WASHINGTON STATE UNIVERSITY EXTENSION

ON-FARM VINEYARD TRIALS: A GROWER'S GUIDE

WSU PEER REVIEWED

EM098E

ON-FARM VINEYARD TRIALS: A GROWER'S GUIDE

By
Dr. Hemant Gohil, formerly, Technology Transfer Specialist, WSU Viticulture & Enology Program, Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, WA. **Dr. Markus Keller**, Professor, Department of Horticulture, Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, WA. **Dr. Michelle M. Moyer**, Assistant Professor / Viticulture Extension Specialist, Department of Horticulture, Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, WA

Abstract

On-farm research offers many opportunities to understand the effectiveness of various management practices and products. However, how these trials are designed can alter the observed results. This guide summarizes the concepts of experimental design in vineyards and how those concepts are important in conducting field trials and understanding their results. It also describes specific examples of trial design and explains how to collect data relevant to vineyard research. Finally, the guide explains simple statistical tests that are used to help interpret results.

Table of Contents

On-Farm Vineyard Trials: A Grower's Guide

Introduction

On-farm trials can be an effective tool for evaluating new practices or products in the vineyard (Veseth et al. 1999). While universities and government laboratories around the country provide information on viticulture research trials, some of this information may not be readily applicable to varying local conditions (Wuest et al. 1995). Designing on-farm trials does not have to be complicated; however, basic components need to be included in every trial in order to properly interpret the results (Koenig et al. 2000; Ketterings et al. 2012).

This guide provides an overview of the basic concepts and principles that are important in conducting research on farms. These principles are also applicable to the successful implementation and interpretation of vineyard trials.

This guide also provides information on the kinds of data to collect and when and how to collect them, since data collection and interpretation are integral to drawing informed conclusions on vineyard trial results. This discussion includes examples using specific viticulture on-farm trials and highlights how to organize and interpret trial results. Finally, the guide provides information on how to make decisions based on trial results.

Basic Components of an Experiment

Most experiments have five basic components:

1. A ***control***;
2. One or more ***treatment(s)***;
3. ***Replication*** of the control and treatments;
4. ***Randomization*** of the control, the treatments, and their associated replications; and
5. ***Repetition*** of the experiment.

These components are discussed in detail below.

Control. A "control" is the practice or product that is in current use on the farm. A control provides a baseline or benchmark against which to compare a new practice or product. All experiments need a control. In some cases, a single control is sufficient for multiple treatments; in other cases, more than one control is necessary. Either way, the control needs to be fully integrated in each experiment. It should be applied within the same area that the experiment is conducted and applied at the same time as the experiment.

Comparing trial results with last year's vineyard performance or with a neighboring vineyard will result in misleading conclusions.

There are two main types of controls: 1) A positive control, which in the case of on-farm trials is generally the use of a standard practice or product and 2) a negative control, which is generally not doing or applying anything at all. For example, when testing a new fertilizer, a section of the vineyard should be treated with the standard fertilizer as a positive control. An alternative testing method would be not to treat a section of the vineyard at all even if you normally would, which would be considered a negative control.

If space and time allow, both types of controls can be used in an experiment. The use of controls allows for more definitive comparisons to know if the results from using the new practice or product were really due to that practice or product rather than due to other external factors (e.g., differences in weather or a change in irrigation practices).

Treatment. A treatment is a new or different practice or product to be tested. Treatments are used for a variety of things, but in viticulture they are often applied to evaluate new varieties, trellising systems, canopy management practices, irrigation practices, fertilizer regimes, cover crops, pesticides, or the effectiveness of new equipment (for examples see Appendix 1). Experiments can be designed with as few as one treatment or as many as can be physically handled.

However, the more treatments are evaluated, the more input is needed in terms of labor, land, capital, and data collection and evaluation. Sometimes the addition of multiple treatments may also require the addition of multiple controls, thus taking an on-farm trial from manageable to unruly. The most common mistake in many on-farm trials is the desire to evaluate too many treatments in a single experiment.

Replication. Replication of both treatments and controls within an experiment is necessary for appropriate interpretation of the results. Even though a vineyard may be planted to a single clone of a specific grape variety, there can be substantial variation within that vineyard due to differences in soil type or depth, slope, or other site aspects either inherent to the site or constructed, such as the location of vineyard infrastructure.

Consequently, it is important to test both control(s) and treatment(s) in multiple locations within a vineyard to reduce results that might be affected by pre-existing site conditions.

What if the particular vine selected for observation was next to a broken irrigation line? What if that vine has a pest problem or cold injury? What if that vine was planted directly over a shallow layer of rock? By collecting data from multiple units (e.g., rows, vines, shoots, or clusters receiving particular treatment) that are spatially separated across a vineyard (i.e., the units should not be adjacent to each other), one can calculate what the *average* response is and how variable the response is. Each of these multiple units are referred to as a replicate. In other words, a replicate is an independent experimental unit that is treated separately throughout the entire experiment.

Typically, a field trial should have a minimum of three replicates of a given treatment or control; however, more replicates are encouraged. The more variable a vineyard, the more replicates are needed. Applying treatments to a larger area of a vineyard does not compensate for lack of replication (Gates 1991).

Replication can come in multiple forms provided these multiple units are treated independently. These forms include:

1. Applying the treatment(s) and control(s) to different sections within a vineyard.
2. Applying the treatment(s) and control(s) to different rows or portions of rows within a vineyard.
3. Making observations on multiple vines receiving a particular treatment.
4. Making observations on multiple units on a single vine (e.g., different shoots, leaves, or clusters).

In many cases, each experimental treatment and control will have two forms of the above replication:

1. The treatment or control is applied to multiple rows; and
2. multiple vines within each of those rows are assessed.

Randomization. Randomization is the process of assigning control(s) and treatment(s) to different units (e.g., rows, vines, clusters) within the vineyard in an unbiased fashion. If a trial is conducted on rolling terrain with all replicates of a treatment in the same area (e.g., a slope base), then how does one know if the results seen are due to the treatment or due to the terrain?

Like replication, randomizing where treatments and controls are applied in a vineyard helps reduce bias. How the treatment and control replicates are randomized will be based on the uniformity of the vineyard and the chosen experimental design (see *Experimental Designs).*

To randomize a treatment, first identify all the units (vineyard block, row, vine, or cluster/plot) that will receive the treatment (or the control). Then, assign a treatment/control to those units by using something such as a random number generator. For information on manually randomizing treatment location, see *Randomizing Treatments.*

Randomizing Treatments

Step 1: Number small sheets of equal sized and shaped paper with a number for each corresponding treatment and control unit. If you have a treatment and a control, and are replicating each 5 times, you would have 10 sheets of paper, numbered 1 through 10 (a single number per sheet).

Step 2: Place the numbered sheets in a container. Then draw five sheets and place them in a pile for the control and five sheets in a pile for the treatment. For example, if you drew the numbers 4, 7, 2, 5, and 10 for the control pile, this means that you would assign the control to the row/vine/cluster that is equivalent to the 2^{nd}, 4^{th}, 5^{th}, 7^{th}, and 10^{th} unit in your designated area. The remaining sheets are for the treatment, which would be applied to the row/vine/cluster that is equivalent to the 1^{st}, 3^{rd}, 6^{th}, 8^{th}, and 9^{th} units.

Repetition. Repetition of the entire experiment helps in reducing the bias a particular vintage or vineyard location might impart on the trial outcomes. For example, if an irrigation trial is conducted in an unusually cool year, then the results might differ considerably from those observed in a very warm year. Experiments should be repeated in a different location, on a different variety, or in a different year before drawing conclusions.

General Strategies for Conducting On-Farm Trials

The key to successful on-farm trials is to change only one or two variables at a time. For example, when testing a new product (e.g., pesticide, fertilizer) or cultural practice, make sure the only difference between treatments is the actual product or specific practice. The following tips can improve the process of designing, establishing, and conducting an on-farm experiment.

1. *Decide what to test.* When selecting treatments that are intended to test a range of options (e.g., irrigation volumes, pesticide rates, fertilizer rates), avoid looking at very narrow ranges (e.g., 10 lbs of nitrogen versus 15 lbs of nitrogen per acre). Starting with a larger range (within legal limits, such as 10 lbs of nitrogen versus 40 lbs of nitrogen per acre) will help to determine where future focus should be placed.

   If treatment options are too narrow, you may not see a noticeable difference. This lack of difference may cause you to assume that the treatment at any rate has no effect.

While that may be true for the range of treatments you selected, it might not be true for a larger range of the same treatment. If you are unsure about what ranges might be appropriate for different treatments, consult your regional extension specialist or extension publications.

2. *Mark the experimental units and replicates.* Data vines (vines used for sampling/data collection) and rows should be marked for easy identification. This will save time in locating a specific treatment or control. It will also reduce confusion as to where a particular treatment should be applied and reduce the risk of inadvertently applying management practices that might interfere with the outcome of the trial.

3. *Time and plan data collection appropriately.* The time at which data are collected or observations are made will depend on the trial. Typically, observations should be made at key stages in vine development as they relate to the purpose of the experiment. For example, if you are interested in how a product or practice affects canopy density, observations should be made prior to applying a treatment and again after the application (typically after the canopy is fully established) to determine changes in growth.

While it is helpful to make observations and collect data multiple times over a growing season, a single appropriately timed collection may be all that is needed. For example, rather than taking repeated canopy measurements during the growing season, measuring pruning weights and counting canes in the winter can give a good indication of the effects of a treatment on vigor and canopy density.

If you are interested in how a product or practice affects fruit composition, then cluster, berry, or juice samples can be taken at or just prior to harvest. If, by contrast, you want to know the influence of a product or practice on the rate of berry growth or fruit ripening, then repeated sampling, starting just prior to véraison (i.e., during the lag phase of berry growth), should be conducted.

4. *Consider the types of data to collect.* The types of data to be collected should be considered *before* the experiment is underway. There are many types of data that can be collected, including:

a. The timing of vine development (i.e., phenology);

b. vine growth (e.g., shoot length, shoot number, lateral shoot number, canopy density);

c. pruning weight (i.e., total weight of one-year-old canes pruned off in winter) and cane number;

d. pest damage on the fruit or foliage;

e. weed growth or regrowth;

f. total yield and/or yield components (e.g., berry weight, berries per cluster, cluster weight, clusters per shoot or vine); and

g. fruit composition (e.g., sugar, titratable acidity, pH, color, tannins).

h. Additional data that are often helpful include:

i. vineyard design (e.g., planting year and distances, row direction, slope, aspect, trellis type, pruning level);

ii. weather data (e.g., temperature, rainfall, relative humidity, wind speed);

iii. soil data (e.g., type, texture, rooting depth, organic matter, pH);

iv. other cultural practices that are being applied in the vineyard that are not directly related to the trial (e.g., irrigation amount and timing, pest control, canopy management, cluster thinning); and

v. who is conducting those other vineyard practices and when. This information is useful in determining the possible causes of a particular result, especially if the result is unusual or varies between years.

5. *Determine your sampling strategy.* When making observations or measurements in the vineyard, there are several factors to consider.

a. Sampling should be performed randomly to avoid favoring one portion of a vineyard area, vine, shoot, or cluster over another. For example, if you only sample fruit from sun-exposed clusters in one of the treatments while collecting samples from both sun-exposed and shaded clusters in another treatment, then the samples may not give an accurate estimate of fruit composition. If you are sampling berries instead of clusters, then collect berries from the top, middle, and bottom as well as from the front and back of the cluster (Figure 1). Avoiding sampling from the back of the cluster will not provide a representative estimate of fruit composition.

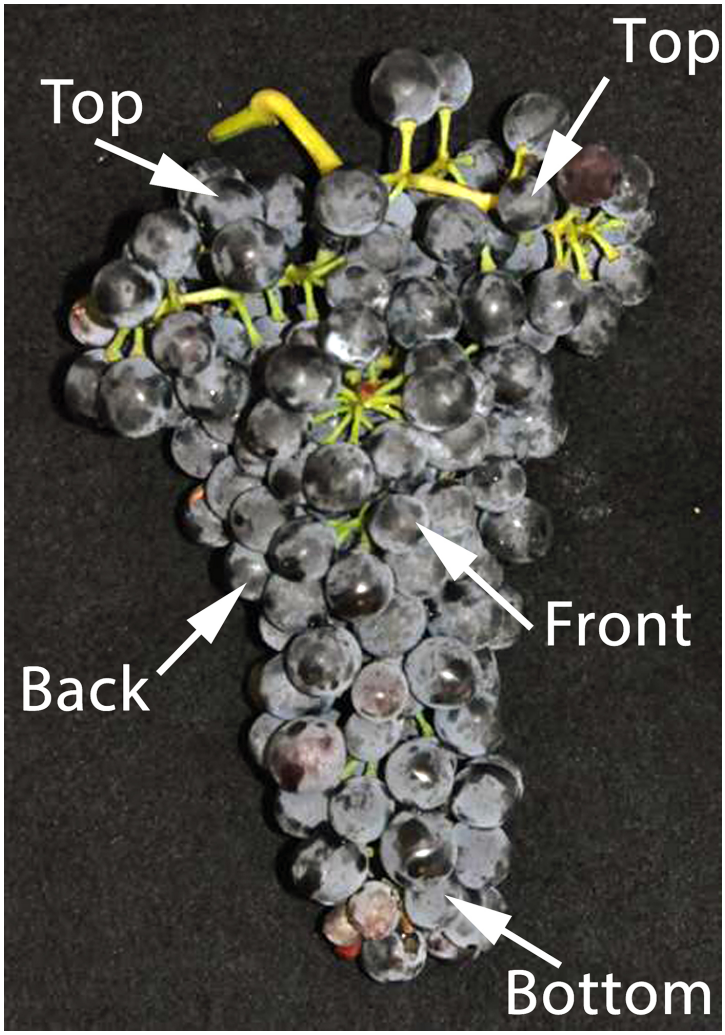

Figure 1. Different portions of a cluster from which representative berry samples should be collected. Photo by Hemant Gohil.

Similarly, on shoots with two clusters, samples should be collected from both clusters (Wolpert et al. 1980). If a vineyard is not uniform, then sampling equal numbers of berries or clusters from each of a set number of vines will overestimate the true maturity status of the vineyard because the fruit of lightly cropped vines tends to be more mature compared with those of heavily cropped vines (Rankine et al. 1962). Consequently, more berries or clusters should be sampled proportionately from high-yielding vines than from low-yielding vines to compensate for the differences in crop loads among vines. For example, if the crop on one sampling vine is twice as much as on another sampling vine, twice as many berries should be collected from the former.

Samples should also be collected from multiple locations within a given experimental unit (i.e., multiple clusters per vine or multiple vines per row). Multiple samples within a treatment replicate can be used to create the average response for the treatment replicate. Avoid sampling from areas such as the end of rows or rows that are on the vineyard border. In many cases, these are sources of large variability due to the vines having access to a greater soil volume, frequent vehicle or equipment movement, or inconsistent application of irrigation or pesticides.

b. If observations need to be made multiple times on the same cluster/vine, it is important to have the same person make those observations because people vary in their visual assessment ability. For example, if sunburn over time is being rated, one person may view a 50% cluster surface area sunburned whereas another may rate it as 65%. In this case, the absolute level of sunburn (i.e., 50% or 65%) is not as important as the relative difference between treatments (i.e., 0% vs. 50%). Keeping the same person rating each replicate or the whole trial helps reduce unwanted variability in the data and will help capture the relative, rather than absolute, difference.

6. *Determine what to sample.* All measurements within each experimental unit should be collected separately.

Phenology. Depending on the size of a trial, use a minimum of three data (observational) vines per treatment replicate. Make visual assessments at key stages of development, such as budbreak, bloom, véraison, and harvest (Figure 2). Rating can be performed using a standard numbering system for each stage, such as the modified Eichhorn-Lorenz (EL) system (Coombe 1995) or the Bayer, BASF, Ciba-Geigy, and Hoechst (BBCH) system (Meier 2001). Always note the date at which a specific stage of interest is reached. The vines may need to be inspected frequently (sometimes multiple times per week) near those key stages to determine their precise timing.

Pest damage. Pest damage includes damage inflicted by any vineyard pest, including mammals, birds, insects, mites, fungi, bacteria, viruses, and other organisms. Depending on replicate size, mark a minimum of 15 clusters or 15 leaves per replicate for observation. In some cases, repeated observations on the same clusters or leaves are necessary to determine the progression of damage over time. In other cases, randomly sampling the clusters or leaves per replicate is appropriate; however, sampling from more than 15 clusters or leaves is recommended in this scenario.

Figure 2. Examples of phenological stages of grapevines in terms of the BBCH scale; BBCH 07 is when green tissue is just visible; BBCH 55 is when the rachis of the flower cluster (inflorescence) begins to elongate; BBCH 69 or 100% bloom is when all the flowers are open; BBCH 83 or 50% véraison is when 50% of the berries are changing color; BBCH 89 is when fruit is ready for harvest. Photos by Michelle Moyer and Hemant Gohil.

Pest damage can be recorded as *incidence*, which is presence or absence of the damage, or as *severity*, which is typically the percentage of surface area on a leaf or cluster affected. In the case of severity, record observed percentage area rather than using a rating scale for ease of future calculations (e.g., 50% of the cluster is covered in mildew, or 30% of the leaf is mottled due to leafhopper feeding). Incidence data allows you to determine how widespread a particular type of pest damage is in a vineyard; severity data allows you to determine how bad that damage is on individual clusters or leaves.

Yield. For information on yield estimation, see the WSU Extension Publication *Vineyard Yield Estimation* (Komm and Moyer 2015). Yield is often measured as pounds of fruit per vine. From the known planting density, this measurement can be converted to estimated tons per acre.

Yield is made up of separate components that, multiplied together, give the total yield of a vine or vineyard (Keller 2015). If yield components are to be evaluated, count and weigh all clusters as they are removed from a vine. This practice will permit the estimation of average cluster weight from the cluster count and the yield per vine. Clusters per shoot (a measure of bud fruitfulness) can be counted at the time of harvest or estimated later by counting the number of canes during pruning weight collection in winter.

Average berry weight is determined by weighing the samples collected for fruit analysis and dividing by berry number (i.e., if 100 berries are being used for analysis, weigh all 100 berries as a whole, and then divide that weight by 100 for average berry weight).

Harvest all of the data vines, preferably individually, as close to commercial harvest of the vineyard as possible.

Sampling/harvest should be finished on the same day as time and weather can impact the physical or physiological condition of the berries and potentially alter results. Advanced preparation, such as labor requirements, labeled bags, charged batteries for weigh scales, and pre-printed data sheets, can save time.

Because grapevine yield is formed over two growing seasons, changing a cultural practice or product in one year may influence the crop of the following year. Consequently, it is good practice to assess treatment effects on yield over at least two years.

Fruit composition. A minimum sample size of 100 berries per treatment replicate is desirable for fruit composition analysis (Nail 2012). Whole cluster samples can also be used, and depending on the size of a trial, three to ten clusters should be sampled from each treatment replicate. Cluster samples often provide composition data that are closer to those seen in a commercial harvest since they represent all the berry positions within clusters (Hellman 2004).

It is also easier to sample clusters compared to sampling individual berries. However, the variability in fruit composition typically decreases as the grapes ripen and more and more of the berries approach their sugar maximum and acid minimum (Rankine et al. 1962; Calderon-Orellana et al. 2014).

Regardless of sampling strategy, samples should always be collected at the same time of day and processed as soon as possible after collection, preferably on the same day. To minimize changes in composition before processing, samples should be collected into a cooler and stored on ice or in a refrigerator. Store samples in a freezer (0ºF or colder) if you are planning to process them after more than a few days. For specific details on how to conduct the composition tests outlined below, see Iland et al. (2013).

Soluble solids. In ripening grapes, soluble solids is a reliable and simple measure of fruit sugar concentration and, thus, of fruit maturity. It is usually measured using a hand-held, temperature-compensated refractometer and expressed in °Brix (equivalent to % w/w or g/100 g of liquid). Grape berries typically undergo the color change associated with véraison once they have reached about 9–10 °Brix, and, depending on variety, they continue to accumulate sugar up to 23–25 °Brix. Higher soluble solids may be achieved by berry dehydration (Keller 2015).

pH. The pH is a measure of the total free hydrogen ions available in a solution ($pH = -\log_{10} [H^+]$, where […] denotes concentration in mol/L); it provides information on fruit acidity (and hence maturity) and the favorability of the juice for the growth of microorganisms. The pH of pre-véraison grape berries is below 3.0 but increases during ripening. The pH of mature fruit is around 3.5 and can sometimes exceed 4.0 in overripe fruit (Keller 2015).

The typical pH range for grape juice is 3.4 to 3.8 for red wine grapes and 3.2 to 3.6 for white wine grapes. However, this can change based on wine style preference. The pH is measured using a pH meter or litmus paper. A pH meter is the preferred and more accurate method. To avoid inaccurate readings, make sure the temperature of the grape juice and the buffers used for calibration is the same (Iland et al. 2013).

Note that due to the log scale, it is mathematically meaningless to directly calculate pH averages. The error is small for pH values that are similar but gets larger if the samples are more variable. For example, the average pH of two juices with pH 3.0 and pH 4.0 is 3.26 and not 3.5.

For maximum accuracy, individual pH values should be converted to H^+ concentrations ($[H^+] = 10^{-pH}$) and the pH mean recalculated from the mean of $[H^+]$.

In our example, $[H^+]$ for pH 3.0 is $10^{-3.0}$, or 0.001 mol/L, and pH 4.0 is $10^{-4.0}$, or 0.0001 mol/L. The average of 0.001 and 0.0001 is 0.00055. To convert that back to pH, you take the negative log of 0.00055, or -log(0.00055) = pH = 3.26.

Titratable acidity (TA). The TA is an approximate marker of the concentration of organic acids (mainly tartaric and malic acids) and measures all the hydrogen ions (free and bound) in a solution. The TA of grape juice at maturity ranges from less than 4 g/L to over 10 g/L depending on variety and environmental conditions. The typical TA range for white wine grape juice is 4 to 10 g/L; the typical TA range for red wine grape juice is 4 to 8 g/L. These ranges can change, however, based on wine style preference.

The TA is expressed in the US as tartaric acid equivalents. Measuring TA by hand typically requires the use of a pH meter or a pH indicator dye, such as phenolphthalein. The most common method in the US is to use a pH meter and titrate to a pH endpoint of 8.2 using sodium hydroxide (Iland et al. 2013).

Pruning weight. An estimate of vine vigor can be obtained by measuring the amount of seasonal vegetative growth. This can be done by collecting pruning weights in winter. During winter pruning, collect all of the cut canes (but not older wood that has been pruned off) and weigh them on a per-vine basis. Pruning weights are easily measured using a spring scale (e.g., a fish weigh scale). Avoid using luggage scales since they lack precision. Note, however, that pruning weight *per se* is not a measure of vigor. A vine with few but vigorous shoots can have the same pruning weight as a vine with many non-vigorous shoots.

Shoot and cane counts can occur at any time prior to collecting pruning weight, including during the growing season. However, it is often easier to collect shoot count numbers just prior to pruning. If canes are counted during pruning, then the average cane weight can be used as a fairly reliable marker of vigor.

You may also calculate the yield-to-pruning weight ratio as a measure of crop load or vine balance (Smart et al. 1990; Skinkis 2013; Komm and Moyer 2015).

A ratio of 5 to 10 lbs of fruit for every pound of pruning weight is often used as an indicator of so-called balanced vines capable of producing high-quality fruit. Vines with ratios below five are said to be under-cropped, whereas vines with ratios above ten are said to be over-cropped.

Weather data. In Washington, area weather data can be downloaded from Washington State University's AgWeatherNet (http://weather.wsu.edu). AgWeatherNet weather stations record information such as air and soil temperatures, relative humidity, precipitation, and wind speed. Weather data can be accessed on 15 minute, hourly, and daily bases. Historical data are also accessible.

General Experimental Design

Vineyard uniformity influences experimental design. Since most modern vineyards are planted to clonal material (i.e., all vines in a block are genetically identical), uniformity typically refers to the variation due to differences in soil depth, texture, drainage, pH, as well as site slope and aspect, in addition to the status of vine health or canopy structure. In many vineyards, the canopy is so variable on individual vines that cluster and berry weight or fruit composition vary more between clusters on the same vine than between vines (Pagay and Cheng 2010; Calderon-Orellana et al. 2014).

Depending on the experimental purpose, all of the vines within a given field trial should be of equivalent vigor and health status. However, there are very few vineyards that are completely uniform (Figure 3). In situations where the lack of uniformity is due to factors that are difficult to change (e.g., slope or variable soil depth), the design of the experiment needs to account for that variability, otherwise results will be difficult to interpret.

The size of the vineyard that will be used for the trial can also influence the experimental design. If the area of the vineyard available for experimentation is small, experimental units might consist of a few vines or partial rows. If the area available is larger, then the experimental units might be whole rows or multiple rows. If space is not limited, experimental units can be entire vineyards, but one must consider the total space needed to properly replicate those units for each treatment and control.

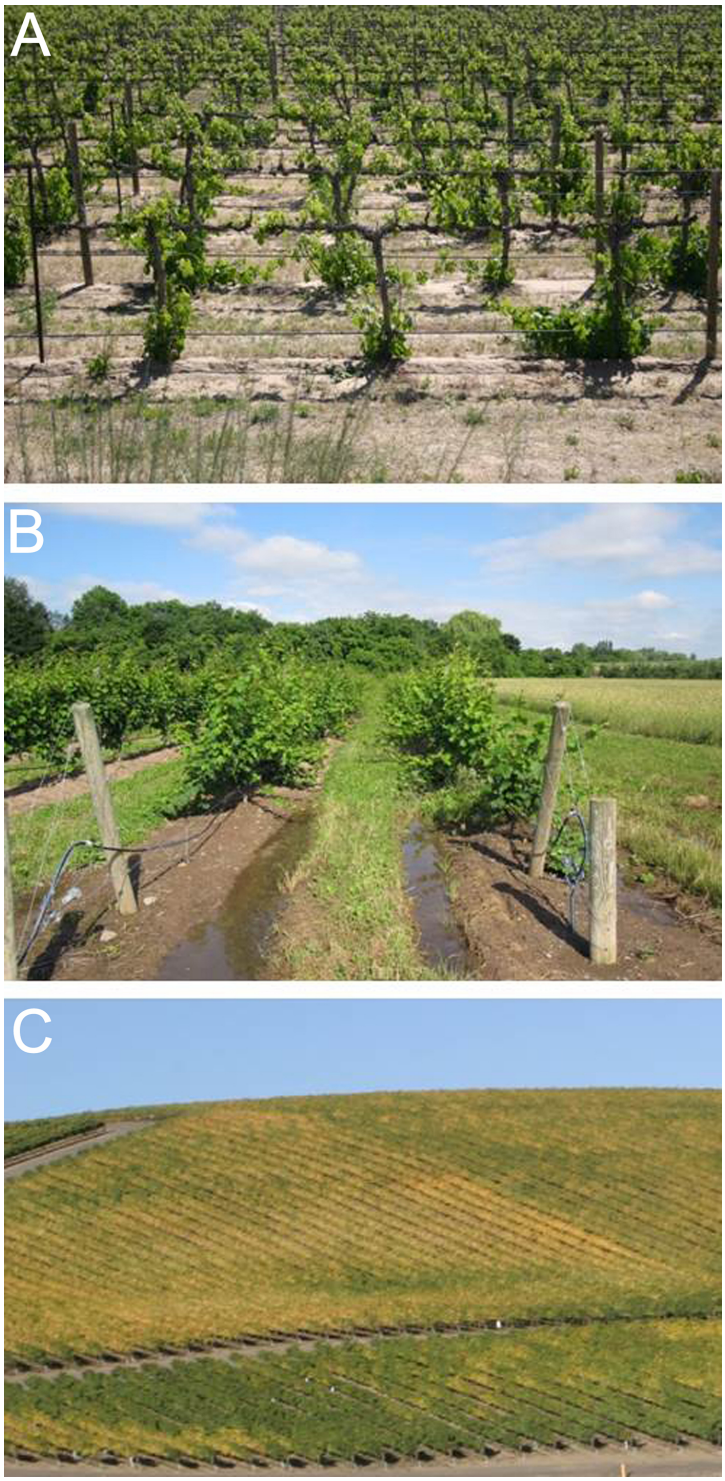

Figure 3. Examples of non-uniform vineyards: A) Cold injury can damage vascular tissues of vines which can result in uneven budbreak, stunted shoot growth, and suckers growing from the base of the trunk (Moyer et al. 2011). B) Poor drainage can lead to differences in vine health and vigor. C) Differences in soil depth or texture across slopes can result in uneven growth and fruit development across a vineyard. Photos by Michelle Moyer and Markus Keller.

Depending on the size and uniformity of the vineyard as well as the experiment objectives, there are several experimental designs available for on-farm trials. The most common designs in agricultural experiments are the Completely Randomized Design (CRD) and Randomized Complete Block (RCB) design (Little and Hills 1975), which are described below.

Completely Randomized Design (CRD). If the vineyard is uniform, then a CRD is the preferred design choice. In a CRD, the different replicates of the treatment(s) and control(s) are assigned in a completely random fashion throughout the vineyard. Figure 4 shows a CRD with one treatment and one control in both small and large area vineyards.

Randomized Complete Block (RCB) Design. If there is predictable variability within the vineyard, such as a noticeable slope or known soil differences, then the RCB should be used (Figure 5). It is the most widely used experimental design in agricultural research because of its ability to address site variability. In a RCB, unlike a CRD, all *zones* (i.e., areas of known differences, also referred to as ***blocks.***

We will use the term "zone" to avoid confusion with the term "block" commonly used to describe vineyard planting design). Each zone must have each treatment/control replicated at least once and each treatment/control must be assigned to a random location in that zone.

Two examples of the RCB design, grouped into different zones due to a slope, are shown in Figure 5. For example, if you are conducting a fertilizer trial in a vineyard that has an area with high water-holding capacity and an area of low water-holding capacity, then divide the vineyard in half and make sure the *experimental replicates* are represented in both of these zones in the vineyard.

Common On-Farm Trials and Their Design

For the scope of this guide, a few select examples of common on-farm trials are described below. Additional examples, along with the types of data collection required and general challenges in conducting the trials, are listed in Appendix 1.

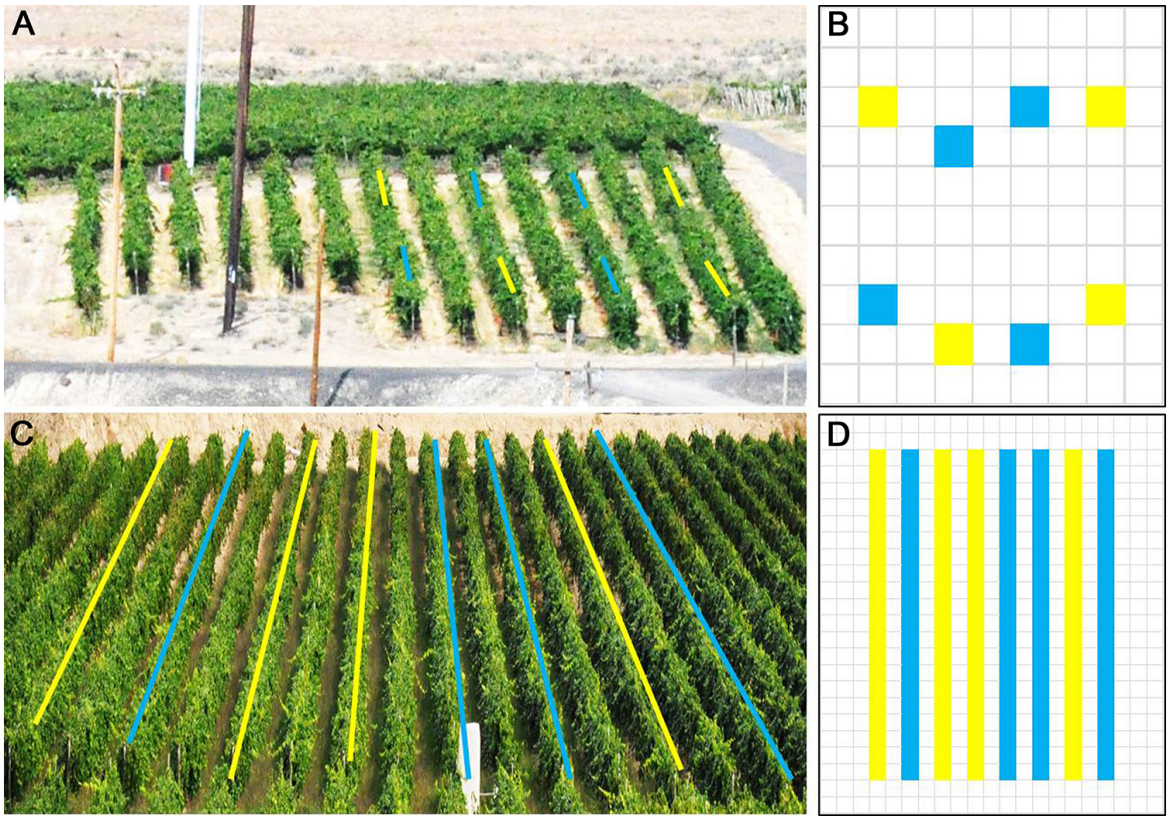

Figure 4. Example of a Completely Randomized Design (CRD) with four replicates of a single treatment and a control in (A) a small and (C) a mid-size vineyard block. Yellow lines and blue lines indicate the control and treatment, respectively. Note: Both examples show buffer rows between treatments. In a CRD, sometimes the same treatment may end up in adjacent rows. Schematic diagrams of the trials shown in A and C are shown in B and D, respectively. Photos and illustrations by Hemant Gohil.

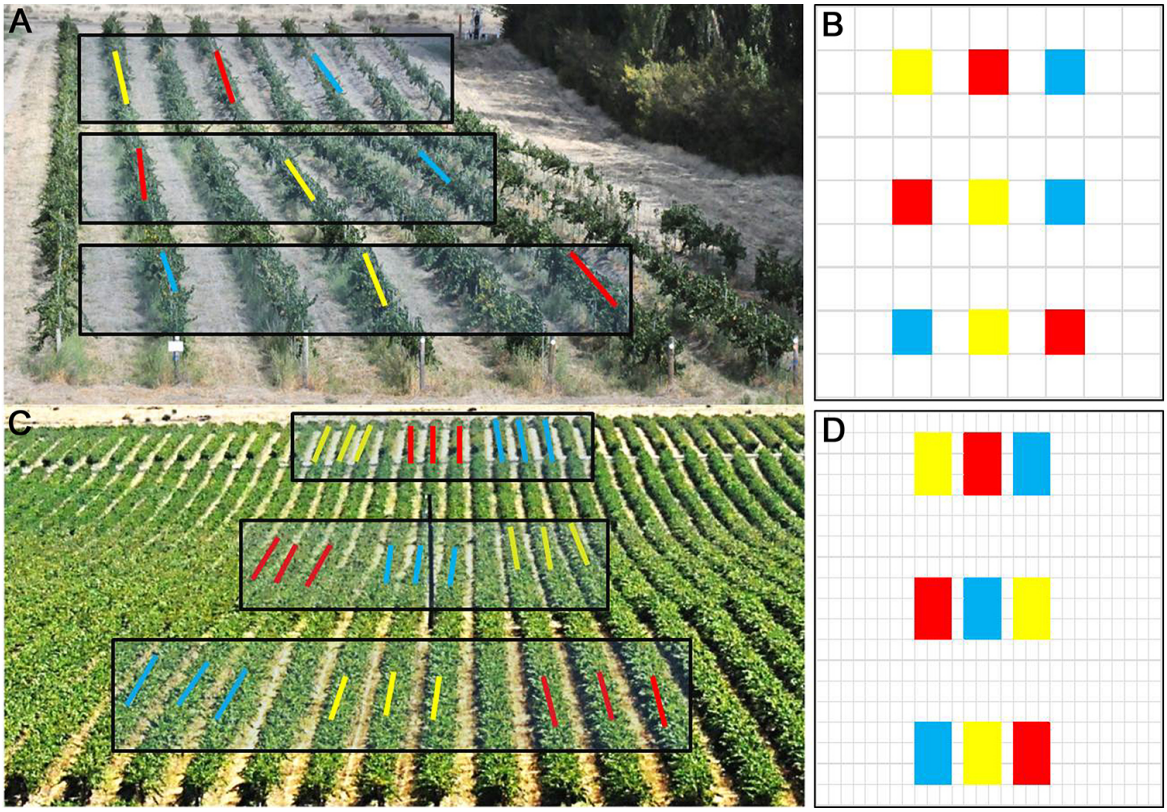

Figure 5. Example of a Randomized Complete Block (RCB) design with three replicates of two treatments and one control. The black frames in A and C represent the grouping of different zones that was done to account for variability due to slope. In this design, the treatments and control (red, blue, and yellow lines) are randomized within each zone. In the smaller field design (A), two to three vines in a single row might be the replicate unit. In a larger design (C), three to five vines across multiple rows might serve as the replicate unit. Schematic diagrams of the trials shown in A and C are shown in B and D, respectively, where individual cells represent a single vine. Photos and illustrations by Hemant Gohil.

Evaluation of New Pesticides or Fertilizers

Pesticide or fertilizer evaluation trials are a simple way to determine if new products on the market are more effective than what is currently being used. Due to the complexity of evaluation and interpretation, it is recommended to only test one or two new products at a time. In most pesticide and fertilizer trials, the vintage can significantly influence results. In these trials, controls are absolutely necessary along with repetition of the experiment in either climatically distinct areas or over at least two years.

The RCB design is generally necessary in these types of experiments due to the high level of influence the site variability has on vine responses, as described in *General Experimental Design*. For example, changes in slope and soil type can change vine-available water thus influencing growth. Such differences in growth might be misconstrued as a change in response due to a fertilizer application.

When testing new products, always place buffer vines or rows between treatments to help reduce compromising treatment integrity that might otherwise result from run-off, drift, leaching, or other sources of cross-contamination.

For example, if you apply a new pesticide in a small experimental area, and you only want to collect data from three adjacent vines within a treatment replicate, then apply that pesticide to at least five adjacent vines. Only record observations from the center three of the five vines (Figure 4A and 5A). If larger areas are available, then treat vines in three adjacent rows for a single replicate and only record observations from the center row. The flanking rows will be considered buffers, along with the outside vines in the center row (Figure 5B).

Spray coverage and application uniformity are very important to maximize the outcome of a management program. Spraying or ground-applying products without proper pre-calibration can dramatically change the amount of pesticide or fertilizer applied. Information on how to calibrate sprayers can be found in the annually updated *Pest Management Guide for Grapes in Washington* (Hoheisel and Moyer). Calibration is required for both machine (sprayer) and hand (backpack) application.

Understanding the type and nature of pest damage is very important before any changes to a control program are implemented.

The *Field Guide for Integrated Pest Management in Pacific Northwest Vineyards* (Moyer and O'Neal 2013) provides comprehensive information on distinguishing differences between symptoms caused by various pests and disorders. The *Field Guide* also provides information on how to develop management strategies.

When conducting new product trials, make sure that the only difference between your treatments is the actual products. In the case of pesticide trials, this means that the same surfactant and application method are used for both the treatment and the control. In the case of fertilizer trials, this means that irrigation practices (e.g., amount, timing) are the same between all treatments and controls.

The style of product application is important. If only a few vines are selected for treatment, application using a backpack sprayer might be the best option. In fertilizer trials, consider whether applications will be made through fertigation, soil application, or foliar application. If applications are made to the soil, consider the timing of water application as well, as most fertilizers need moisture to be moved into the soil. However, be cautious about over-application of water that may result in fertilizer leaching and excessive plant vigor.

When and what to sample. Pest ratings are important to take during the growing season, but the most important time to collect these data is directly before harvest. Whole clusters or leaves are assessed for both incidence and severity. Since the human eye is naturally drawn to differences, tagging clusters that will be rated should be done ahead of any product application to avoid bias in selection.

In the case of fertilizer trials, canopy development over the duration of the treatment, pruning weight, yield and its components, and fruit composition may be important in determining the effectiveness of a new fertilizer. Tissue nutrient tests at either bloom or véraison may also be helpful, especially if a fertilizer is applied to alleviate a deficiency situation (Davenport and Horneck 2011).

Interpreting results. In an example fertilizer trial, you decided to compare the ability of a foliar-applied zinc fertilizer (treatment) to improve fruit set relative to your standard soil-based zinc application (control). You estimated fruit set indirectly by comparing number of berries per cluster within each of your treatment and control replicates. If there were no differences in fruit set, this indicates that the foliar-applied zinc product is no better (and no worse) than the soil-applied product (standard practice).

If the foliar-applied zinc fertilizer is cheaper, easier to apply, or allows for other practices to take place normally, then its use should be considered. If the foliar-applied zinc did not improve fruit set relative to your standard soil-applied zinc and the foliar application process is more expensive or hinders other activities, then it may not be advantageous to adopt its use.

If the foliar-applied zinc resulted in reduced fruit set, get the tissue tested for potential nutrient toxicity (i.e., too much zinc made it into the plant) or deficiency (i.e., not enough zinc made it into the plant).

This will allow you to decide whether to repeat the trial with a lower or higher rate or to not use the product altogether. Always consider the product and application cost to make sure that any additional cost of the new product or method is offset by the income achieved using the product or method.

Evaluation of Canopy Management Techniques

Canopy management practices are designed to promote sunlight interception by the canopy, optimize light penetration into and air movement through the fruiting zone, reduce excessive shoot growth, divert photosynthetic products (sugars) to developing and ripening berries, and improve the ease of certain vineyard operations. Canopy management trials may be designed to test different pruning strategies, such as cane versus spur pruning, mechanized versus hand pruning, number or position of retained buds, and/or the timing of pruning. Canopy management trials can also include the timing and degree of shoot positioning, thinning, or hedging, or the timing, extent, and location of leaf removal or cluster thinning.

The RCB design is typically the most appropriate for these trials. Canopy management trials should be conducted for at least two years to test the consistency of results and account for potential carryover effects (e.g., bud fruitfulness).

When and what to sample. In canopy management trials, observations typically consist of changes to the canopy itself as well as impacts of a practice on yield, yield components, and on fruit composition. Point quadrat analysis (Smart and Robinson 1991) can be used to determine canopy density. Shoot length or the number of count and non-count shoots are appropriate measures of canopy development, and winter pruning weight may be used as a measure of overall effects. Fruit composition can be monitored from véraison through harvest or assessed at harvest.

When collecting clusters or berries for analysis, be sure to include fruit from both sides of the canopy as well as from sun-exposed and shaded positions in the canopy, unless canopy side or sun exposure are factors you are specifically trying to differentiate between. Additional important observations include: (i) the timing of key phenological stages, such as bloom, lag phase, véraison, and commercial fruit maturity; (ii) yield components, such as berry and cluster weight and number of clusters per vine; and (iii) pesticide penetration and spray coverage (Figure 6).

Figure 6. The deposition of a pesticide can be assessed using food-grade fluorescent dyes or water-sensitive card, as shown here. Cards should be placed in the specific areas of the canopy you are interested in assessing. When designing an experiment, each treatment replicate should have the same distribution of spray cards. For general spray assessment purposes, most cards should be placed within the fruit zone, with the water-sensitive side of the card facing outward to the vineyard row. Ideally, cards should be placed vertically in the middle of the fruit zone. Photo by Hemant Gohil.

Interpreting results. Interpreting canopy management trial results can be challenging due to myriad factors that influence vine growth and fruit composition. If a specific treatment consistently results in improved fruit quality relative to your control and the cost to execute that practice is not limiting to production, consider adopting it. If quality is similar between the treatment and the control, then additional factors such as higher yield, advanced or delayed harvest date, reduced cost of application, or improved pest management should be considered in order to decide whether to adopt or discard the new practice. In any case, the trial should be repeated the following year to rule out the effects of weather (e.g., rain or heat). If results hold true after repeating a trial, then you have a strong case for adopting the practice.

Evaluation of Irrigation Management Strategies

Common irrigation trials focus on the impact of different rates and/or timing of irrigation on control of canopy development, yield components (e.g., berry weight), and fruit composition. Regulated deficit irrigation (RDI), or intentionally inducing mild to moderate drought stress on the vine after fruit set, is a commonly applied irrigation management strategy (Moyer et al. 2013). Irrigation trials are best designed on a large scale, with irrigation treatments (regimes) applied to full rows (Figure 7).

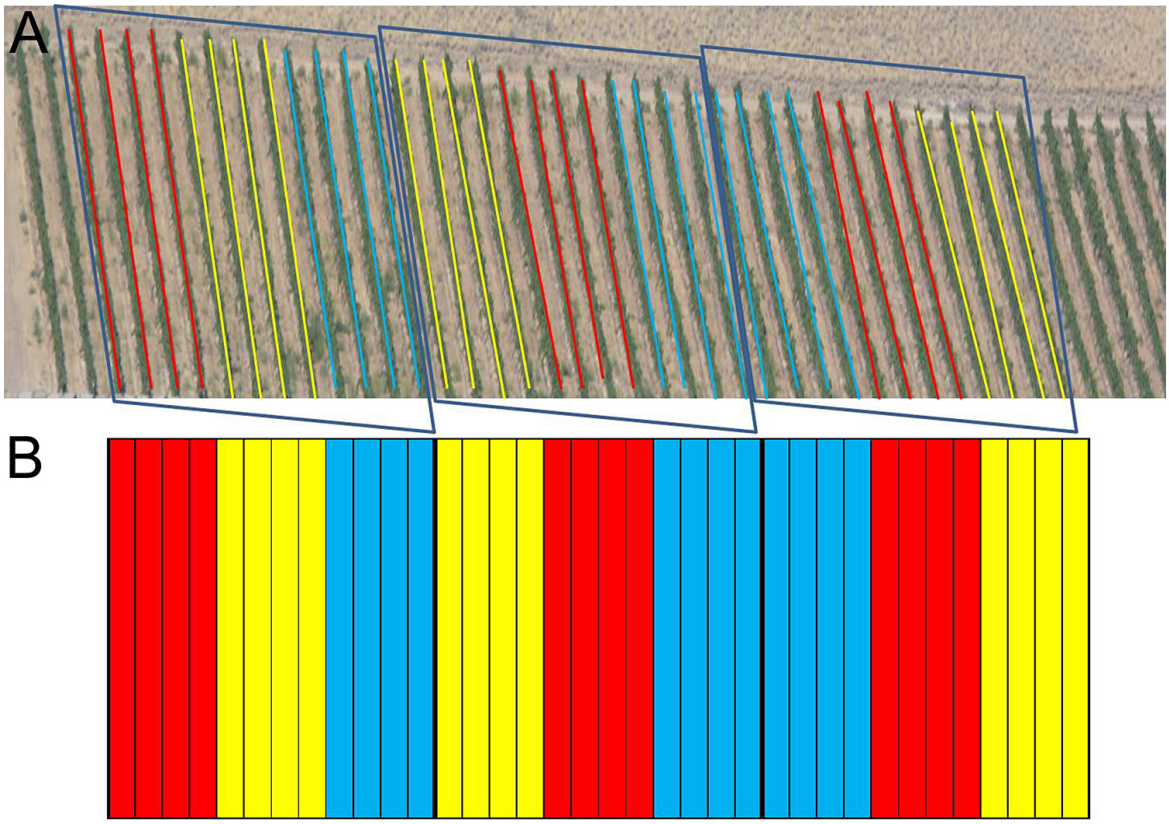

Figure 7. (A) Arial view of an irrigation trial design in a 10-acre vineyard using a RCB design that assures each zone receives all treatments. Different colors represent different irrigation treatments. Blue boxes represent "blocking" (zoning) to account for soil variability. Entire rows are irrigated for ease of management. In this design, four adjacent rows receive the same treatment and sampling should be done from one or both of the middle rows. (B) Schematic diagram of the same design is shown. Photo and illustration by Hemant Gohil.

Nevertheless, small-scale trials may be implemented by using drip emitters of different flow rates to alter the amount of water applied within a row. It is important, in this case, to change emitters repeatedly during the growing season to ensure that water amounts differ only during the intended deficit period. Irrigation trials are typically laid out using the RCB design in the Pacific Northwest where soil types tend to vary, thus requiring blocking. As with pesticide and fertilizer trials, treatment replicates in irrigation trials should have buffer rows and vines to reduce treatment "drift."

When and what to sample. Vegetative measurements, such as shoot length and canopy density, are appropriate measurements for assessing the impact of irrigation treatments. Total yield, berry weight (and/or diameter) are also appropriate measurements. As with other experiments, fruit composition is typically assessed either from véraison to harvest or at harvest only. If possible, wine can be made using the fruit from each replicate to determine whether there are any differences in wine quality. Additional measurements of interest might be pest pressure as estimated by incidence and severity and the progression of phenological stages (e.g., timing of véraison).

Interpreting results. When assessing the influence of irrigation regimes on yield or its components, there are several considerations to make (Table 1).

For example, if a certain RDI treatment (e.g., replacing 30% of crop evapotranspiration or ET_C) results in smaller berries relative to the control (e.g., standard RDI at 70% ET_C) and this is something you desired, then it may be a good practice to adopt. If the treatment results in marginal loss of yield but that loss is associated with losses in fruit quality, then you may need to reconsider adopting this deficit regime.

Factors that can influence the results of an irrigation trial include:

1. Unexpected high precipitation could result in insufficient water deficit to achieve desired outcomes; and/or
2. in a warm year, reduced irrigation (severe RDI) may impart greater stress, which may compromise canopy development and yield formation.

In all cases, repeat the trial at least one more year before making a decision regarding implementation of a new irrigation regime.

Table 1. Example of yield per vine from a large-scale irrigation trial in a wine grape vineyard. There are three irrigation treatments: 30% ET_C, 30%–70% ET_C, and a control of 70% ET_C. Treatments were assigned to three entire rows, and each replicate was comprised of three vines within the middle row. For more information on calculating means and standard deviations, see the section Basic Interpretation of Results.

Replicate	Sample vine	Yield (lbs/vine)		
		30% ET_c	30–70% ET_c*	70% ET_c (control)
1	1	9.1	18.0	22.9
	2	12.2	21.2	26.0
	3	14.1	25.6	19.9
	Replicate mean	*11.8*	*21.6*	*22.9*
2	1	11.0	20.1	20.9
	2	1.1	19.0	23.9
	3	14.2	23.6	27.0
	Replicate mean	*8.8*	*20.9*	*23.9*
3	1	14.6	22.5	30.2
	2	15.2	25.2	32.1
	3	16.1	27.0	27.9
	Replicate mean	*15.3*	*24.9*	*30.1*
4	1	12.1	18.0	19.9
	2	8.0	22.1	22.9
	3	10.6	24.2	25.0
	Replicate mean	*10.2*	*21.4*	*22.6*
	Treatment Mean	*11.5*	*22.2*	*24.9*
	Treatment Standard Deviation	*2.8*	*1.8*	*3.5*

* 30% pre-véraison and 70% post-véraison

Evaluation of Varieties, Clones, or Rootstocks

Variety, clone, and/or rootstock trials are conducted to evaluate suitability of a particular vineyard location for a desired variety, clone, or rootstock. Each variety, clone, and/or rootstock has horticultural traits that may be more or less suitable for the area. These traits may be related to differences in phenology, vigor, canopy or cluster architecture, cold hardiness, fruit composition, or wine quality or style (Keller 2015). Such trials are long term, taking three years to establish and another three or more years of data collection to effectively evaluate the plants selected and their interactions with variable weather year to year.

Site uniformity is often hard to assess in a new vineyard planting, so the RCB design is recommended. Standard protocols must be followed for cultural practices as well as winemaking in order to make accurate comparisons. In addition to the typical assessment of vine growth, yield components, and fruit composition, key factors to consider when evaluating a variety, clone, and/or rootstock in Washington are winter survival and growing season heat requirements (e.g., growing degree days, or GDD, needed to adequately ripen the fruit). Enough plants should be grown per replicate to produce at least ten gallons of wine, which is about 4 cases (48 bottles of 750 mL). Ten gallons of wine requires approximately 130 lbs of fruit per replicate (10 to 20 vines per replicate, depending on planting density and yield).

The control in such trials is usually a variety or clone (if known) that performs well in the area or, for rootstock trials in eastern Washington, own-rooted plants. This is one experimental type where there may not be a true control, but rather, the interpretation is based on side-by-side comparisons within the experimental area.

When and what to sample. It is not feasible to sample for every aspect of growth and development. During the first three years, grapevines may not produce sufficient fruit for winemaking. However, an annual assessment of cold damage and disease incidence/severity can be made to eliminate susceptible varieties/clones early on. Cold hardiness should be considered as a key selection criterion in Washington and other northern production regions. The WSU Extension manual *Assessing and Managing Cold Damage in Washington Vineyards* explains in detail how to conduct damage assessments (Moyer et al. 2011).

If you are testing new varieties or rootstocks for pest susceptibility, periodical rating of pest incidence and severity is a direct indicator of success. Final harvest ratings, however, are easier to make and are recommend if time is limited.

Varieties vs. Clones

Varieties are types within a species which evolved in nature or were bred or selected by humans. Each variety has a distinct genetic makeup that imparts diverse characteristics in terms of leaf shape, fruit color, cluster size, canopy structure, and other attributes. Botanically, grape varieties are called “cultivars” because only the latter remain true-to-type owing to vegetative reproduction (Keller 2015). Many varieties are grown based on consumer demand or requirement from the winery. Such varieties may or may not be adapted to the site on which they are grown.

Clones are the result of years of selection for desired traits and are propagated vegetatively, often by cuttings from the original mother vine. That trait could be cluster size or compactness, vigor, fruit composition, or timing of ripening (Keller 2015). A variety may give rise to multiple clones that are somewhat different from each other. A specific clone may perform differently on different sites due to variances in local climate, soil, or cultural practices.

Vigor can be estimated from winter pruning weight and cane number. Yield measurements at harvest may be supplemented with data on yield components (e.g., cluster number, cluster weight, or berry weight).

Considering fruit composition as specified by the intended end user (e.g., winery, juice processor, or consumers) can provide an idea of when and what to sample. For example, different varieties may have different target sugar or acid levels for harvest depending on the intended end use. When harvesting fruit for wine production, harvest different varieties, clones, or rootstocks at the same sugar level rather than on the same calendar date, as sugar content influences many aspects of winemaking and sensory evaluation. This might mean that harvest within a trial occurs over the course of several days or even weeks. Wine samples can be sent to commercial laboratories for compositional analysis. Blind tasting, preferably involving winemakers, is also helpful in wine evaluation.

Interpreting results. In many cases with variety or clone trials, particular selections may not survive the duration of the study; that is to be expected if you are conducting the trial in a location where grape production is relatively unknown or if you are testing a variety or a rootstock that is entirely new to a region. If this is the case, the instant interpretation is that the failed selections are not suitable for your location. In the following example, used to help outline how to interpret trial results (Figure 8), the assumption is that you are evaluating varieties that survived beyond the third year of vineyard establishment.

Figure 8. Variety trial in a small vineyard, using RCB design. A single zone as marked by blue borders represents a single replicate of four varieties in four adjacent rows, with five vines per variety. Photo by Hemant Gohil.

In an experiment with four treatments (varieties, clones, and rootstocks) and four replicates, each comprising 10 to 20 vines per treatment, you may be forced to limit data collection to key stages such as véraison and harvest. However, take notes of general growth and appearance throughout the growing season. After seven to eight years of cultivation, the final assessment should consider all the parameters discussed above.

If pest resistance is the primary objective for selecting a new variety or rootstock, then healthy leaves and clusters are the signs of disease resistance or that your spray program is effective, particularly in a year with high disease pressure. Additional factors such as cluster compactness and berry size should also be considered, as they may impact pathogen susceptibility. In a clonal trial where all clones of the same variety were subjected to similar fertilizer, irrigation, and spray regimes, differences in canopy size are the result of inherent differences in vigor among clones. If two clones are more or less the same in yield, then additional factors such as advancement or delay in sugar accumulation, cluster compactness, and fruit quality should be compared.

Basic Interpretation of Results

After you have invested your time and money in the collection of data from a trial that was carefully designed to account for spatial variation across the vineyard and temporal variation between years, it is equally important to ensure that the conclusions you draw from those data are robust and reliable. Drawing conclusions based on trends or average numbers alone can be misleading. Regardless of the calculated average, data that are highly variable within a treatment tend to be unreliable and suggest that something other than your treatment is influencing, and perhaps distorting, the results.

Generally speaking, you want to maximize the variation in your data that is due to the imposed treatment(s) and minimize the variation within each treatment (i.e., between replicates; see also section *Advanced Statistical Analysis*). Statistical tests provide assessments of both central tendency (e.g., average or mean) and variability in the data. In general, simple statistics such as mean and standard deviation, which are explained below, are necessary to initially interpret results (Vaux 2012).

Mean

The mean is a measure of central tendency; it is also called the average of a series of numbers that, in our case, are derived from measurements or observations. Mathematically, the mean is the sum of all the individual numbers or values, divided by the number of values used for that sum.

Calculating the mean is shown in Equation 1.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{(x_1 + x_2 + x_{3\ldots} + x_n)}{n} \qquad \textbf{(Equation 1)}$$

Where $\bar{x}$ = mean, Σ = an operational symbol indicating "to sum" or add all of the subsequent values x_i = (x_1, x_2, *etc.*), and n = total number of values (in an experiment, this is usually the number of replicates).

Standard Deviation

The standard deviation is a measure of variation within the observations; it is usually written with a "±" symbol to indicate the fact that the measured values scatter both above and below the mean. It is useful for deciding whether apparent differences between means are likely to be real.

For data that follow a typical bell-shape (i.e., normal) distribution, about two thirds of the measured values are within one standard deviation of the mean and about 95% of the values are within two standard deviations. Consequently, a small standard deviation relative to the size of the measurements suggests that the measured values are close to the mean (i.e., there is a small variation) while a large standard deviation indicates that the measured values are more scattered around the mean (i.e., there is a large variation).

In other words, data with a high standard deviation are highly variable. Such data tend to be less reliable than data with a small standard deviation. The means of two different treatments may not actually be different if the standard deviations of one or both of these means are high. The calculation for standard deviation is presented in Equation 2.

$$SD = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n - 1}} \qquad \textbf{(Equation 2)}$$

Where SD = standard deviation, x_i = an individual value within a treatment, $\bar{x}$ = the mean of the treatment, n = the total number of measurements or observations (replicates) within the treatment. Equation 2 is shown below using an example of calculating the standard deviation for four different data values (replicates).

$$SD = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2}{3}}$$

The mean and standard deviation statistics can be very powerful in aiding the basic interpretation of trial results. For example, looking at the fertilizer trial shown in Table 2, we see that the highest yield based on the mean alone was reached in the 30 lbs N treatment applied at three times (90 lbs N total) (control), while the 20 lbs N treatment applied at 3 times (60 lbs N total) resulted in the second highest yield. If we were only to look at these means, then the control was the "best performing" fertilizer regime. However, the rather large standard deviations suggest that the top two regimes are not that different from each other, and one can achieve similar yield results with less N applied.

Table 2. Example of yield per row from a fertilizer trial in a vineyard. Three fertilizer treatments: 10 lbs N, 20 lbs N, and a control of 30 lbs N per acre applied three times over a growing season. An entire row each comprising 80 vines was treated using a granular application of fertilizer.

	Yield (lbs/row)		
Replicate	10 lbs N/acre 3 times	20 lbs N/acre 3 times	30 lbs N/acre (control) 3 times
1	3520	3360	4080
2	3520	4240	3840
3	3600	4080	4000
4	3760	3680	4400
Mean	*3600*	*3840*	*4080*
Standard Deviation	*113*	*397*	*236*

Other examples of common mean and standard deviation outcomes from experiments are highlighted in Figure 9, using a pruning trial as an example. In both scenarios presented in Figure 9A and 9B, the mean yield of the treatment (mechanized pruning) is almost double that of the control (manual pruning).

In Figure 9A, however, that high treatment mean is associated with a large standard deviation. If high yield were the main goal, it would be tempting to select the treatment with the highest average response in this scenario. This would also mean that the response would be more variable. In Figure 9B, the standard deviation is fairly consistent between the two treatments, indicating that the high yield with mechanized pruning will likely be consistently higher relative to the control.

Nevertheless, further analysis using a *t*-test or ANOVA as described in Appendix 2 should be performed to confirm this conclusion. Remember that the standard deviation only indicates the spread among the observations within a treatment; it does not indicate statistical significance.

Standard Error of the Mean

The standard error of the mean is similar to the standard deviation, but it estimates how close the measured treatment average (e.g., based on four vines or four rows) is to the actual treatment average (e.g., all of the vines that received that treatment). The standard error is estimated by dividing the standard deviation by the square root of the number of observations or replicates (Equation 3).

$$SE = \frac{SD}{\sqrt{n}} \qquad \textbf{(Equation 3)}$$

Where SE = standard error of mean, SD = standard deviation, and n = number of observations.

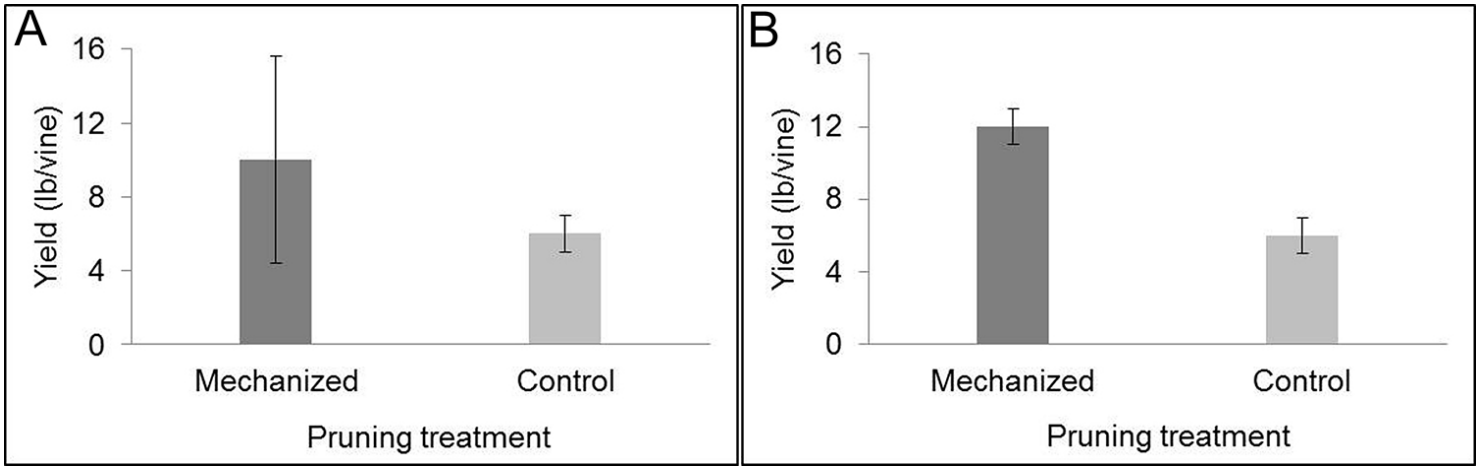

Figure 9. Interpreting the mean and standard deviation of the results of an experiment can be complicated. Here, two potential outcomes of a pruning trial comparing mechanized pruning (treatment) and manual pruning (control) are shown. Lines on the top of each bar represent the standard deviation. (A) The mean yield in mechanized pruning is almost double compared to the control; however, it has a standard deviation which is almost five times larger than that of the control, suggesting high variability in yield compared with the control, making any conclusions unreliable; (B) the mean yield in mechanized pruning is almost double compared to the control, and the standard deviations are similar, indicating that the variability within mechanized pruning and control is similar and that mechanized pruning clearly outperformed the control in terms of yield.

Using Micro-Processing Sheets for Statistics

If you have a computer with productivity software that includes spreadsheets (e.g., Microsoft Excel, Apache Calc, Google Docs Spreadsheet, Zoho Sheet, etc.), many statistical calculations and tests are built directly into the programs. This makes calculations fast and easy, provided data are arranged in a specific way (in columns and rows). In many cases, outputs include the mean, standard deviation, standard error of the mean, ANOVA, and *p*-values (see the help information for your software of choice):

Microsoft Excel

Apache Calc

Google Docs Spreadsheet

Zoho Sheet

Conclusion

Successful on-farm trials require a statistically valid experimental design, which incorporates key components such as the use of a control, randomization, replication, and repetition to separate the natural vineyard variability from the effects of the treatment(s) being tested. The examples presented here are of common on-farm trials. While they are not all inclusive, they can serve as templates for other types of experiments.

Careful planning of on-farm trials is needed, as each experiment requires specific types of data to be collected and different strategies for collecting these data. Of course, simply collecting data does not provide insight into the value of a change in cultural practice or product. Collected data need to be analyzed and interpreted properly in order to draw meaningful conclusions.

Even after analysis, placing the results of a trial in context with practical vineyard operations is needed; statistical significance in a trial does not always mean practical significance when considering other factors such as cost and time.

References

Calderon-Orellana, A., L. Mercenaro, K.A. Shackel, N. Willits, and M.A. Matthews. 2014. Responses of Fruit Uniformity to Deficit Irrigation and Cluster Thinning in Commercial Winegrape Production. *American Journal of Enology and Viticulture* 65:354-362.

Coombe, B. G. 1995. Adoption of a System for Identifying Grapevine Growth Stages. *Australian Journal of Grape and Wine Research* 1:104-110.

Davenport, J. and D. Horneck. 2011. Sampling Guide for Nutrient Assessment of Irrigated Vineyards in the Inland Pacific Northwest. *Pacific Northwest Extension Publication* PNW622. Pullman, WA: Washington State University.

Gates, C. E. 1991. A User's Guide to Misanalysing Planned Experiments. *HortScience* 26:1262-1265.

Hellman, E. 2004. How to Judge Grape Ripeness Before Harvest. Southwest Regional Vine and Wine Conference. Albuquerque, NM, Feb 27-28.

Hoheisel, G. A. and M. M. Moyer. 2015 (Updated Annually). Pest Management Guide for Grapes in Washington**.** *WSU Extension Publication* EB0762. Washington State University.

Iland, P., N. Bruer, G. Edwards, S. Caloghiris, and E. Wilkes. 2013. Chemical Analysis of Grapes and Wine: Techniques and Concepts. 2nd Ed., Adelaide: Patrick Iland Wine Promotions.

Keller, M. 2015. The Science of Grapevines: Anatomy and Physiology. 2nd Ed., London: Academic Press.

Ketterings, Q., K. Czymmek, and A. Gabriel. 2012. On Farm Research. Nutrient Management Spear Program. *Cornell University Cooperative Extension* Agronomy Factsheet #68. Cornell University.

Koenig, R. T., M. Winger, and B.C. Kitchen. 2000. Simple Low Cost Data Collection Method for Agricultural Field Studies. *Journal of Extension* 38:2-6.

Komm, B. L. and M.M. Moyer. 2015. Vineyard Yield Estimation. *WSU Extension Publication* EM086. Washington State University.

Little, T.M. and F.J. Hills. 1975. Statistical Methods in Agriculture Research. Davis: Univ. of California Press.

Meier, U. 2001. Growth Stages of Mono- and Dicotyledonous Plants: BBCH-Monograph. Berlin: Blackwell Wissenschafts-Verlag.

Moyer, M., L. Mills, G. Hoheisel, and M. Keller. 2011. Assessing and Managing Cold Damage in Washington Vineyards. *Washington State University Extension* EM042E. Washington State University.

Moyer, M. M. and S. O'Neal (eds.). 2013. Field Guide for Integrated Pest Management in Pacific Northwest Vineyards. *Pacific Northwest Extension Publication* PNW644. Pullman: Washington State University.

Moyer, M.M., R.T. Peters, and R. Hamman. 2013. Irrigation Basics for Eastern Washington Vineyards**.** *WSU Extension Publication* EM061e. Washington State University.

Nail, W. 2012. Collecting Berry Samples to Assess Grape Maturity. eXtension.org.

Pagay, V. and L. Cheng. 2010. Variability in Berry Maturation of Concord and Cabernet Franc in a Cool Climate. *American Journal of Enology and Viticulture* 61:61-67.

Smart, R. and M. Robinson. 1991. Sunlight Into Wine. A Handbook for Wine Grape Canopy Management. Adelaide: Winetitles.

Rankine, B. C., K. M. Cellier, and E. W. Boehm. 1962. Studies on Grape Variability and Field Sampling. *American Journal of Enology and Viticulture* 13:58-72.

Skinkis, P.A. 2013. Basic Concept of Vine Balance. Oregon State University via *eXtension.org*.

Smart, R. E., J. K. Dick, I. M. Gravett, and B. M. Fisher. 1990. Canopy Management to Improve Grape Yield and WineQ: Principles and Practices. *South African Journal of Enology and Viticulture* 11:3-17.

Vaux, D. L. 2012. Know When Your Numbers are Significant. *Nature* 492:180-181.

Veseth, R., S. Wuest, R. Karow, S. Guy, and D. Wysocki. 1999. On-Farm Testing – A Scientific Approach to Grower Evaluation for New Technologies. *Pacific Northwest Conservation Tillage Handbook.* Washington State University.

Wolpert, J. A., G. S. Howell, and C. E. Cress. 1980. Sampling Strategies for Estimates of Cluster Weight, Soluble Solids and Acidity of 'Concord' Grapes. *Journal of the American Society of Horticultural Science* 155:434-438.

Wuest, S. W., S. O. Guy, L. J. Smith, and B. C. Miller. 1995. On-Farm Tests as a Tool for Extension Programming. *Journal of Extension* 33:4-8.

Appendix 1

Examples of on-farm trials and guidelines on the types of data that may be collected, as well as potential challenges involved with each type of experiment.

Category	Experiment Examples	Measurements	Challenges
Pruning	Mechanization Cane vs. spur pruning Delayed dormant pruning	Phenology, canopy density, yield, fruit composition, pruning weight	Getting uniform levels of pruning, especially with machine. Maintaining vine balance. Crop load differences affecting time of maturity. Time of budbreak affecting subsequent phenology.
Canopy Management	Shoot thinning	Phenology, canopy density, yield, fruit composition, pruning weight	Costs associated with manual labor. Inaccuracy of machine intervention.
	Fruit-zone leaf removal		Costs associated with manual labor. Excessive fruit exposure leading to sunburn. Inadequate timing (before rachis elongation or after bunch closure) compromising fruit set or fruit composition.
	Cluster thinning		Costs associated with manual labor. Seasonal weather conditions altering the optimum crop load. Inadequate timing leading to compensatory berry growth or fruit set or failure to improve fruit composition.
Irrigation	Timing of deficit irrigation	Phenology, shoot length, yield, fruit composition, pruning weight, soil moisture, plant water status	Seasonal variation in weather masking effects of irrigation. Inadequate timing (pre-bloom or pre-harvest) compromising fruit set or ripening or leading to berry dehydration.
	Severity of deficit irrigation		Calculating the exact amount of water to be delivered. Ability to deliver desired amounts of water. Excessive water stress compromising yield formation, fruit development, vine capacity, or cold hardiness.
Variety, Clone, or Rootstock	Fruit quality	Yield, fruit composition, blind tasting	Complex experimental design required for multiple selections. Evaluations are long-term and may be time-consuming. Different varieties may require different cultural practices. Adjusting management practices in subsequent years of trial.
	Pest resistance	Pest ratings (% surface area of clusters damaged)	
	Vine development	Phenology, shoot length, yield, pruning weight	
New Products	Pesticide rates, application timing, alternative products	Pest ratings (% surface area of clusters damaged), yield	Poor spray coverage. Not driving the tractor at the appropriate speed. Sprayer not calibrated.
	Fertilizer rates, application timing, type of delivery, alternative products	Tissue nutrient analysis, yield, cluster weight, pruning weight, fruit composition	Uniformity of application. Inadequate timing leading to nutrient leaching. Low soil moisture at the time of application compromising fertilizer uptake. Administering the same amount of fertilizer using different methods of application. Excess vegetative growth, pest susceptibility, or toxicity due to too much fertilizer.
Vineyard Floor Management	Beneficial insect attraction Weed control Building organic matter Managing soil compaction	Beneficial insect counts, weed growth, insect damage ratings (% canopy lost, % feeding damage), yield	Knowing what attracts beneficial insects. Excessively tall floor vegetation interfering with other vineyard activities and increasing frost risk. Resilience to tolerate traffic. Cover crop may need supplemental water to survive. Erosion and poor water infiltration associated with bare soil surface.

Appendix 2

Advanced Statistical Analyses

For readers who are interested in a "behind-the-scenes" look at statistical analysis, there are more advanced approaches to analyzing and interpreting results. Some of these approaches are briefly described below.

Hypothesis testing

A hypothesis is a definitive statement about how something is expected to work or how a test will perform before the evidence is in. It typically reads something like: "This new fertilizer will increase my per-acre yield by 2 tons." Hypothesis testing is a way to design and analyze experiments to determine if that hypothesis is true or false. It can be compared to the US justice system, where a person is considered innocent until evidence (data) provides enough support beyond a reasonable doubt that they are guilty. In statistics, typically the *null hypothesis* is a hypothesis that presumes that there is no difference between two or more treatments (or treatment/s and a control) or sets of observations. The null hypothesis is taken to be true until statistical tests demonstrate otherwise. Whether to accept or reject the null hypothesis is based on our pre-defined threshold of statistical significance (see below under *p*-value). The *alternative hypothesis* is the opposite of the *null* hypothesis, and is the hypothesis that states there is a difference between two or more treatments and/or the controls. When people discuss the "statistical significance" of a result, they are often referring to the results obtained when evaluating the null or alternative hypothesis. One way to evaluate whether a hypothesis should be accepted or rejected based on the data collected is to conduct a *t*-test.

t-test

When only two treatments (or a treatment and a control) are compared, a *t*-test (technically, a two-tailed *t*-test) would tell us if there is a statistically significant difference between means of the two sets of measurements or observations. Once you have the mean and standard deviation values for the treatments, it is easy to perform a *t*-test. Use equation 4 to calculate the *t*-value.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}} \qquad \text{(Equation 4)}$$

Where, $\bar{x}_1$= mean of treatment 1
$\bar{x}_2$= mean of treatment 2
SD_1^2 = variance (standard deviation squared) of treatment 1
SD_2^2 = variance of treatment 2
n_1= number samples (replicates) in treatment 1
n_2= number of samples in treatment 2

After the *t*-value is calculated, it may be compared to an expected or tabulated *t*-value. Such values are listed in Table 3. The so-called degree of freedom (df) that is required to choose the correct *t*-value is calculated using the following equation: $df = n_1 + n_2 - 1$. Following the row corresponding to your degree of freedom, the expected *t*-value is found in the corresponding column of the desired level of significance. The most common level is a two-tailed *t*-value where $p = 0.05$.

If the calculated *t*-value is greater than the tabulated value, then the two means are said to be significantly different. In effect you are asking the question: "What is the likelihood I would see this result by chance?" In the case of a *t*-value associated with a *p*-value of 0.05 the answer is: 5% or less. This can be interpreted as the response you are seeing is most likely due to the treatment effect, since randomly seeing this type of response is rare (only 5% of the time).

p-Value

The *p*-value is the probability of the so-called *null* or *alternative* hypothesis to be true. For example, in our irrigation trial (Table 1), the *null* hypothesis is that yield does not differ among the treatments. Probability values range from 0 to 1 (i.e., 0 to 100%). The lower the *p*-value, the lower the probability that the difference is due to chance alone (e.g., 1% for a *p*-value of 0.01). Consequently, the smaller the *p*-value, the stronger the case for the treatment causing an effect. In most cases, a threshold *p*-value of 0.05 is used to determine whether an effect is statistically significant (i.e., $p \leq 0.05$) or insignificant (i.e., $p > 0.05$). While the *p*-value is important for statistical inference, it does not always indicate *practical* significance. For example, if a new treatment results in a statistically significant increase in fruit sugar content by 0.5 °Brix at harvest, but the treatment costs more than your standard practice, then you should determine whether that 0.5 °Brix improvement is really worth the extra money.

Analysis of Variance (ANOVA)

A number of factors can influence the outcomes of an experiment and the behavior of grapevines in general. All of these factors are thus sources of variation in measured or observed traits (e.g., see Keller 2015 for a discussion on sources of variation in fruit composition).

In most vineyard trials, there are two main sources of variation: 1) the variation that is the result of using different treatments (termed "variation among/between"), and 2) the variation that might be present irrespective of any treatment due to inherent (i.e., genetic or environmentally caused) differences between experimental units, such as individual vines or blocks (termed "variation within"). Given that most modern vineyard blocks are planted to genetically homogenous clonal material, most of the variation within is caused by differences in soil conditions and microclimate.

An ANOVA test is designed to identify to what extent each of these sources of variation are influencing the results. If, for example, the vine-to-vine variation is dominant, that indicates that while we might see differences in means between treatments it is possibly due to differences in how the individual vines are growing. If the treatment-induced variation is the main reason a difference in treatment means are seen, then that indicates that the treatments themselves are causing a response.

An ANOVA is typically used to test for differences in experiments with more than two treatments (or more than one treatment and a control). Note that the analysis says nothing about whether that response or difference is desirable or undesirable; it simply tests whether the response is likely to be statistically significant.

For example, in the fertilizer application trial (Table 2), variation within is the variation in yield due to pre-existing conditions across a block/plot, whereas variation among/between is the variation in yield due to different rates of fertilizer. An ANOVA calculates the ratio between these two sources of variation (variation among:variation within). This ratio is called the "*F*-ratio" or "*F*-value."

If the *F*-value is large, then we say that the treatment had a "significant effect." If the *F*-value is small, the treatment did not have an effect, and any apparent differences were due to those other sources of variation. In many cases, along with the *F*-value, ANOVA outputs will also provide a *p*-value.

Calculating an ANOVA and *F*-value by hand is cumbersome. However, these can be easily calculated in micro-processing spreadsheets (see *Using Micro-Processing Sheets for Statistics*).

Interpreting an ANOVA: If the calculated *F*-value (the ratio of the variation between treatments and the variation within treatments) of an experiment is larger than the so-called "standard *F*" or "critical *F*" value, this indicates that at least one of the treatments in the experiment is causing a different response relative to another treatment or relative to the control—a real or 'significant' difference. Critical *F*-values are found in tables and are presented in a similar way as *t*-values (Little and Hills 1975).

Because ANOVA is generally used for experiments with more than two treatments (or more than one treatment and a control), additional statistical tests are needed to determine how these treatments differ. Such tests are termed *post-hoc* tests; examples include Fisher's LSD test, Tukey's HSD test, or Duncan's test, among others. A treatment of these tests is beyond the scope of this manual, but they are often provided as an automatic ANOVA output by statistical software packages and micro-processing spreadsheets.

Table 3. The t distribution values at degrees of freedom for typical vineyard trials at two-tail *p*-values of 0.05 and 0.1 (Little and Hill 1975).

Degrees of freedom	Highest t value for obtaining a *p*-value of:	
	0.05	0.1
1	12.706	6.314
2	4.303	2.92
3	3.182	2.353
4	2.776	2.132
5	2.571	2.015
6	2.447	1.943
7	2.365	1.895
8	2.306	1.86
9	2.262	1.833
10	2.228	1.812
11	2.201	1.796
12	2.179	1.782
13	2.161	1.771
14	2.145	1.761
15	2.132	1.753
16	2.121	1.746
17	2.11	1.74
18	2.101	1.734
19	2.093	1.729
20	2.086	1.725
Larger than 30	2.04	1.645

Issued by Washington State University Extension and the U.S. Department of Agriculture in furtherance of the Acts of May 8 and June 30, 1914. Extension programs and policies are consistent with federal and state laws and regulations on nondiscrimination regarding race, sex, religion, age, color, creed, and national or ethnic origin; physical, mental, or sensory disability; marital status or sexual orientation; and status as a Vietnam-era or disabled veteran. Evidence of noncompliance may be reported through your local WSU Extension office. Trade names have been used to simplify information; no endorsement is intended. Published February 2017.