparisons of n and δ and CVs was written for IBM-compatible microcomputers. (The program inputs and outputs are described in the Appendix.) The calculations are based on formulas listed in Odeh and Fox (1975). Example graphical output from the program is contained in Fig. 2). The graphs depict the number of replicates, or n, necessary to detect a significant treatment effect for difference ranges in size of effects and CVs for a completely randomized design at different levels of α and β. In the construction of these graphs it was assumed that

$$Y_{ij} = \mu + \delta_i + \varepsilon_{ij}$$

where

Y_{ij} is the value (efficacy) of observation j in treatment i

μ is the grand, experiment-wide mean

δ_i is the effect of treatment i

ε_{ij} is the error, $\varepsilon \sim N(0, \sigma^2)$

and

δ_i', which equals $(\delta_i/\mu)100$, is the size of effect (or effect size) expressed as a percent of the mean.

For an experiment with T treatment levels, we entertained the null hypothesis (H_0: $\delta_i' = 0$, $i = 1, 2, 3, \ldots, T$) against an alternative hypothesis (H_1: $\delta_1' \neq 0$). To construct the curves, an array of specific alternatives was considered and for each specific alternative the noncentrality parameter was evaluated. There are countless ways that this could be done. We chose the alternative hypothesis that the δ_i' were uniformly distributed over the range $\max(\delta_i')$–$\min(\delta_i')$. For example, if our experiment had five treatment levels, and the range in size of effects was 50, the δ_i' would be 25, 12.5, 0, −12.5, and −25.

For discussion purposes, the conditions exhibited in Fig. 2*a* represent the base case where a one-factor completely randomized design with two treatment levels and $\alpha = 0.05$ and $\beta = 0.20$ is analyzed for three CV values (covering the majority of values obtained in Table 1, where $CV = (\sigma/\mu)100$). With a CV of only 25%, the number of replicates necessary to guard against a type II error (achieve an 80% power) is reasonable (less than 30) even for a small range of effects (10%). Increasing the experiment-wide CV to ≥50% makes the detection of small differences exceedingly difficult without very large (>100) sample size. If we increase the probability of a type II error to 50% (decrease our power to detect differences), we reduce the sample size necessary for detection by approximately one-half (Fig. 2*b*). Such low power is likely the rule, rather than the exception, for woody plant herbicide trials, judging by the sample sizes commonly reported in the woody plant efficacy literature and common CV values (Table 1). Increasing the probability of a type I error to 0.20 (Fig. 2*c*) has a nearly equivalent effect. Increas-

ing the number of treatment levels to five (Fig. 2*d*) increases the necessary sample size.

When plots or tracts are used as the experimental units, CVs are substantially reduced (Table 2). D'Anieri and Zedaker (1990) found that the use of tracts, as opposed to individual trees, as replicates reduced CVs 30 to 70%. However, it is common to install herbicide experiments on replicated plots on a single tract. Since plots within tracts might be expected to be even less variable than whole tracts, most herbicide efficacy experiments conducted on single tracts of land would have even lower experiment-wide CVs. For secondary field evaluations, where the range of efficacies is expected to be high (80–100%), the CVs using plots as replicates would be relatively low (perhaps <10%) and the corresponding precision of the experiment would be improved.

Conclusions

To detect small differences in the efficacy of herbicide treatments, within the range of typical variation in efficacy observations and for generally acceptable α and β probabilities, surprisingly large sample sizes are needed when individuals are the experimental unit. When plots (within which many (10 or more) individuals have been measured) are the experimental units, fewer replicates are needed because CVs are reduced. However, common herbicide trials with three or four replicate plots have relatively low power. Woody plant herbicide researchers would greatly improve their chances of conducting successful experiments (designing an experiment that would be able to detect differences in the treatments they are evaluating) if they more carefully assessed the trade-offs between precision and sample size. A listing of the program used to generate the sample size – range of effects graphs can be obtained by writing the senior authors at Virginia Polytechnic and State University, Blacksburg.

Acknowledgments

The authors thank N.S. Nicholas for providing assistance with computer analyses, DowElanco and the USDA Forest Service for providing financial resources, and the three referees who provided substantive comments to improve the manuscript.

Avery, T.E., and Burkhart, H.E. 1983. Forest measurements. 3rd ed. McGraw-Hill, New York.

Burkhart, H.E., Glover, G.R., and Sprinz, P.T. 1987. Loblolly pine growth and yield response to vegetation management Chap. 8. *In* Forest vegetation management for conifer production. *Edited by* J.D. Walstad and P.J. Kuch. John Wiley & Sons, New York.

Camper, N.D. (*Editor*). 1986. Research methods in weed science. 3rd ed. Southern Weed Science Society, Champaign, Ill.

Cochran, W.G., and Cox, G.M. 1957. Experimental designs. John Wiley & Sons, New York.

Cohen, J. 1988. Statistical power analyses for the behavioral sciences. 2nd ed. Lawrence Erlbaum Assoc., Hillsdale, N.J.

Dallal, G.E. 1986. PC:SIZE USERS manual. [Available from G.E. Dallal, 53 Beltran Street, Malden, Mass. 02148, U.S.A.]

Dallal, G.E. 1991. PC-SIZE: Consultant—a program for sample size determinations. Am. Stat. **44**: 243.

D'Anieri, P., and Zedaker, S.M. 1990. Variation in glyphosate efficacy in commercial loblolly-pine release operations in Virginia. For. Ecol. Manage. **31**: 233–243.

Glover, G.R. 1991. Experimental designs and sampling methods for forest herbicide studies. Chap. 2. *In* Standard methods for forest herbicide research. *Edited by* J.H. Miller and G.R. Glover. Southern Weed Science Society, Champaign, Ill. pp. 10–21.

Gregoire, T.G., and Driver, B.L. 1987. Type II errors in leisure research. J. Leisure Res. **19**(4): 261–272.

Han, C. 1985. NFD: noncentral F distribution. Am. Stat. **39**: 211.

Kline, W.N., Fears, R.D., and Zedaker, S.M. 1985. Procedures for evaluating woody plant control in forest management systems. Down Earth, **41**: 9–15.

Knowe, S.A. 1991. Comparisons of expressions for crown size of woody competitors in herbicide efficacy studies. For. Sci. **37**(6): 1664–1670.

Knowe, S.A., Shiver, B.D., and Borders, B.E. 1990. Evaluation of four estimators of herbicide efficacy for woody competition control studies. For. Sci. **36**(2): 201–211.

Little, T.M., and Hills, F.J. 1978. Agricultural experimentation: design and analysis. John Wiley & Sons, New York.

Mendenhall, W., and Scheaffer, R.L. 1973. Mathematical statistics with applications. Duxbury Press, Boston.

Mendenhall, W., and Scheaffer, R.L. 1986. Mathematical statistics with applications. 2nd ed. Duxbury Press, Boston.

Miller, J.H., and Glover, G.R. (*Editors.*) 1991. Standard methods for forest herbicide research. Southern Weed Science Society, Champaign, Ill.

Neill, R.H., Glover, G.R., and Campbell, T.E. 1982. Standardization of forestry herbicide research techniques in the south. Proc. South. Weed Sci. Soc. **35**: 143–155.

Nicholas, N.S., Gregoire, T.G., and Zedaker, S.M. 1991. The reliability of tree crown condition classification. Can. J. For. Res. **21**: 698–701.

Odeh, R.E., and Fox, M. 1975. Sample size choice. Marcel Dekker, Inc., New York.

Sokal, R.R., and Rohlf, F.J. 1981. Biometry. 2nd ed. W.H. Freeman & Co., San Francisco.

Steel, R.G.D., and Torrie, J.H. 1960. Principles and procedures of statistics. McGraw-Hill, New York.

Zedaker, S.M., and Freyman, R.C. 1988. Backpack foliar treatments for pine release. Proc. South. Weed Sci. Soc. **41**: 179.

Zedaker, S.M., and Lewis, J.B. 1983. Measurement and rating of target and crop species. Proc. South. Weed Sci. Soc. **36**: 280–289.

Zedaker, S.M., and Miller, J.H. 1991. Measurement of target and crop species. Chap. 4. *In* Standard methods for forest herbicide research. *Edited by* J.H. Miller and G.R. Glover. Southern Weed Science Society, Champaign, Ill. pp. 22–40.

Zedaker, S.M., and Nicholas, N.S. 1991. Quality assurance methods manual for forest site classification and field measurements. U.S. Environ. Prot. Agency Res. Dev. EPA 1600/13-90/082.

Appendix

The SAMSIZ program is written in SAS. No input data are required, but prior to invocation, there are a number of macro variables that must be initialized by the user. SAMSIZ produces an output data file. Each data record in the output file consists of four values. The first is the effect size expressed as a percent, i.e., the abscissa value in Fig. 2. The remaining values are the three coordinates corresponding to the three CV levels.

SAMSIZ invokes two SAS macro program segments that must reside on the disk drive indicated by macro-variable DRIVE. The other macro variables stipulate (*i*) the value of α; (*ii*) β, which is used to determine sample size; (*iii*) the experimental design; (*iv*) the number of treatments; and (*v*) the CV levels. Comments at the beginning of the program define these variables and stipulate permissible values.