Vigor or Rigor? The Competing Goals of Variety Trials

Anusuya Rangarajan

Additional index words. cultivar, evaluation, extension, vegetables, experimental design

Summary. Both growers and vegetable seed companies have had long-term historic relationships with public agriculture extension educators and faculty to conduct unbiased evaluation of vegetable varieties. Reductions in both the number of vegetable seed companies as well as university human resources has led to questions about the viability and appropriateness of publicly-funded variety evaluation programs. Field based extension educators and regional staff have taken more leadership to evaluate varieties, but this often results in fragmented or repetitive trials with limited long term integration of data. Statewide vegetable extension specialists must provide the leadership in coordinating these trials to enhance the rigor of data collection and analysis. Fundamental to enhancing rigor is improving regional coordination and collaboration. The calculation of stability estimates for new and older varieties is most efficiently and quickly achieved through regional collaborations. Initial efforts should improve uniformity of trials by creating common evaluation methods for yield and qualitative evaluations (e.g., color, appearance), including two standard varieties (one local and one regional, long-term standard), standardizing field establishment practices, and selecting experimental designs and plot sizes to improve labor efficiency. These regionally coordinated trials will improve the ability to publish this type of applied research and demonstrate new levels of efficiency for university administrations. In the long term, carefully designed comparisons of genotypic performance among different environments could suggest new directions for university breeding programs as well as cropping systems research.

Variety evaluation has long been an essential activity of agricultural extension efforts in vegetable crops. Growers trust the unbiased evaluation conducted by public institutions, and seed companies learn how particular climatic and market conditions affect acceptability of varieties. For faculty working in vegetable production, variety trials provide quick research results and excellent extension opportunities to engage growers. However, university administration and department chairs reiterate the opinion that this research is not rigorous and therefore, may not enhance tenure and promotion packages. In addition, there is no legislated mandate to provide these evaluations as is done with pesticides. With the decreasing human capital supporting applied production research, many college and university administrators and faculty question whether variety trials are a wise use of resources. In reviewing papers published from the 1986 ASHS symposium, “Vegetable Cultivar Testing” (Price and Zandstra, 1987; Thomas, 1987; Wehner, 1987), one is reminded that the issues surrounding variety evaluations are not new. In 1986, speakers remarked on the strengths of the public trust, but outlined the struggles with competing goals, loss of personnel, and low value to university administration. Since 1986, new and perhaps unexpected factors have negatively affected the implementation of effective variety evaluation programs. I will review some of these hindering factors and then offer strategies for university extension faculty to enhance the rigor of these trials while still providing grower-valued vigor assessments.

Department of Horticulture, Cornell University, Ithaca NY 14853.

Current leadership of extension variety evaluations

Extension educators and researchers based in counties or at regional stations have increasingly taken the lead for variety evaluations, moving the activity away from campus and university-based staff. The benefit is that growers have more information on local adaptability and educators strengthen their ties with growers. However, these educators usually have no mandate to conduct trials that have an impact beyond their home county or local region. Crop selection and trial design become locally isolated. This diffuse type of variety evaluation results in a lack of uniformity of trial design and poor coordination in selection of standard varieties and evaluation strategies, minimizing the opportunity to enhance rigor of the results and, perhaps, the real long-term value to the industry.

Changes in the vegetable seed industry

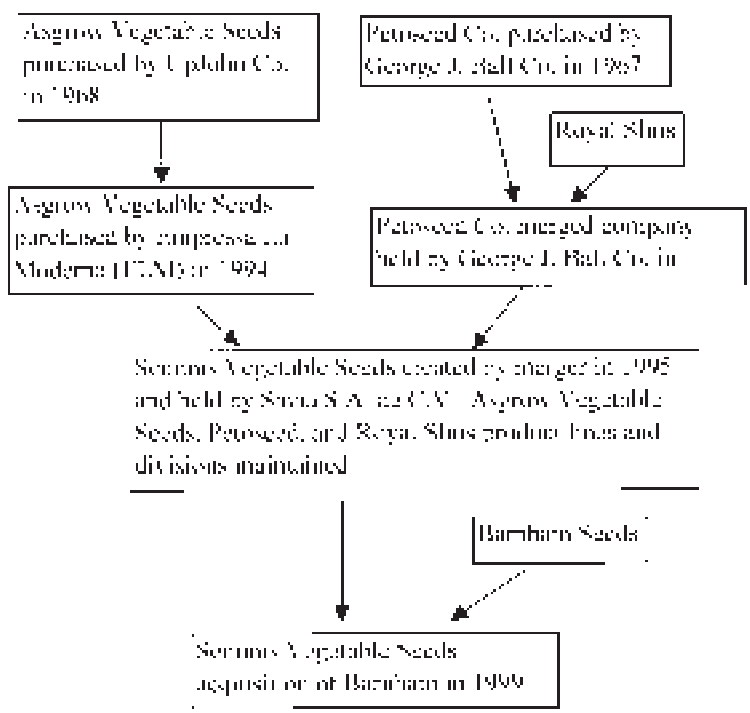
What may not have been expected at the 1986 ASHS variety trial symposium were the accelerated changes in the seed industry over the last 14 years, particularly the changes in investment in biotechnology. For example, consider the Asgrow-Petoseed company complex (Fig. 1). This genealogy was adapted from the full history of the commercial vegetable seed industry as documented by James Baggett at Oregon State University (unpublished data). Only those consolidations that have occurred since the 1986 workshop are provided in Figs. 1 and 2. Seminis Vegetable Seeds (Oxnard, Calif) was created in the mid-1990s from the merger of three separate companies, Asgrow Vegetable Seeds (Ventura, Calif.), Petoseed Co. (Ventura, Calif.) and Royal Sluis (Ventura, Calif.) (Fig. 1). Since 1986, five seed companies, Sunseeds Co. (Morgan Hill, Calif.), Northrup King (Syngenta Seeds, Golden Valley, Minn.), Rogers Brothers Vegetable Seeds (Syngenta Seeds, Boise, Idaho), Pioneer Vegetable Genetics (Sunseeds Co., Morgan Hill, Calif.) and Nunhems Zaden (Haelen, The Netherlands) have merged into two companies (Fig. 2). The result of these mergers across the seed industry has reduced the number of companies actively involved in variety testing and increased the pressure on company representatives to conduct field trials. These representatives also must prioritize where they will support public variety trials based upon their own reductions in budget and personnel. Thus, the constraints to variety evaluations are coming from both the public and private sectors.

Strategies to enhance rigor of variety trials

In attempting to increase the rigor of our variety trials, we must examine our decisions in designing trials. These include the selection of cropping systems, plot size, evaluation criteria, and integration over locations and years. Most often we report single year and site evaluations to growers. While lines may be evaluated over several years, and integration of these data is conducted in an informal manner, and analysis and publication of multi-year performance data are rare but increasing (Cushman and Simonne, 2002). A natural role for campus-based specialists is to assist with coordination and implementation of variety trials across multiple locations or several years. Modifying trial designs to facilitate integration of performance data across years, locations, and potentially farming systems presents a viable strategy for increasing rigor of these experiments and utility to seed companies (Thomas, 1987). It would also allow us to explore genetic by environment interactions. This integration is essential to enhance opportunities for multidisciplinary research with breeders, pathologists, entomologists and even human nutritionists on adaptation of these materials across environments.

Some fairly basic experimental design issues require consensus among collaborators, including careful definition of the target market for the crop to be evaluated and subsequent desirable crop traits. This integration of market traits is commonly illustrated through the selection of standard varieties for comparison both within and across environments and years. Inclusion of two standard varieties may be the minimum to enhance long-term integration of data. One standard, perhaps termed the long-term research

Fig. 1. Three companies, Asgrow Vegetable Seeds (Ventura, Calif.), Petoseed Co. (Ventura, Calif.), and Royal Sluis (Enkhuizen, The Netherlands) merged in 1995 to form one company, Seminis Vegetable Seeds (Oxnard, Calif.). Barnham Seeds (Gilroy, Calif) was acquired by Seminis in 1999. Savia S.A. de C.V. (Monterrey, Mexico) holds majority interest in Seminis. Other firms that have held these seed companies since 1986 include UpJohn Co. (Peapack, N.J.), and George J. Ball Co. (Chicago, Ill.). This figure was adapted and updated from the full history of the commercial vegetable seed industry by J.R. Baggett, Oregon State University, Corvallis.



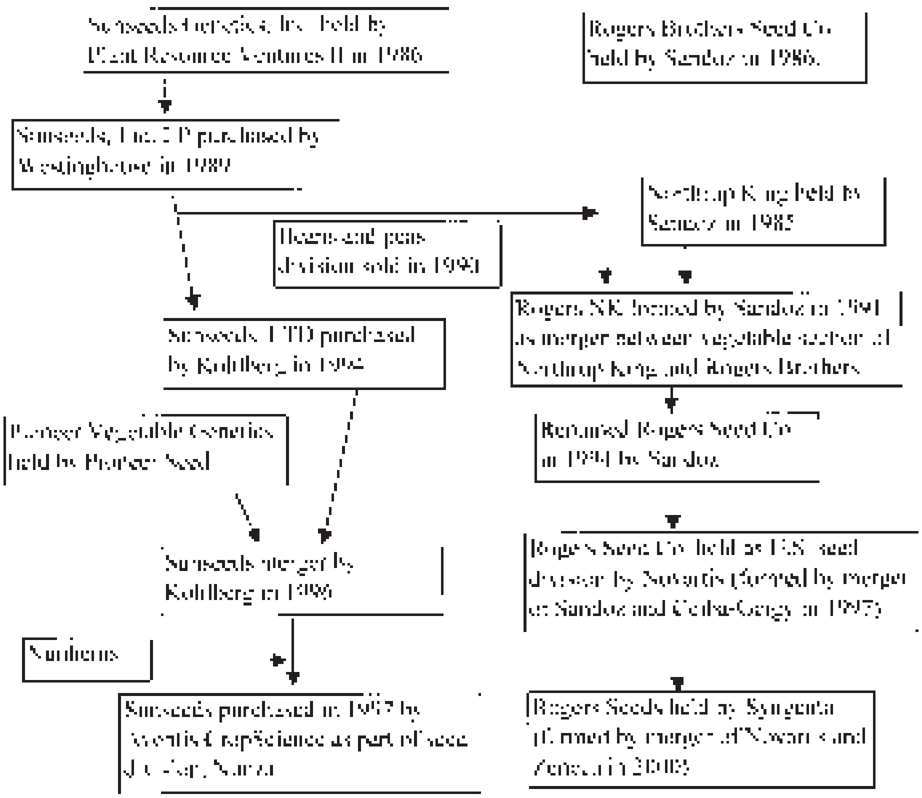


Fig. 2. Since 1986, five seed companies, Sunseeds Co. (Morgan Hill, Calif.), Northrup King (Syngenta Seeds, Golden Valley, Minn.), Rogers Brothers Vegetable Seeds (Syngenta Seeds, Boise, Idaho), Pioneer Vegetable Genetics (Sunseeds Co., Morgan Hill, Calif.) and Nunhems (Haelen, The Netherlands) have been merged and held by two companies. Several other holding companies, including Sandoz (Basel, Switzerland), Kohlberg Investment Group (location unknown), Plant Resource Ventures II (investment group, location unknown), and Westinghouse (Pittsburg, Pa.) have acquired and sold these seed firms since 1986. Sunseeds is now held by Aventis CropScience (Lyon, France) as the U.S. division of their seed subsidiary, Nunza (Haelen, The Netherlands). Nunhems is also retained as the brand for sales to Europe, Asia, the Middle East and Africa. Rogers Brothers Vegetable Seeds is now held by Syngenta Seeds and responsible for the seed division in the U.S. and S&G Seed Co. (Basel, Switzerland) targets Europe, Asia and Africa. This figure was adapted and updated from the full history of the commercial vegetable seed industry by J.R. Baggett, Oregon State University, Corvallis.

standard, would be a variety produced historically, for which there already is multi-year performance data. This standard provides a baseline to compare environmental conditions as well as yield improvements provided by new genotypes over time. The second standard, termed the industry standard, would be the current regional or local standard as viewed by the industry. Some researchers advocate the inclusion of one poor performer (prone to poor quality under certain conditions) to assess local environmental potential. In the northeastern U.S., the chipping potato (*Solanum tuberosum*) variety Katadin is included as a standard in trials because it is prone to internal necrosis (J. Sieczka, personal communication). By agreeing upon some of these types of standards, integrated data sets would be more rigorous for identifying poor performers or optimum environments for particular genotypes. While these standards will enhance rigor of the experiment, gains must be balanced with increased costs.

In New York, the processing vegetable industry has had a long history of supporting public evaluation of crops. Selection of standards for the long term is particularly difficult for some crops, such as sweet corn (*Zea mays*), where new genotypes are constantly being introduced. Over the last 10 years, the standard varieties for these trials have changed. In 1990, 'Jubilee', 'Reward' and 'Stylepack' were the standards. In 1994, this changed to 'Reward', 'Rival' and 'More'. In 2000, the standards were 'Rival', 'Bonus' and 'Sprint'. While the standard changed over time, continuity does exist through overlap of these standards. In snap beans (*Phaseolus vulgaris*), 'Labrador' has been a standard for 10 years, providing an excellent opportunity for analysis of yield stability over time. This selection of standard varieties does depend upon those genotypes being stable through seed production cycles. In cabbage (*Brassica oleracea* Capitata group), some have questioned if varieties have maintained their characteristics when grown in similar environments over multiple years (R. Becker, personal communication). Integrated data analysis, over multiple years, could provide an answer to the perennial question: Was it the seed?

Selecting plot size is an art that balances resources (land, labor, equipment, costs) plant compensational ability and nature of border effects, local soil heterogeneity, yields needed for post-harvest processing, and experimental error. In most cases, we seek a plot size that can be conveniently managed and allows some flexibility for locating harvest area, in case of stand losses. Using this approach, plots are commonly 50% longer than the area to be harvested, providing adequate buffers and room to adjust data rows. Square shaped plots are preferred, to reduce variation among plots within a block (Wehner, 1987). Buffers are included around the data area when the crop yield is affected by adjacent rows, such as with tall or short sweet corn. Bell peppers (*Capsicum annuum*) and most other short-stature vegetable crops, on the other hand, are commonly tested using single row plots, with buffers at plot or row ends. With some crops, particularly cucurbits (Cucurbitaceae) or other vining crops, use of a phenotypically different variety with similar growth habit in buffer areas helps reduce harvesting errors and improve harvest efficiency for labor.

Large plots can have higher experimental error since more land area must be used. Small plots can also have high variability for several reasons, such as 1) harvest or recording errors from small plots may have a large effect on final yields, 2) plant to plant variation may be high, 3) plant losses may contribute to greater error, particularly for plants that are sensitive to border effects, and 4) competition or border effects may be larger in small plots (Petersen, 1994). Different strategies have been suggested to estimate minimum plot size needed to minimize variance per plot based upon soil heterogeneity, error estimates from previous experiments and costs of the trial (Petersen, 1994; Sills and Nienhuis, 1993; Smith, 1938). Developing these types of localized estimates for vegetable experimental plot sizes would improve the efficiency of testing and rigor of experimental design.

Results from previous experiments can be used to determine the number of replicates needed to detect predetermined differences in yield. Statistical programs, such as SAS (SAS Institute, Cary, N.C.) and Minitab (Minitab Inc, State College, Pa.), can integrate power, standard error and desired differences to provide a recommended replication number needed to detected that difference. It is important to remember that precision of a trial is usually improved faster by increasing the number of replicates, not increasing plot size (Petersen, 1994) and general performance information is optimized by increasing the number of locations, not the number of replications within a location (Talbot, 1997). Because information gained on variety performance from multiple sites is more useful than from individual sites, two replicates per location is commonly adequate. One can examine the variance components to optimally allocate resources to plot size, replication within location or multiple locations.

Crop performance measurements, such as yield or size, are almost always affected by cultural practices used in the trial. All reasonable efforts should be made to standardize production and harvesting strategies, particularly plant populations and maturity indices used to initiate harvest. Measurements such as yield per plant or per acre, nutritional composition expressed with both fresh and dry weight, or nutritional quality analysis following nutritional science standards (mg/100 g fresh weight) can facilitate comparison of data across sites, if production practices are similar. In cases where production practices are different, performance of the new variety could be compared to standard varieties. Unusual spatial differences not accounted for in the design of blocks may be addressed through covariance analysis. These types of differences may include losses in plant stand or yield due to pests, unanticipated hedge effects or other environmental heterogeneity that can be scored. Use of covariance analysis can reduce the standard error of differences among means and affect the relative ranking of varieties in trials (Mean, 1997). However, the researcher must verify that the factor to be adjusted (e.g., yield) with covariance analysis is not actually due to other genetic traits (e.g., pest resistance).

For most fresh market vegetable crops, there are usually more varieties of interest than can be effectively managed. One strategy is to evaluate new, often numbered, lines using only one or two replicates in the first year. If promising, these lines can then be included in full replicated evaluation. If a large number of entries is unavoidable, then the experimental design should be carefully selected. The randomized complete block design is the most common experimental design used in variety evaluations, but should only be selected if there is clear indication of a gradient of some type at the trial location. A completely randomized design may be suitable in cases where blocking is not justified, or in certain controlled environments.

Experimental designs underused in vegetable variety evaluation programs include the incomplete block and lattice designs. These designs may reduce standard error and increase labor efficiency, depending on the number of lines being tested (Petersen, 1994). When a large number of varieties are to be tested, the blocks of a randomized complete block design may become quite large, reducing their effectiveness for improving precision as well as resulting in high costs. Incomplete block designs (ICBD) can help reduce block size, by not including all varieties in every block (Mean, 1997). Generally, varieties are arranged to allow an equal number of direct comparisons for all pairs of varieties. An example ICBD is the lattice design, in which the number of varieties must be a perfect square (4, 9, 16, etc.), so that the number of incomplete blocks and the number of plots in a block are equal to the square root of the number of varieties. Each variety appears in a block with every other variety the same number of times. Because the relative size of each block is decreased, the standard errors can be reduced, improving precision, with these lattice designs compared to the randomized complete block.

Yield, uniformity and quality stability are more important than the "best in show for 2001." Phenotypic stability for a variety is evaluated by comparing traits over a range of environments or years. Implicit in considering stability is the concept of ranking—good varieties consistently perform well under certain conditions compared to other varieties. Some traits, such as fruit number in cucumbers, have been observed to be more stable than yield (Ells and McSay, 1981), suggesting careful consideration of trait to be evaluated across time or location. Several of the stability analyses commonly used by plant breeders may help evaluate the types of conditions under which varieties perform best, which varieties are least affected by environment, and those most responsive to high-yielding environments (Lin et al., 1986; Pritts and Luby, 1990). Collective wisdom gained from long-term agronomic trials suggests that year-to-year effects are greater than location effects for stability analysis (Talbot, 1997). Thus, three trials run over 3 years will provide more information on performance than one trial run in nine locations for 1 year. The challenge to the vegetable industry is to balance evaluation of new varieties with assessment of longtime good performers. In some crops, a 3-year evaluation cycle would produce results too late to be of use to growers, or varieties evaluated may no longer be important to the industry. If state specialists were able to coordinate variety selection for trials to allow a certain number to be compared across several locations within a region, calculation of these stability factors could occur in a more timely fashion and could greatly enhance utility to both growers and seed companies.

There are excellent examples of regional integration available for us to model, including the Northeastern Region Potato Variety Trial (NE 184) (Maine Agricultural and Forest Experiment Station, 2001). This multidisciplinary regional effort evaluates new breeding lines and cultivated potatoes across northeastern states and eastern Canadian provinces. Their objectives include developing and evaluating pest resistant, early maturing, long dormant and specialty varieties and determining climatic effects on performance. These researchers have standardized their qualitative and quantitative ratings of both plant and tuber characteristics, and include the same standard varieties (Atlantic or Superior) so that their results can be analyzed as a group, across the region. One statistician integrates data across trials to provide stability factors and genetic by environment evaluation of variety performance. These results are published annually for grower and industry assessment of new and old variety performance.

Conclusions

The opportunity to increase the rigor of variety trials lies with the statewide extension specialists. The collaborations needed to achieve regional integration will, in all likelihood, have to proceed with no additional outside funding. These regional collaborations and multi-year evaluations will, however, provide the rigor and the impact we desire: 1) improved assessment of genetic stability will more thoroughly describe variety performance, 2) these multi-location and multi-year evaluations can be published, 3) growers are provided with improved information for variety performance across a range of environments, and 4) comparison local to regional variety performance could suggest areas for more multidisciplinary research. In addition, regional collaboration is recognized as positive by university administrations looking to highlight efficiency.

In the northeastern U.S., we are increasingly evaluating crops for niche markets, including specialty ethnic or certified-organic markets. Organic production systems can provide a unique environmental assessment of variety performance and pest resistance, since traditional pest control options are limited. Currently, we are examining maturity and yield of sweet corn in organic certified production systems. While maturity, yield and quality are standard evaluation criteria, we are also assessing nitrogen uptake efficiency, to further understand crop performance and selection for this farming system. We have included several breeding lines that are either adapted to low nitrogen environments, heavy nitrogen users, or nitrogen stable (or nonresponsive), to define future breeding strategies for organic systems. In addition, we are exploring alternative on-farm evaluation strategies (using few replicates across many sites) to refine strategies for variety trials of niche crops. Other examples may include examining crop variety performance in no-till or reduced tillage systems, in which emergence and stand establishment are more challenging than in conventional tillage systems, or selection of varieties to serve other roles, such as trap crops or refuges for beneficial insects. As we look to the future, new crops or cropping systems may become more important focal points for some of our variety evaluation programs as we try to anticipate changing markets or production environments for vegetables.

Literature cited

Cushman, K.E. and E.H. Simonne. 2002. Beyond station reports: Conducting scholarly vegetable variety trials and publishing the results: Workshop introduction. HortTechnology 12(4):552.

Ells, J.E. and A.E. McSay. 1981. Yield comparisons of pickling cucumber cultivar trials for once-over harvesting. HortScience 16:187–189.

Lin, C.S., M.R. Binns, and L.P. Lefkovitch. 1986. Stability analysis: Where do we stand? Crop Sci. 26:894–900.

Maine Agricultural and Forest Experiment Station. 2001. Performance evaluations of potato clones and varieties in the northeastern states–1998. Misc. Publ. 747. Maine Agr. For. Expt. Sta., Orono.

Mean, R. 1997. Design of plant breeding trials, p. 40–67. In: R.A. Kempton and P.M. Fox (eds.). Statistical methods for plant variety evaluation. Chapman & Hall, London.

Petersen, R.G. 1994. The field plot, p. 19–34. In: R.G. Petersen (ed.). Agricultural field experiments: Design and analysis. Marcel Dekker, New York.

Price, H.C. and B.H. Zandstra. 1987. Cultivar testing: Public point of view. HortScience 22:1211–1212.

Pritts, M. and J. Luby. 1990. Stability indices for horticultural crops. HortScience 25:740–744.

Sills, G.R. and J. Nienhuis. 1993. Field plot technique affects snap bean yield evaluation. J. Amer. Soc. Hort. Sci. 118:672–674.

Smith, H.F. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. J. Agr. Sci. 28:1–23.

Talbot, M. 1997. Resource allocation for selection systems, p. 162–174. In: R.A. Kempton and P.M. Fox (eds.). Statistical methods for plant variety evaluation. Chapman & Hall, London.

Thomas, P. 1987. Cultivar testing, A seed industry perspective. HortScience 22:1212–1213.

Wehner, T.C. 1987. Efficient methods for testing vegetable cultivars. HortScience 22:1220–1223.