INSTALLING A PRACTICAL RESEARCH PROJECT AND INTERPRETING RESEARCH RESULTS

R. KASTEN DUMROESE AND DAVID L. WENNY

When this paper was presented, Kas Dumroese was Research Scientist at the University of Idaho Forest Research Nursery in Moscow. At the time of publication, Kas is Plant Physiologist, USDA Forest Service, Southern Research Station, 1221 S. Main Street, Moscow, ID 83843; (208) 883-2324. Dave Wenny is professor and director at the Forest Research Nursery, Department of Forest Resources, University of Idaho, P.O. Box 441137, Moscow, Idaho 83844-1137.

kdumroese@fs.fed.us and dwenny@uidaho.edu

Abstract

We review the basic concepts of science and research and the scientific process. Using an example from a bareroot nursery, we show how a practical research project can be done at any type of nursery, meshing sound statistical principles with limitations of busy nursery managers.

Key Words

Experimental design, research, bareroot nursery, container nursery, statistics

Most growers already do nursery research, whether they realized it or not. Have you ever done the following: 1) contemplate a problem at the nursery; 2) after reading an article or discussion with a colleague, had an idea how that problem might be corrected; 3) put in trials to test your guess; 4) decide if your idea solved the problem? If so, you have done scientific research following the research process. Depending on how well the research is done, the research process can provide accurate and useful information, or it can be a waste of time yielding papers and conclusions that mean nothing. Our objective is to help growers design projects that yield meaningful results. The bonus is that if you can design a good experiment, you can also tell if published research results were generated through a good experiment and 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 distinct steps (Table 1).

Table 1. Steps in the scientific method.

Steps	Example
Natural phenomena are observed	Ships sailing from port gradually disappear from sight, the tops of the masts being the last part seen.
The problem is defined	If the world is flat, why do ships gradually disappear from bottom to top?
A hypothesis (a guess) is made	The world is not flat, but round.
Test	Sail west from port and see if you return to where you started without falling off the end.
Theory	The world is round.

Theories are general explanations for natural events, useful to understand, predict, and control natural phenomena. For installing practical research projects at our nurseries, we're probably less concerned with developing broad, sweeping theories of the universe, but rather, for example, whether or not it's cost effective, in terms of improved growth, to double the amount of magnesium (Mg) we apply to 1+0 black cherry. The research process we generally follow looks like this: 1) *Observation:* practical experience, literature review, conversations with other nursery managers; 2) *Problem definition:* specific question(s) asked; 3) *Hypothesis generated-* objectives defined and methods selected for testing the hypothesis; 4) *Testing-* data collected, analyzed, and interpreted; 5) *Hypothesis accepted, rejected, or modified* (Stock 1985). Although this process seems straightforward, and when reading published research it sounds as if the researchers followed these steps systematically, the research process is dynamic and subject to change as more information is discovered.

FOLLOWING THE PROCESS-AN EXAMPLE

It really doesn't matter whether you grow bareroot or container seedlings-experiments are designed in the same manner. To see how an experiment might be designed, let's follow an example of doubling the rate of magnesium (Mg) to improve growth of 1+0 black cherry.

OBSERVATION

When a usually competent technician accidentally applied 2X the normal amount of Mg to a bareroot bed of 1+0 black cherry, those seedlings appeared taller than an adjacent bed. After measuring 100 random seedlings from each bed we noted that those receiving 2X Mg were 12 in (30 cm) taller. What can we conclude from this? Not much. This is an observational study-there was no control over which seedlings were in each treatment (1X or 2X Mg). Were growth differences due to 2X Mg? Maybe, but growth might have been effected by seed source, soil conditions, or because weeders didn't get to the 1X bed for an extra 3 weeks because of wet weather. Seed source, soil conditions, and weeds have **confounded** the issue of whether or not it was solely the Mg fertilizer. We can't tell treatment

effects for sure, all we can say is that 2X Mg was associated with increased growth. However, in talking with other nursery managers, they also report intriguing observations that extra Mg increases growth. After checking issues of Forest Nursery Notes, we find some papers dealing with Mg nutrition. Based on our personal observations, discussion with colleagues, and reading papers (see sidebar), we think seedling growth will benefit from increasing the Mg fertilization rate.

Problem Definition

With what we've seen and heard, our problem statement is: Our 1+0 black cherry seedlings don't get enough Mg fertilizer.

Hypothesis

From this 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. Oh yeah, and we'd have to do it over a couple of different growing seasons to make sure weather didn't affect our results too. Often the hardest part of the research process is defining a concise, achievable objective. Perhaps this hypothesis more succinctly states our . best guess:
Doubling the amount of Mg applied to 1 +0 black cherry grown infields 6 and 14 at the Iowa DNR State Forest Nursery in Ames increases seedling height.
Being good statisticians, we then formulate (why we do this will become apparent later) *a null hypothesis* for testing: *Seedling heights of 1 +0 black cherry grown infields 6 and 14 at the Iowa DNR State Forest Nursery in Ames fertilized with 1 X and 2X rates of Mg* are **the same.**

Testing

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

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, don't assume the information is correct; 2) don't assume that authors citing previous research did so correctly; 3) don't 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 section for titles of similar work and read them too.

From our hypothesis, we want to test if 2X Mg increases seedling height over that of our usual practice (1X Mg). Therefore, we have 2 treatments (1X and 2X Mg). Our 1X treatment serves as our "control" because this is the usual fertilization rate. Without a control to compare we can't 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's not realistic to try 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 2X Mg is beneficial to all seed sources of black cherry when in fact it only favors that particular seed source. In our hypothesis, we also want to check the effects of Mg in the 2 fields we grow black cherry (6 and 14). Let's assume that the soil in field 6 is fairly uniform and the soil in field 14 is also fairly uniform, although they are not necessarily uniform with each other.

In order to determine that it's the Mg level that is affecting growth, we must design our experiment so that factor (Mg level) is represented and not confounded. We need a location where the entire test plot has similar conditions so the only outside variable is in the treatment (Columbo 1999). Okay, how about we put 1X Mg on all the black cherry in field 6 and 2X Mg on seedlings in field 14? Bad idea. Differences in soil conditions between the 2 fields would confound with the Mg level. We won't be able to tell if growth differences are due to Mg level or inherent differences in soil conditions. And, if Illinois seed sources were grown in field 6 while field 14 had Iowa seed sources, we wouldn't be able to tell if any growth affects were due to Mg levels or the inherent genetic differences between black cherry

from Illinois and Iowa. Again, the experiment would be confounded.

To avoid confounding, researchers generally design experiments into blocks. In our test, we could call each seed source a block. Each block would receive both treatments (levels of Mg). We could also call our fields blocks, and each field block would receive each seed source which we just decided would receive both levels of Mg. Each field (2) - seed source (3) - Mg level (2) combination (there are 12; 2 x 3 x 2 = 12) is a plot. Plots require replication so we can conclude with certainty whether the treatments are actually different. Growth differences between the 1X and 2X Mg rates should be larger than the growth differences between replicates of the experimental units in order for the Mg rates to be considered different. A minimum of 3 replicates of each plot is encouraged--4 to 6 is better.

Okay, we have 12 plots replicated 4 times-48 distinct experimental units. Our next step is lining them out in the fields. Think in terms of dividing the field into a grid with an equal number of plants in each grid (Columbo 1999). In a perfect world with the perfect study, seed sources and Mg levels would be randomly assigned across each field (throughout the grid; Figure 1). Portions of several beds would have multiple seed sources and Mg levels, allowing us to compare seedling growth among seed sources and magnesium levels with the same precision. In real life, this would make lifting and maintaining lot integrity difficult, and cause employees to grumble. We assumed earlier that soil conditions within each field are similar, and because we are less interested in comparing growth among the seed sources and wish to focus mainly on magnesium levels, we can manipulate the design. Let's assume we plan to plant 100 bed ft (30.5 m) of 3 sources of black cherry. Although not statistically perfect, we can lay our 8 experimental units (2 levels of Mg, 4 replicates) in each seed source (Figure 2). If we divide 100 by 8, each unit could be 12.5 ft (3.8 m) long. However,

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

Figure 1. A completely randomized layout of plots. The 6 combinations of magnesium levels (1X, 2X) and seed source (A, B, C) are randomly assigned within each bed (column).

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

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

we want to avoid using the first and last 6 ft (1.8 m) of the bed because of the variability in seedbed density caused by starting and stopping the seed drill. That leaves us 88 ft (26.8 m). We should also have a buffer (3 ft [0.9 m]) between treatments to adjust the fertilizer rate of our equipment. That leaves us 67 ft (20.4 m), or about 8 ft (2.4 m) per experimental unit. After drilling the black cherry, measure the bed 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. Our Mg levels are then randomly assigned to each plot (Figure 3). We repeat the process for each of

the remaining 2 seed sources in field 6. Moving our equipment over to field 14, we repeat the process with the same 3 seed sources, 2 Mg levels and 4 replicates. **Make a detailed map and stop it in a safe place.**

When the Mg is applied, appropriate plots are fertilized with 1X and 2X rates. Buffer strips between plots serve as the transition zone between fertilizer levels. It's often best to use a code to identify plots, hiding the treatment identity helps reduce any bias that might occur during data collection and evaluation (Columbo 1999). **Add the codes to your map and store them in a safe place.**

A		
	6-ft (1.8-m) buffer at end of bed (untreated seedlings)	
2X	8-ft (2.4-m) plot with 2X Mg	Replicate 1
	3-ft (0.9-m) buffer (untreated seedlings)	
1X	8-ft (2.4-m) plot with 1X Mg	
	3-ft (0.9-m) buffer (untreated seedlings)	
1X	8-ft (2.4-m) plot with 1X Mg	Truncated portion of replicate 2

Modified from Sandquist and others (1981).

Figure 3. Spacing and location of the first 3 plots for Seed Source A shown in Figure 2.

From the time of drilling until the end of the growing season, implement cultural treatments to the experiment concurrently. That is, if you top dress with ammonium sulfate, top dress all of the plots with the same fertilizer rate. Root prune or apply pesticides to all plots on the same day. The more uniform cultural practices are applied to test plots, the more likely we'll be able to measure treatment effects-what we really want to know.

Measuring Seedlings

We hypothesized that 2X Mg would increase height growth. At the end of the growing season we'll need to measure seedlings to see if indeed this is true. In a perfect world, the number of seedlings to measure is determined by statistical methods. Often, the perfect statistical answer must be tempered by real-world considerations of time and money. Assuming seeds were drilled to achieve 5 seedlings per square foot ($55/m^2$), each plot has about 160 trees. We want to avoid measuring seedlings around the outer edges 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. So if we planted on 6 in (15 cm) rows in a 4-ft-wide (1.2-m-wide) bed we have 7 rows. We should 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 for measurement. If we measure every seedling in every plot in both fields, we'll have to measure 3360 seedlings (2 Mg levels x 3 seed sources x 4 replicates x 2 fields x 70 seedlings = 3360). That's a lot of seedlings. We could sub-sample each plot by systematically measuring every fifth seedling we come to in each row (5 per row, 3 interior rows; 15 seedlings per plot). We would then measure a more realistic 720 seedlings. Have the same person collect data from each treatment at the same time to reduce adding unwanted variability (Columbo 1999).
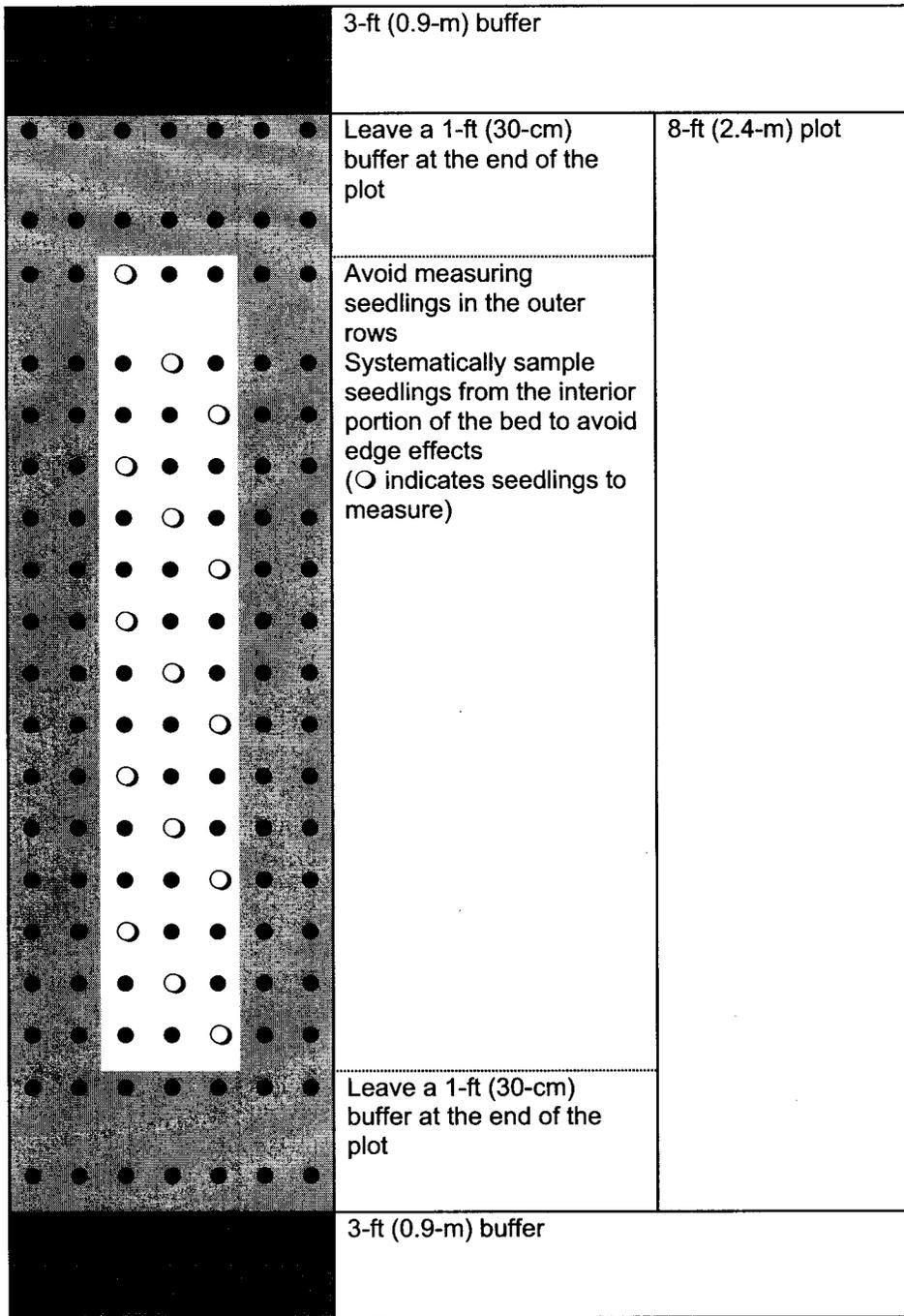
ACCEPTING, REJECTING, OR MODIFYING THE HYPOTHESIS

Statistics-Striking Fear Into The Hearts of Many

For many of us, statistics seems like a foreign language. Looking at a statistics textbook becomes a substitute for insomnia drugs. However, remember we use statistics to do 2 things: 1) estimate population parameters; and 2) test hypotheses about those parameters. In other words and for our example, we use statistics to estimate the heights of our seedling populations that received 1X or 2X Mg, and then use those estimates to see if our null hypothesis is correct (all seedlings had the same height regardless of Mg rate).

Remember that statistics don't prove anything-statistics only deal with probabilities, not immutable absolutes (Freese 1980). In other words, statistics only compute the probability of something happening and leave it to the researcher to draw conclusions from that probability. Usually in designing an experiment, a researcher will select a probability to use for testing the null hypothesis (for example, 0.05 or 0.01). Often researchers use the 0.05 level of probability. If statistics show that the probability of the null hypothesis occurring is < 0.05, we infer that differences seen between treatments have a less than one-in-twenty chance of occurring by chance. In other words, 19 out of 20 times the difference is due to the treatment. In our experiment, if we ran statistics and found that the probability of the null hypothesis being true (seedlings have the same height in 1X and 2X Mg treatments) was < 0.05, we could infer this is not true (reject the hypothesis; seedlings in 1X are not the same height as seedlings in 2X) because the heights are different 19 out of 20 times (the probability of the heights being the same is less than one-in-twenty).

Okay, that's nice, but how would most nursery managers that don't have access to powerful statistical analysis software like Systat or SAS analyze data? Both an analysis of variance or t-test can be done by hand, and hand calculations are explained well in Freese (1980). Some spreadsheet programs like Excel have built-in statistical packages, but if you don't understand the process by which the computer is generating the results, how will you know if the answer is correct?

This hypothetical design allows us to partition variation to different sources (fields, seed sources, Mg fertilizer levels). It is a powerful tool for evaluating treatment effects. If complete data analysis cannot be done by hand or computer program, and if a good statistician isn't available to

		8-ft (2.4-m) plot
	3-ft (0.9-m) buffer	
	Leave a 1-ft (30-cm) buffer at the end of the plot	
	Avoid measuring seedlings in the outer rows Systematically sample seedlings from the interior portion of the bed to avoid edge effects (O indicates seedlings to measure)	
	Leave a 1-ft (30-cm) buffer at the end of the plot	
	3-ft (0.9-m) buffer	

Figure 4. 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.

help with complete analysis, some basic comparisons can still be made.

Means-a Basic Comparison

An easy way to compare treatments is to compare arithmetic means. Means are the average value of all the measured values in our experimental unit, or plot. Spreadsheets and calculators can give us the mean, along with a standard deviation and confidence intervals.

Standard deviation characterizes the dispersion of individuals around the mean. It provides some idea 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, and 95% will be within 2 standard deviations, and 99% within 2.6 standard deviations.

Confidence intervals (CI) provide a range of values inside which the true mean of the population resides. CI provide an indication of the reliability of the mean. Usually the limits are set at a 95% or 99% level. If you choose a 95% CI (0.05 level of probability), the result the spreadsheet gives you tells you that unless a one-in-twenty chance has occurred in sampling, the population mean is somewhere within the CI (Freese 1980). Very wide CI indicate a lot of variability. Collecting more samples from the treatment plots may, or may not, yield a better estimate of the mean (narrower CI).

IS IT SIGNIFICANT?

For most growers, the whole idea of statistical significance and comparing means boils down to 1 simple question: *What's important to me, the grower?* Sometimes treatments can be significantly different, but not biologically or economically significant. If 2X-treated black cherry were 2 in (5 cm) taller than the 1X, 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 increased height but made them more susceptible to insects? As growers we might be interested in both the qualitative and quantitative aspects of our results.

SUMMARY

Define your problem, and subsequent hypothesis, concisely-have 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 and should become the new nursery standard. Growers should share their results by publishing.

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. Agricultural 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 0-07-061583-7J