Proc. Fla. State Hort. Soc. 128:86–89. 2015.

Relevance of Epidemiology to Identifying Huanglongbing Resistance: The Power of Experimental Design

Daniel J. Anco*[1,2] and Tim R. Gottwald[2]

[1]North Carolina State University, National Science Foundation Center for Integrated Pest Management, 1730 Varsity Drive, Suite 110, Raleigh, NC 27606

[2]USDA–ARS, U.S. Horticultural Research Laboratory, 2001 S. Rock Road, Fort Pierce, FL 34945

Additional index words. sample size, power analysis

Biological phenomena are influenced by numerous factors and interactions. As such, their observation as affected by different treatments often takes on a distribution of responses, the perceived form of which depends on aspects of experimental design. If sampling sizes or replicates are too few, then misleading conclusions may ensue, since relatively limited data present only a slice of the full range of responses that, under varying conditions, an individual treatment might contribute toward. An analysis involving simulated subsampling of actual huanglongbing data was conducted to illustrate the effect of varying sample sizes and replicates on results. At one end of the spectrum, increased sample sizes (while maintaining only one replicate) increased the rate of significantly different ($\alpha = 0.05$) estimates of disease incidence under one treatment as compared to the control (complete sampling: 150 trees per treatment, three replicates) to ~33%. Conversely, with a fixed per treatment sample size of 10 trees, estimates of disease incidence were respectively up to 75%, 40%, or 25% different from complete sampling estimates when one, two, or three replicates were utilized. Thus, too few replicates or too few samples per replicate can lead an investigator to infer apparent differences among treatments when larger sample sizes and/or more replicates would demonstrate a lack of statistical difference. Though the analyzed data were based on the effects of various control strategies on development of huanglongbing disease incidence, results are analogously applicable toward alternative investigations, such as evaluation of resistant lines.

As most citrus growers know too well, huanglongbing is a serious problem. In order to increase certainty of effectiveness of disease management tools such as durable resistance, it is important for their experiments to be conducted in a manner that yields robust results. The purpose of this study was to assist citrus growers and scientists in addressing issues of experimental design with particular reference toward evaluating citrus lines for huanglongbing resistance. While experimental designs can be limited by physical, logistical, or monetary constraints, occasionally samples sizes and number of replicates are determined by selecting "round" numbers (e.g., 10 samples or 2 replicates). While these practices can be commonplace and can under certain situations (i.e., those with minimal variation) yield representative results, allowing statistical theory to inform experimental design is more scientifically sound, less subjective to chance, ensures a given experiment performs with the desired level of efficiency, and avoids the possibility of drawing unsound, statistically invalid conclusions.

Materials and Methods

This study utilized simulation to illustrate the effect of varying sample sizes and number of replicates on experimental conclusions. The data that were analyzed were actual huanglongbing disease incidence data collected from a commercial citrus orchard in Collier County, FL. The original experiment compared five treatments for management of huanglongbing utilizing a randomized complete-block design. Treatments were replicated in space three times, with each replicate containing 150 trees. Descriptions of actual treatments are omitted from the current paper to avoid unnecessary distraction.

Simulation was performed by randomly sampling without replacement 2 to 146 tree samples per treatment for scenarios involving 1, 2, or 3 replicates. Each sampling scenario was conducted with 1000 iterations. Linear models (for simulated experiments with one replicate) and linear mixed models (for simulated experiments with more than one replicate) (Littell et al., 2007) were fit to resulting data to determine presence of significant treatment differences, and comparisons of estimated treatment effects (BLUPs) were made to estimates from the complete dataset by comparing 95% confidence intervals. Power analyses were conducted as previously described (Stroup, 2011) utilizing variance terms estimated from the complete dataset (Table 1) and allowed examination of scenarios including more samples or replicates than originally present in the utilized dataset. All analyses were conducted in SAS (SAS 9.4, Cary, NC).

Results and Discussion

Analysis of the complete dataset determined the treatments were not significantly different ($P = 0.4783$) (Table 1). The effect of varying sample sizes and replicates on concluding significant treatment differences can be seen in Fig. 1. With one replicate, increased sample sizes increased chances of conclud-

*Corresponding author. Email: danco@clemson.edu; phone: 803-284-3343. Current address: Clemson University, Edisto Research and Education Center, 64 Research Road, Blackville, SC 29817

Table 1. Analysis of complete dataset of treatments evaluated for managing huanglongbing.

Effect[z]	Test statistic value[y]	*P*	Estimate[x] (SE)
Treatment	0.94	0.4783	…
Treatment x replicate	–	–	0.01962 (0.008523)
Replicate	–	–	0.02221 (0.02161)
Residual	–	–	0.1848 (0.005563)
Treatment 1	5.65	0.0002	0.6769 (0.1198)
Treatment 2	4.57	0.0010	0.5477 (0.1198)
Treatment 3	5.27	0.0004	0.6313 (0.1198)
Treatment 4	6.25	< 0.0001	0.7491 (0.1198)
Treatment 5	4.78	0.0007	0.5732 (0.1198)

[z]Effects are a fixed effect (Treatment), covariance parameters (Treatment x replicate, Replicate, and Residual), and individual treatments (Treatments 1–5).
[y]Test statistics are *F* values (Treatment) or *t* values (Treatments 1–5).
[x]Estimated variance (covariance parameters) or huanglongbing disease incidence (all others).

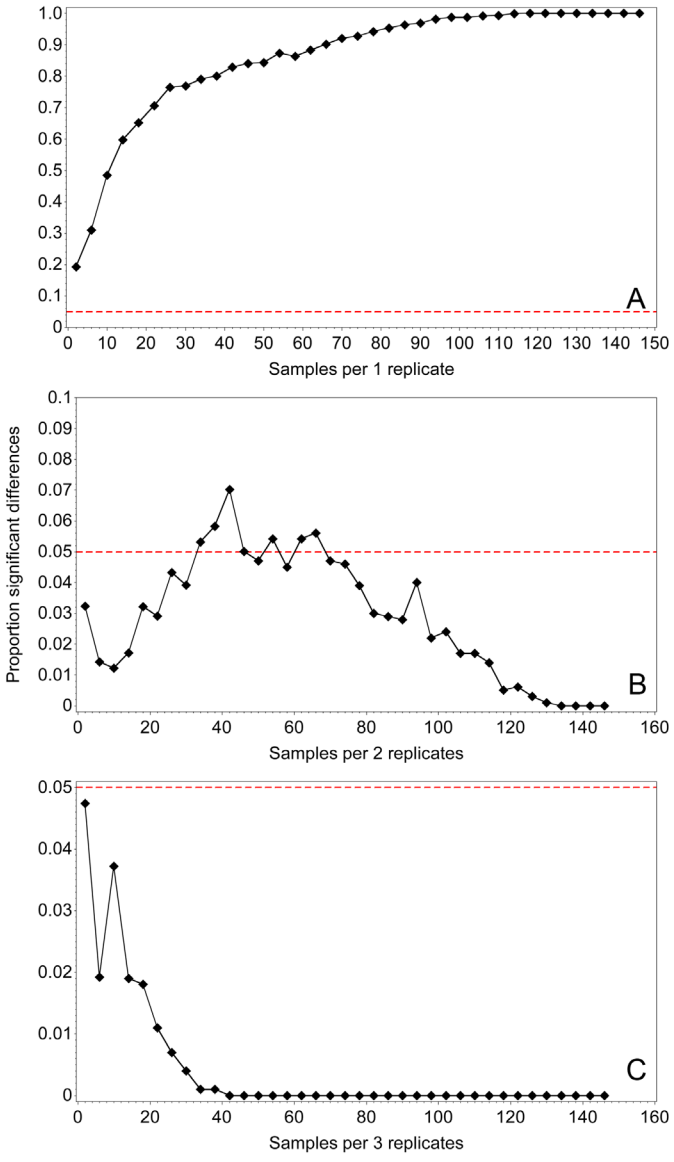

Fig. 1. Effect of sample size on proportion of times experiments concluded significant treatment differences with one (**A**), two (**B**), or three (**C**) replicates. Each point represents results from 1000 iterations. Red reference lines indicate the commonly chosen alpha of 0.05 (Type I error rate: the chance of an experiment if repeated many times concluding significant differences when no true significant differences exist).

ing significant differences ($\alpha = 0.05$). This is because greater sample sizes increased precision of the estimated mean of each treatment-replicate (Fig. 2), which consequently made it easier to detect differences between individual treatment-replicates. With two replicates, increased sample sizes up to 42 samples per replicate increased chances of concluding significant differences, with further increases in sample size resulting in greater frequencies of concluding no significant differences. In the presence of three replicates, increased sample sizes clearly corresponded to an increased proportion of times concluding no significant differences, with 100% of simulated experiments concluding no significant differences after ≥ 42 samples. These points confer how too few samples or replicates can produce spurious results, as well as how the effect of increased sample size stabilized with increasing number of replicates (at least three).

Figure 2 illustrates how increased sample sizes in experiments with one replicate can actually increase the chance of obtaining incorrect (significantly different) treatment estimates as compared to the mean from the complete dataset, with up to ~33% of simulated experiments having produced significantly different estimates of Treatment 1 when sample sizes approached 146. This demonstrates how means of some replicates are closer to the collective mean of all replicates than that of other replicates. The unknown difference of individual treatment-replicates' means compared to the true mean before an experiment is conducted is a primary motivating factor for replicating treatments multiple times. Along those lines, when a fixed per-treatment sample size of 10 trees was examined, estimates of disease incidence were respectively up to 75%, 40%, or 25% different from complete sampling estimates when one, two or three replicates were utilized, based on 95% confidence interval differences.

How then do we determine how many samples or replicates are adequate? The old saying commonly attributed to Francis Bacon, "knowledge is power", here is quite fitting. In the experimental sense, power refers to the ability of a study to detect meaningful effects that exist (more formally, the probability of avoiding a Type II error, which is the chance of failing to reject the null hypothesis when the null hypothesis is in fact incorrect) (SAS Institute Inc., 2011), and it depends on sample size, number of replicates, experimental design, replicate variation, and the magnitude of difference to be detected (Fig. 3). While the power of a given study depends on several factors, it can fortunately be estimated before an experiment is conducted, provided estimates of experimental variance (e.g., of replicates) are available (Stroup,

2011). In order to obtain a power of at least 80% in the examined dataset, ≥ 30 samples would be needed in each of three replicates to detect a difference of 0.25, and ≥ 50 samples would be needed to detect a difference of 0.20. As is shown in Table 2, increasing the number of replicates can actually decrease the total number of samples required for a given experiment to perform with a specific level of power. Compared to basing sample size and number of replicates for an experiment off of "round" numbers, it is better to let intentional experimental design by way of power

Table 2. Samples required to detect selected differences with experimental power of ≥ 80%.

Difference of mean from 0.65 disease incidence	Number of replicates		
	2	3	4
	Samples per replicate (total per experiment)		
0.15	368 (736)	106 (318)	64 (256)
0.20	122 (244)	50 (150)	33 (132)
0.25	66 (132)	30 (90)	21 (84)

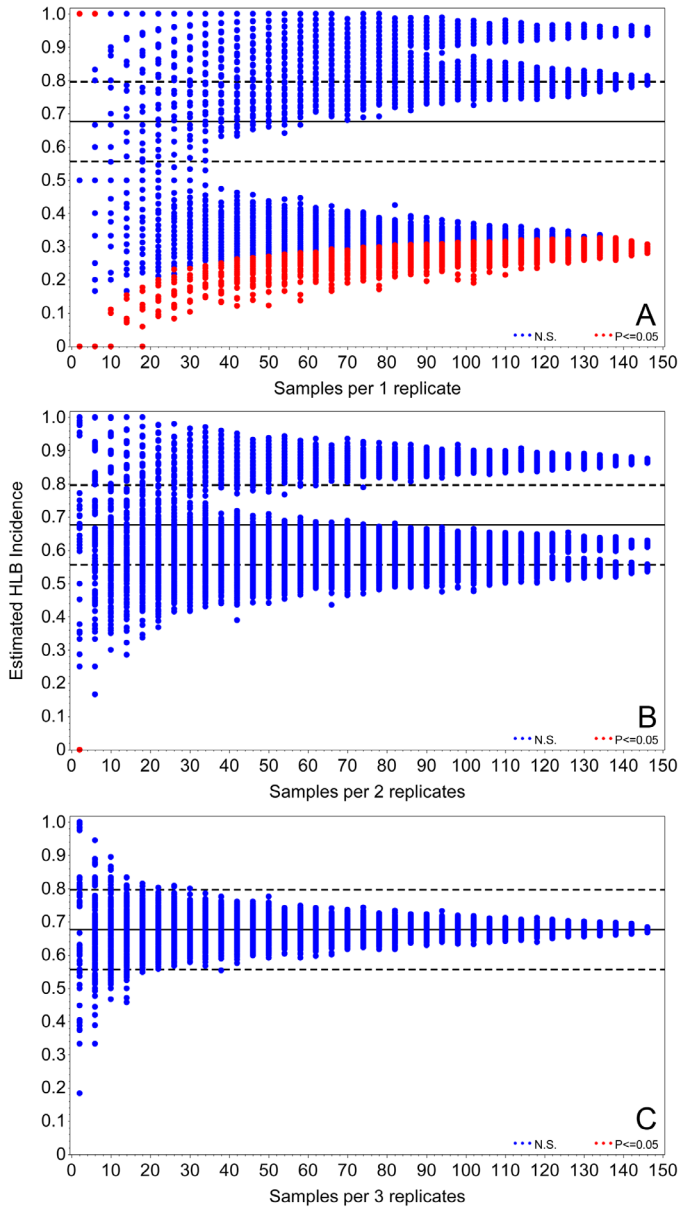

Fig. 2. Effect of sample size on variability of estimated means of huanglongbing (HLB) disease incidence from experiments with one (**A**), two (**B**), or three (**C**) treatment replicates. Plots represent distributions of possible estimates from individual (**A**) or combinations of (**B, C**) replicates of Treatment 1 as an example. Each sample size contains results from 1000 iterations. Red dots are significantly different from the mean estimated from the complete dataset based on 95% confidence intervals, whereas blue dots are not significantly different. Solid reference lines indicate the mean of Treatment 1 estimated from the complete dataset, with the dashed reference lines indicating its standard error.

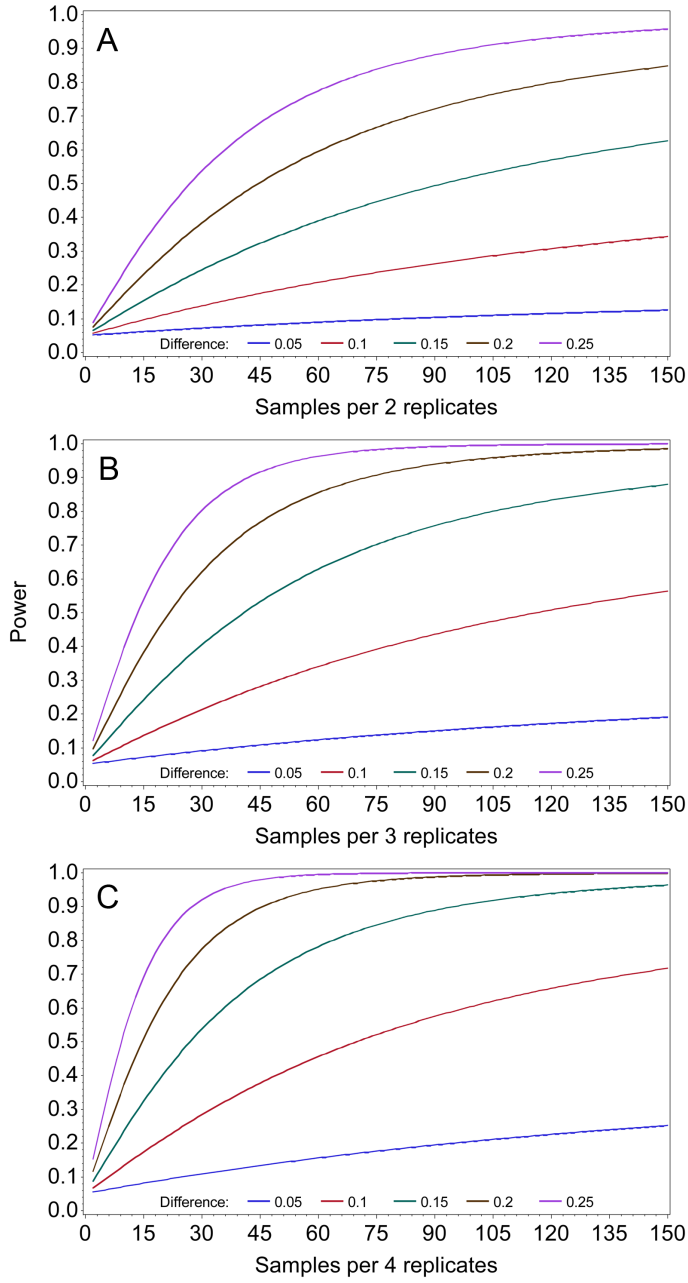

Fig. 3. Relationship of sample size, number of replicates, and power of an experiment to detect treatment differences ranging from 0.05–0.25. Results were calculated using variance terms estimated from the complete dataset and represent experiments with two (**A**), three (**B**), or four (**C**) replicates.

analysis inform sample size and number of replicates. This allows for the most efficient use of resources and results in more effective studies whose conclusions are increasingly robust. In the context of examining lines for huanglongbing resistance, this translates into lines classified as being resistant having been evaluated with appropriate rigor, which in turn provides more sustainable management tools for growers.

Literature Cited

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2007. SAS system for mixed models, 2nd ed. SAS Inst. Inc., Cary, NC.

SAS Institute Inc. 2011. SAS/STAT 9.3 user's guide. SAS Inst. Inc., Cary, NC.

Stroup, W.W. 2011. Living with generalized linear mixed models. SAS Global Forum 2011. Paper #349. 18 p. <http://support.sas.com/resources/papers/proceedings11/349-2011.pdf>