Dumroese, R. Kasten; Wenny, David L. 2003. Installing a practical research project and interpreting research results. Tree Planters' Notes 50(1): 18-22

Installing a Practical Research Project and Interpreting Research Results

R. Kasten Dumroese and David L. Wenny

Research Plant Physiologist, USDA Forest Service, Southern Research Station, Moscow, ID
and Professor/Director, Forest Research Nursery, Department of Forest Resources, University of Idaho, Moscow, ID

The basic concepts of the scientific method and research process are reviewed. An example from a bareroot nursery demonstrates how a practical research project can be done at any type of nursery, meshing sound statistical principles with the limitations of busy nursery managers. Tree Planters' Notes 50(1): 18-22; 2003.

Although they may not realize it, most growers already do nursery research. Have you ever done the following: (1) contemplated a problem at your nursery, (2) had an idea how that problem might be corrected after reading an article or discussing it with a colleague, (3) put in trials to test your guess, and (4) decided if your idea solved the problem? If so, you have done scientific research. Depending on how the research is done, the process can provide accurate and useful information, or it can yield conclusions that are meaningless. Our objective is to help growers design projects that yield meaningful results. Once you can design a good experiment, you can also tell if published research results are generated by a well-designed experiment and are worthy of consideration.

What Is Research?

Science is the possession of knowledge attained through study or practice. Research is the systematic search for new knowledge. Scientific research, simply stated, "is the testing (systematic, controlled, empirical, and critical investigation) of ideas (hypothetical propositions about presumed relations among natural phenomena) generated by intuition" (Stock 1985). Scientific research is carried out using the scientific method, which has 5 distinct steps (table 1). The process begins with observation, which can be practical experience, a literature review, or conversations with other nursery managers. It is followed by problem definition: specific questions are asked that you hope to answer. Third, the hypothesis is formulated and methods are selected for testing the hypothesis, based on the defined objectives. The 4th step, testing the hypothesis, involves collection, analysis, and interpretation of data. And finally, the hypothesis is accepted, rejected, or modified (Stock 1985).

Table 1—*Steps in the scientific method*

Steps	Example
1. A natural phenomenon is observed.	Ships sailing from port gradually disappear from sight, with the tops of the masts being the last part seen.
2. The problem is defined.	If the world is flat, why do ships gradually disappear from bottom to top?
3. A hypothesis (a guess) is made.	The world is not flat, but round.
4. A test is conducted.	Sail west from port and see if you return to where you started without falling off the end.
5. A theory is formulated.	The world is round.

When sufficient investigation is completed, a theory may be formulated. Theories are general explanations of natural events that are useful to understand, predict, and control natural phenomena. When installing practical research projects at our nurseries, we are probably not concerned with developing broad, sweeping theories of the universe. But we are interested, for example, in whether or not it is cost effective, in terms of improved growth, to double the amount of magnesium (Mg) we apply to 1+0 black cherry (*Prunus serotina* Ehrh.). To illustrate how an experiment is designed, we offer an example to answer this question using the steps of the scientific method. The same approach can be applied to any number of similar questions. Experiments are designed the same way whether you grow bareroot or container seedlings.

Following the Scientific Method—An Example

Observation. After a usually competent employee accidentally applies twice (2×) the normal amount of Mg to a bareroot bed of 1+0 black cherry, those seedlings appear taller than an adjacent bed. After measuring 100 random seedlings from each bed, we note that those receiving 2× Mg are 12 in (30 cm) taller. What can we conclude from this? Not much. This is an obser-

vational study: the study lacked control over which seedlings were in each treatment (1× or 2× Mg). Are growth differences due to the 2× Mg? Possibly, but growth might be affected by seed source, soil conditions, or because the 1× Mg bed was weeded 3 wk after the 2× Mg bed. Seed source, soil conditions, and weeds *confound* the issue of whether or not it is solely the Mg fertilizer. We cannot be certain about the treatment effects, only that 2× Mg is associated with increased growth. However, when talking with other nursery managers, they also report observing that extra Mg increases growth. Then we read about Mg nutrition. Based on our personal observations, discussion with colleagues, and reading papers (see box 1 at end of paper), we think seedling growth benefits from increasing the Mg fertilization rate.

Problem definition. Our problem statement is based on what we have seen and heard: our 1+0 black cherry seedlings may not be getting enough Mg fertilizer.

Stating the hypothesis. From the problem definition, we *could* state the following hypothesis: doubling Mg fertilizer increases growth of 1+0 black cherry. How would we test this hypothesis? As broad as this statement is, we would have to test all 1+0 black cherry seedlings, in all nurseries, on all possible nursery soil types, and all possible seed sources. And we would have to test several growing seasons to make sure weather did not affect the results! Often the hardest part of the research process is defining a concise, achievable objective. Another hypothesis more succinctly states our best guess: doubling the amount of Mg applied to 1+0 black cherry grown in fields 6 and 14 at our nursery increases seedling height. We then formulate the null (no effect) hypothesis: heights of 1+0 black cherry seedlings grown in fields 6 and 14 at our nursery that are fertilized with 1× and 2× Mg *are the same*. The goal of our experiment is to determine which of these statements is true.

Testing. Randomly assigning seedlings to treatments is the most important part of the design of the experiment. Randomization ensures that, other than the treatment, systematic differences between or among groups of seedlings are lacking, allowing us to conclude the 2× Mg treatment is causing the observed result (increases in seedling height) in the experiment (Ganio 1997).

The 1× Mg application serves as our "control" because this is the usual fertilization rate. Without a control for comparison, we cannot be sure our treatment has an effect. *One of the most common mistakes in installing a practical research study is failure to have an adequate control.* Our hypothesis is rather broad in that we think this will work for 1+0 black cherry, implying all possible seed sources of black cherry we might ever grow at the nursery. It is not realistic to include every possible seed source, but at least 3 should be included in the test. If only 1 seed source is used, and it happens to have a genetic trait that yields a growth response to Mg, we might conclude that 2× Mg is beneficial to all seed sources of black cherry when in fact it only favors that particular seed source. As stated in our hypothesis, we also want to check the effects of Mg in the 2 fields (6 and 14) in which we grow black cherry. We assume that soil in field 6 is fairly uniform and soil in field 14 is also fairly uniform, although the soils are not the same.

To determine that the Mg level is affecting growth, we must design the experiment so that the Mg level is not confounded. A location where the entire test plot has similar conditions is needed so that the only variable is the treatment (Columbo 1999). We *could* put 1× Mg on all the black cherry in field 6 and 2× Mg on seedlings in field 14, but this is the incorrect approach because differences in soil conditions between the 2 fields would confound the Mg level. In other words, it would be impossible to determine if growth differences were due to Mg levels or soil conditions. Similarly, if Illinois seed sources were grown in field 6 while field 14 had Iowa seed sources, we would not be able to tell if any growth effects were due to Mg levels or the genetic differences between seed sources. Again, the experiment would be confounded.

To avoid confounding, researchers generally design experiments into blocks determined by the potentially confounding factors. In our test, these factors are the fields and the seed sources. Each field — seed source combination is a block, and each block receives both levels of Mg. Each field (2) — seed source (3) — Mg level (2) combination (we have 12; 2 × 3 × 2 = 12) is a plot. Plots must be replicated and their differences assessed to conclude with certainty whether the treatment and control seedlings are actually different. Growth differences between the 1× and 2× Mg rates must be larger than the growth differences among replicates of the plots for the Mg rates to be considered different. A minimum of 3 replicates of each plot is encouraged; 4 to 6 are better.

If the 12 plots are each replicated 4 times, we have 48 distinct experimental units. The next step is lining these out in the fields. Think in terms of dividing the fields into grids with an equal number of plants in each grid (Columbo 1999). In a perfect study, the seed source — Mg level combinations would be randomly assigned across each field throughout the grid (figure 1). By so doing, portions of several beds would have multiple seed source — Mg level combinations, allowing us to compare seedling growth among seed sources and Mg levels with the same precision. In real life, however, this

Rep 1	Rep 2	Rep 3	Rep 4
1×-A	1×-C	2×-A	2×-B
2×-B	2×-A	1×-C	2×-C
2×-A	1×-B	1×-A	2×-A
1×-C	1×-A	2×-C	1×-B
1×-B	2×-C	2×-B	1×-A
2×-C	2×-B	1×-B	1×-C

Figure 1—*The completely randomized layout of 24 plots that would be installed in each of the 2 fields having dissimilar soils. The 6 combinations of magnesium level (1×, 2×) — seed source (A, B, C) are randomly assigned within each bed (column).*

would make lifting while maintaining seed source integrity difficult. Since soil conditions within each field are similar, and because we are less interested in comparing growth among the seed sources than Mg levels, we can manipulate the design. Although not statistically perfect, we can plant each of the 3 seed sources, 1 seed source per bed, and lay out the remaining 8 experimental units (2 levels of Mg × 4 replicates) in each bed (figure 2). If we plant 100 bed-ft (30.5 m) of each seed source, each experimental unit could be 12.5 ft (3.8 m) long (divide 100 by 8). However, we should avoid using the ends (1st and last 6 ft, 1.8 m) of each bed because of the variability in seedbed density caused by starting and stopping the seed drill. That leaves 88 ft (26.8 m). We should also have a buffer (3 ft, 0.9 m) between treatments to adjust the fertilizer application rate of the equipment. That leaves 67 ft (20.4 m), or about 8 ft (2.4 m) per experimental unit.

After sowing the black cherry, we measure the beds as shown in figure 3. The first 6 ft (1.8 m) is avoided, then an 8-ft-long plot, a 3-ft-long buffer, an 8-ft-long plot (2.4 m, 0.9 m, 2.4 m) and so on is measured. We then randomly assigned the Mg levels to each plot. The process is repeated for each of the remaining 2 seed sources in field 6. We move the equipment to field 14 and repeat the process with the same 3 seed sources, 2 Mg levels, and 4 replicates.

When the Mg is applied, appropriate plots are fertilized with 1× and 2× rates. Buffer strips between plots serve as the transition zone between fertilizer levels. Codes can be used to identify the plots to hide treatment identities and help reduce any bias that might occur during data collection and evaluation (Columbo 1999). *It is essential to make a detailed map of the layout in both fields, add the codes to the map, and store it in a safe place.*

A	B	C	
2×	1×	1×	
1×	2×	2×	Rep 1
1×	2×	2×	
2×	1×	1×	Rep 2
1×	2×	2×	
2×	1×	1×	Rep 3
2×	2×	1×	
1×	1×	2×	Rep 4

Figure 2—*In this layout, magnesium levels (1×, 2×) are randomly replicated 4 times within a bed of each seed source (A, B, C).*

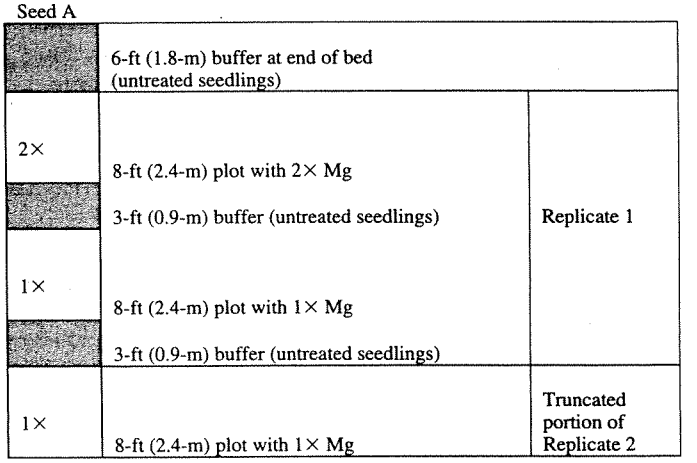

Figure 3—*Spacing and location of the first 3 plots for seed source A shown in figure 2 (modified from Sandquist and others 1981).*

From the time of sowing until the end of the growing season, cultural treatments to the experiment are implemented concurrently. That is, if you add ammonium sulfate, add it to all of the plots at the same application rate. Root prune or apply pesticides to all plots on the same day. The more uniformly cultural practices are applied, the more likely it will be that treatment effects are measured.

Measuring seedlings. At the end of the growing season, seedlings heights must be measured to determine if indeed Mg level affected height growth. In the perfect experiment, the number of seedlings to measure is determined by statistical methods. Often, the perfect statistical answer is tempered by real-world considera-

tions of time and money. Assuming seeds were sown to achieve 5 seedlings/ft² (54/m²), each plot has about 160 trees. Measuring seedlings around the outer edges of the plots should be avoided because of "the edge effect" where seedling growth can be influenced by lower density, higher soil compaction in the wheel ruts, more light, and so on. With 7 rows in a bed, we can avoid measuring seedlings in the 2 outside rows and for at least 1 ft (30 cm) on each end of the plot (figure 4). That leaves about 70 seedlings in the center of each plot to measure for a total of 3360 seedlings in all the plots in both fields (2 Mg levels × 3 seed sources × 4 replicates × 2 fields × 70 seedlings = 3360). That is a lot of seedlings. Sub-sampling each plot by systematically measuring every 5th seedling in each row (5 per row × 3 interior rows = 15 seedlings per plot) would result in measuring a more realistic 720 seedlings. Have the same person collect data from each Mg level at the same time to reduce unwanted variability (Columbo 1999).

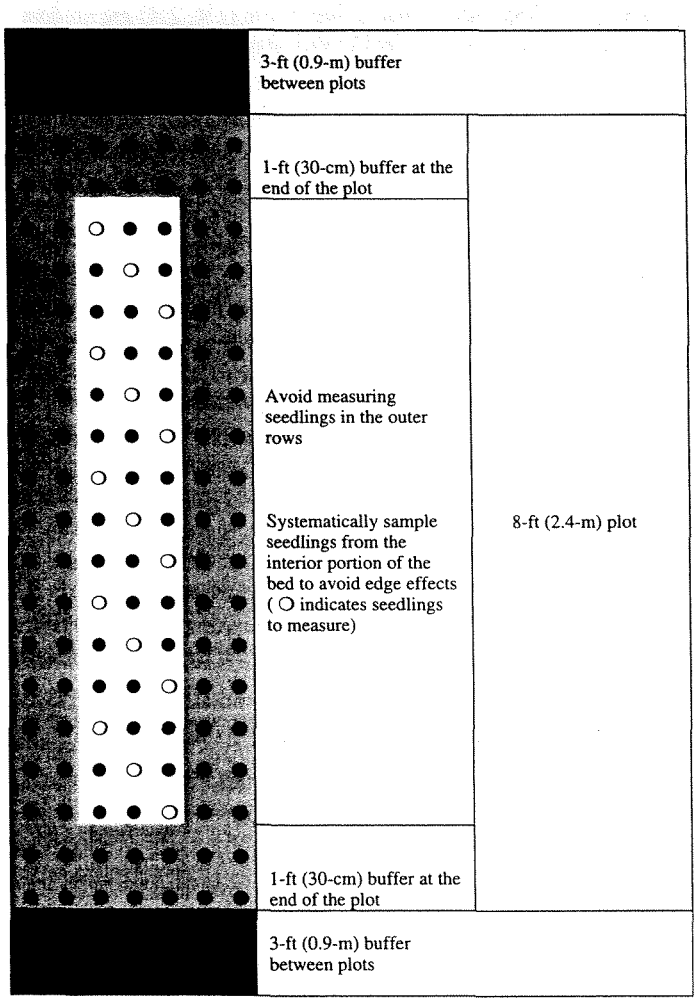

Figure 4—*Measuring seedlings within a plot. To reduce the variability of measured seedlings, avoid measuring seedlings on the edges of the treatment plot. Depending on the number of remaining seedlings in the plot, a systematic sampling of seedlings might be most efficient in terms of labor.*

Statistics: accepting, rejecting, or modifying the hypothesis. *Statistics do 2 things: estimate population* parameters and test hypotheses about those parameters. In our example, we can use statistics to estimate the heights of the seedling populations that received 1× or 2× Mg, and then use those estimates to decide if the null hypothesis is correct (that seedlings have the same height regardless of Mg rate). Statistics do not prove anything: statistics only compute the probability of something happening and leave it to us to draw conclusions from that probability (Freese 1980). Usually the researcher selects the probability to use for testing the null hypothesis, often the 0.05 level of probability. If statistics show that the probability of the null hypothesis occurring is < 0.05, then the difference between treatments has less than 1-in-20 odds of occurring by chance; or stated the other way, in 19 out of 20 instances, the difference can be expected to be due to the treatment. In our experiment, if the probability of the null hypothesis (seedlings in 1× and 2× Mg are the same height) being true is < 0.05, we can infer the alternate hypothesis is true (seedlings in 1× and 2× Mg are not the same height).

Nursery managers have several options for complete analysis of their data. Several powerful statistical software packages are available, and some spreadsheet programs have statistical options. But without an understanding of the process by which the computer is generating the results, it is difficult to know if the answer is correct. An analysis of variance or *t*-test can be done by hand, and hand calculations are explained well in Freese (1980). However, we should not overlook another option. When our experiment is designed well, like the design of our hypothetical Mg experiment, we have a powerful tool to partition the variation in the data to the different sources (fields, seed sources, Mg fertilizer levels) and to evaluate the effects of any of the combinations of these factors. Such an experiment is likely to garner assistance from USDA Forest Service nursery specialists, statisticians, and editors of technology transfer publications who will realize the value of the work and can help you with data analysis.

If a basic evaluation of the data is all that is necessary, an easy way to compare treatments is to compare arithmetic means. Means are the average value of all the measured values in our experimental units. Calculators can generate means, along with the standard deviation and confidence interval. The standard deviation characterizes the dispersion of individuals around the mean. It indicates whether most of the individuals in a population are close to the mean or spread out. If the means are

normally distributed, 67% of all individuals will be within ±1 standard deviation of the mean, 95% will be within ±2 standard deviations, and 99% within ±2.6 standard deviations. A confidence interval provides a range of values inside which the true mean of the population resides. It is an indication of the reliability of the mean. Usually the upper and lower values that define the interval are set at a 95% or 99% level. In other words, if you choose a 95% confidence interval (0.05 level of probability), unless a 1-in-20-chance event has occurred, the population mean is within the specified interval (Freese 1980). A very wide interval indicates a lot of variability in the measurements taken. Collecting more samples from the treatment plots may, or may not, yield a better estimate of the mean, which would be indicated by a narrower confidence interval.

Is it significant? For most growers, the statistical significance of the comparison of means is redùced to 1 simple question: *what is important to me, the grower?* Sometimes treatments can be significantly different from a statistical perspective, but not biologically or economically significant, so not meaningful to us. If 2× Mg-treated black cherry were 2 in (5 cm) taller than the 1× Mg treatment, and that was statistically different, would it be important to you as a grower? What if they were 6 in (15 cm) taller? Or 12 in (30 cm) taller? What if the treatment indeed made them taller, but less sturdy? Or if the treatment increased height but made them more susceptible to insects? As growers, we must interpret the statistical analysis of our data from both the qualitative and quantitative aspects.

Summary

Define your problem and subsequent hypothesis concisely, with very specific objectives of what you want to evaluate. Use blocking to eliminate confounding. Randomly assign seedlings to treatments. Include a control treatment. Treat all seedlings the same, except for the treatment itself, to reduce the chance of confounding. Although powerful statistical packages can be useful, for most growers, a comparison of means between or among treatment populations is probably sufficient enough to determine whether or not the treatment is biologically and economically significant. Growers with well-planned experiments should consider seeking assistance with statistics. Growers should share their results by publishing.

Address correspondence to: Kas Dumroese, USDA Forest Service Southern Research Station, 1221 South Main Street, Moscow, ID 83843; **e-mail:** < kdumroese@fs.fed.us >.

Box 1—*Reading scientific papers.*

Armson (1993) points out several things to consider when reading scientific papers: (1) just because a paper appears in a journal that requires peer review, do not assume the information is correct; (2) do not assume that previous research is cited correctly; (3) do not jump to conclusions. If you only read the abstract or conclusions with the purpose of deciding whether or not the authors agree with your point of view, bias may enter the decision. Papers must be read thoroughly, critically, and with an open mind. Check the references for titles of similar work and read them too.

Acknowledgments

Kas Dumroese was a Research Associate, Forest Research Nursery, Department of Forest Resources, University of Idaho, when this paper was written. The authors thank Dr. Dale Everson, Professor Emeritus, Division of Statistics, University of Idaho and Dr. Andrew Robinson, Assistant Professor, Department of Forest Resources, University of Idaho, for their helpful comments. This is Idaho Forest, Wildlife, and Range Experiment Station Contribution Number 896. The editor thanks Jill J. Williams, statistician (retired), USDA Forest Service Rocky Mountain Research Station, Fort Collins, CO, for her review of this paper.

References

Armson KA. 1993. How to read a scientific paper. Forestry Chronicle 69: 419–420.

Columbo A. 1999. Designing your own research. FarWest Magazine 43(8): 110–111.

Freese F. 1980. Elementary statistical methods for foresters. Washington (DC): USDA Forest Service. Agriculture Handbook 317. 97 p.

Ganio LM. 1997. Designing a nutrient study. In: Haase DL, Rose R, coordinators and editors. Symposium proceedings, forest seedling nutrition from the nursery to the field; 1997 Oct 28–29; Corvallis, Oregon. Corvallis (OR): Oregon State University Nursery Technology Cooperative. p 88–100.

Sandquist RE, Owston PW, McDonald SE. 1981. How to test herbicides at forest tree nurseries. Portland (OR): USDA Forest Service, Pacific Northwest Forest and Range Experiment Station. General Technical Report PNW-127. 24 p.

Stock M. 1985. A practical guide to graduate research. New York (NY): McGraw-Hill Book Company. 168 p. [ISBN 0-07-061583-7]