EXPERIMENTAL DESIGNS FOR COMPARING LANDSCAPE TREE CULTIVARS UNDER CITY CONDITIONS[1]

by W. Donald Bartoe, graduate assistant, University of Florida and Henry D. Gerhold, professor of forest genetics, The Pennsylvania State University, University Park, Pennsylvania 16802.

ABSTRACT.-- Trees in urban environments encounter many unnatural stresses. The effects of restricted growing space, air pollutants, road salts, and other stresses may best be quantified by measurements of height and foliage condition. To develop efficient experimental designs for testing trees in cities, representative age class means and variances were calculated for height, foliage condition, and six other performance traits. Existing plots of 23 different landscape tree cultivars growing in 19 U.S. cities were evaluated. Sizes and arrangements of experiments needed to show that differences of various magnitudes between two cultivars are significant were computed. Several experimental designs which optimize statistical efficiency and precision are recommended. One consists of two sets of plots containing one cultivar each and randomly located in one site category (e.g. residential, business, or industrial) in a city. This design would require at least 24 trees per cultivar and four to eight trees per plot. Other designs which may be more precise are discussed. How tests of trees in cities can be implemented, extended, and augmented by non-urban tests is also considered.

[1] Journal Paper No. 5574 of The Pennsylvania Agricultural Experiment Station, authorized for publication Aug. 17, 1978. Research supported by USDA Northeastern Forest Experiment Station through the Pinchot Institute Consortium for Environmental Forestry Studies.

Trees improve urban environments. They enhance the visual landscape, abate noise, cleanse the atmosphere, ameliorate climate, control erosion, provide recreational sites and wildlife habitat, and increase property values. (Moeller 1976).

A multimillion dollar industry produces landscape trees. New cultivars (cultivated varieties) are bred and mass-produced from seed or more often by budding or grafting, and cultivars are selected, planted, and maintained. But the performance of most landscape tree cultivars has not been tested adequately. A formal urban testing program is needed to provide more detailed knowledge than is currently available about relative landscape tree performance, e.g. growth traits or tolerance to environmental stresses (Gerhold *et al.* 1975; Gerhold and Bartoe 1976; Santamour 1968; Santamour 1971; Steiner and Gerhold 1976; Reisch et *al.* 1971).

STRESSES TREES ENCOUNTER IN CITIES

Trees in urban environments encounter many unnatural stresses. Growing space may be restricted above and below ground. Soils are often compacted and poorly aerated. The atmosphere is polluted. De-icing salts accumulate on aerial and subterranean portions of street trees. Sunlight is blocked by buildings. Ambient temperature is increased by radiant heat from buildings and streets. Vehicular traffic, building construction, road maintenance, and vandals physically damage or destroy landscape trees. Furthermore, unnatural stresses may increase a tree's susceptibility to injury by the natural stresses of weather, disease, and insects.

How, then, can we typify the ability of landscape trees to tolerate the debilitating effects of urban conditions?

METHODS OF TESTING

Landscape tree testing in the past has been virtually limited to non-urban field tests and informal observations in cities (Gerhold and Steiner 1976). Non-urban field tests are not representative of the complex urban planting site (Santamour 1968; Santamour 1971; Steiner and Gerhold 1976). Informal observations of trees in cities enable detection of only the larger, more obvious differences between species or cultivars (Steiner and Gerhold 1976). To detect smaller but important differences between trees in cities, experimental designs must be developed that will provide statistically significant results yet be practical in view of the constraints of urban spaces and managerial practices.

Formal tests of existing tree plantings in cities have been reported by Mower (1973) in New York and Reisch et al. (1971) in Ohio. Detailed comparisons of cultivars when existing plantings are sampled are hampered by differences in planting sites, planting methods, and planting stock origins, and by limited numbers of plots available for measurement. New plantings of trees in cities can be planned to avoid the confounding of environmental and cultivar differences. But few municipal arborists have the resources to plan and implement the needed experiments.

A cooperative performance testing program has been proposed in which municipal arborists, commercial nurserymen, and a coordinator would collaborate (Gerhold and Bartoe 1976). Arborists would plant and measure trees among their regular plantings, nurserymen would provide planting stock through normal channels, and a coordinator would provide guidance in experimental design and analyze the results. Test results would be useful to arborists in selecting planting stock, to nurserymen in developing production and marketing strategies, and to tree breeders in assessing improvement needs. Such a labor-sharing system may make feasible the statistical evaluation and comparison of landscape trees in cities.

The specific objectives of this study were to (1) analyze data from existing urban plantings of landscape trees to estimate sizes, numbers, and arrangements of test plots required to achieve useful levels of precision in a performance testing program; and (2) develop practical urban test planting designs that will provide statistical evidence of performance differences between landscape tree cultivars.

EVALUATING PERFORMANCE

Performance traits like growth rate and disease susceptibility are heavily influenced by environment; appearance traits like flower color and branching habit are relatively stable over a wide range of environments (Steiner and Gerhold 1976). Small genetic differences in performance traits are often useful when selecting among similar cultivars. Only sensitive statistical tests can detect these small performance differences. Appearance trait differences are also important when selecting among cultivars but only when the differences are large and apparent.

Of the many performance traits, height and health of foliage are perhaps the best indicators of a cultivar's tolerance of city conditions. Height is closely related to adaptability and is an important selection criterion. Growth patterns, ultimate size, and general health and vigor in the long run are indicated by height measurements. A cultivar's height growth is affected by its adaptability to environmental conditions, thus height is also a useful index of site quality. Foliage condition is a determinant of aesthetic quality and is a sensitive measurement of short run health and vigor. Signs of stresses encountered by landscape trees are likely to be seen first in the foliage, though cumulative effects of less severe stresses may be reflected only in depressed growth rate Moreover, it is useful when defining the experimental design to consider not just objective traits like total height but subjective traits as well.

EXPERIMENTAL DESIGN

Sizes and Numbers of Plots

Wright and Freeland (1960) expressed experimental efficiency in terms of the numbers of replicates (r, n-tree plots) and total size of experiment (nr) required to show that an arbitrary difference between the means of two cultivars (x_1-x_2) is significant. Assuming each cultivar would have the same number of replicates (plots) and the same variance (V) between n-tree plots, r can be calculated as:

$$(1) \quad r = \frac{2Vt^2}{(\bar{X}_1-\bar{X}_2)^2},$$

where t is Student's t value for n-1 degrees of freedcm (n = trees of a cultivar per plot) and the desired level of significance.

The difference between the means of two cultivars that can be shown to be significant is the least significant difference (LSD = -x). Given a representa tive mean (X) common to both cultivars and an arbitrary fraction (W) of that mean the experimenter wishes to detect, r can be calculated as:

$$(2) \quad r = \frac{2Vt^2}{(W\bar{\bar{X}})^2} \quad ,$$

where V is the representative variance between means of plots containing the same cultivar.

Rather than computing numbers of replicates required to detect an arbitrary LSD it may be more meaningful to determine the LSD detectable for a given number of replicates. Rearranging terms in equation (2), LSD (=WX) can be calculated as:

$$(3) \quad W\bar{\bar{X}} = t\sqrt{\frac{2V}{r}}$$

Representative Means and Variances

Because the proposed performance testing program is meant for broad application, the experimental design must be generalized. Besides, too little information exists to develop refined designs for comparing particular cultivars in particular cities and sites. The idea is to measure existing city plantings of commonly planted cultivars and determine means and variances that are _representative_ of landscape trees in northeast U.S. cities.

Data. Existing plots of 23 different landscape tree cultivars (Table 1) growing in 19 U.S. cities (most in the Northeast) were evaluated. Data also were extracted from a report published by Reisch et _al._ (1971) and from an unpublished 1975 survey of city arborists and landscape specialists conducted by H. D. Gerhold. Trees ranged in age from 1 to 22 years after planting. Two distinct crown forms, spreading and columnar, were recognized and evaluated separately. Data for 19 spreading crown cultivars and 4 columnar crown cultivars were analyzed. 105 spreading crown cultivar plots and 30 columnar cultivar plots were used to analyze total height (feet), trunk diameter at a height of 4 or 4.5 feet (inches), width of live crown (feet), and crown width-to-height ratio. Fifty-seven plots of spreading crown cultivars and 18 plots of columnar trees were used to analyze foliage condition, branch and trunk condition, maintenance needs, and injury incidence. Foliage condition and branch and trunk condition were measured on the continuous scale of 1 = excellent health to 5 = tree dying. Maintenance needs were measured on the continuous scale 1 = very low cost to 5 = very high cost. Injury incidence is number of injury types per tree.

Table 1
Species and cultivars measured in cities

Botanical Name	Common or Cultivar Name
Acer platanoides L.	Norway Maple
	'Crimson King'
	'Cleveland'
	'Columnare'[a]
	'Emerald Queen'
	'Schwedleri'
	'Summershade'
Acer rubrum L.	Red Maple
	'Armstrong'[a]
	'Columnar'[a]
	'October Glory'
	'Red Sunset'
	'Scanlon'
Acer saccharum Marsh.	Sugar Maple
	'Columnare'[a]
	'Green Mountain'
Fraxinus holotricha Koehne.	Baltic ash
	'Moraine'
Fraxinus pennsylvanica Marsh.	Green Ash
	'Marshall Seedless'
Gleditsia triacanthos L.	Honey locust
	'Moraine'
	'Sunburst'
	'Skyline'
	'Shademaster'
Tilia cordata Mill.	Little-leaf Linden
	'Glenleven'
	'Greenspire'
	'Rancho'
Tilia americana L.	American Linden
	'Redmond'

[a]Columnar crown cultivars; other cultivars listed have spreading crowns.

About 50 percent of both spreading and columnar crown cultivar plots were in residential sites, and about 25 percent of each were in business sites. Other plots were in park, highway, industrial, and business/ residential sites.

Representative means. Representative means were chosen for each trait and crown category and at various ages (Tables 2 and 3). First, an age class mean for each trait was computed for species so that constituent cultivars had equal weight. Similarly, an overall mean for each trait and age class was computed for crown types so that constituent species had equal weight. Some computed overall means were adjusted so that the growth traits changed with age in a more systematic manner. Height, trunk diameter, and crown width should follow the typical sigmoid curve as age increases; the chosen representative means for these traits reflect that curve. Crown width-to-height ratio representative means were computed directly from crown width and height representative means. Subjective traits like foliage condition do not have general, predictable functions with age; adjustments of subjective trait means involved only removal of greatly anomalous plots from mean calculations.

Representative variances. Representative variances were chosen for each trait for various age class ranges. The age class ranges for each trait (except crown width-to-height ratio and injury incidence) in which variance was expected to be independent of the means was determined from percent coefficient of variation (CV) values. It was assumed that CV within plots and variances between plots would change similarly. CV for each plot was computed as:

$$\text{CV (plot)} = \frac{\text{Standard Deviation of Plot Mean}}{\text{Plot Mean}}$$

Average CV for age class ranges was computed so it would be representative of the constituent cultivars, species, and age classes (Table 4). A distinct difference in CV within plots between age class ranges 1 to 2 and 3 to 12 was evident for trunk diameter and total height; distinct age class differences were not discernable for the other traits.

Variance between plot means (V(7)) was computed as:

$$V(\bar{X}) = \frac{\Sigma X^2 - (\Sigma X)^2/r}{r-1}$$

Table 2
Representative age class means of cultivars having spreading crowns

Age Class	TRAIT							
	Trunk Diameter (inches)	Total Height (feet)	Width of Live Crown (feet)	Crown Width-to-Height Ratio	Foliage Condition (1 to 5)[a]	Branch and Trunk Condition (1 to 5)[a]	Maintenance Needs (1 to 5)[b]	Injury Incidence (#)[c]
1	1.60	12.0	5.3	0.44	1.30	1.40	1.20	0.50
2	1.70	12.5	5.5	0.45	1.45	1.55	1.80	0.72
3	1.85	13.1	6.1	0.46	1.60	1.70	1.30	0.95
6	2.75	16.1	8.1	0.49	1.20	1.70	1.52	0.60
9	3.80	19.4	10.6	0.55	1.50	1.60	2.62	0.90
12	5.00	21.8	13.3	0.61	1.80	1.90	3.00	1.20
15	5.80	23.4	15.4	0.67				
18	6.55	24.8	17.5	0.72				
21	7.15	26.0	19.3	0.75				

[a]Foliage condition and branch and trunk condition were measured using the continuous scale 1 = excellent health to 5 = tree dying.

[b]Maintenance needs were measured using the continuous scale 1 = very low cost to 5 = very high cost.

[c]Injury incidence equals number of injury types per tree.

Table 3
Representative age class means of cultivars having columnar crowns

	TRAIT							
Age Class	Trunk Diameter (Inches)	Total Height (feet)	Width of Live Crown (feet)	Crown Width-to-Height Ratio	Foliage Condition (1 to 5)[a]	Branch and Trunk Condition (1 to 5)[a]	Maintenance Needs (1 to 5)[b]	Injury Incidence (#)[c]
1	1.18	11.50	1.81	0.16	1.57	1.14	1.78	0.69
2	1.30	11.60	2.15	0.18	1.78	1.51	1.65	0.91
3	1.51	11.90	2.62	0.22	2.00	1.88	1.52	1.15
6	2.26	13.60	4.72	0.35	1.48	1.69	1.13	0.99
9	3.08	16.15	5.67	0.35				

[a]Foliage condition and branch and trunk condition were measured using the continuous scale 1 = excellent health to 5 = tree dying.

[b]Maintenance needs were measured using the continuous scale 1 = very low cost to 5 = very high cost.

[c]Injury incidence equals number of injury types per tree.

Table 4

Average coefficient of variation[a] within plots, all cultivars

Age Class	No. Plots	TRAIT Trunk Diameter (%)	Total Height (%)	Width of Live Crown (%)	Foliage Condition (%)	Branch and Trunk Condition (%)	Maintenance Needs (%)
1	20	10.7	8.1	22.4	28.0	24.5	2.1
2	18	9.7	9.8	22.8	19.9	19.3	3.9
3-4	22	12.1	13.7	22.2	10.6	32.9	4.7
5-7	9	16.5	14.9	27.5	33.7	43.9	21.1
10-12	7	17.4	10.6	14.4	29.2	34.3	0.0
1-4	60	11.0	10.6	24.1	27.1	35.1	3.7
5-12	16	16.9	13.9	21.7	29.0	35.0	7.2
1-12	76	14.9	11.4	23.3	27.6	33.0	3.9

[a]% CV = standard deviation of the plot mean divided by the plot mean, all multiplied by 100.

where X = plot means, 7 = mean of plot means, and r = number of plots. V(7) was calculated for pairs of plots (r = 2) of a cultivar having the same age and in the same city and site category. Plot errors were assumed to be independent. Representative variances for age class ranges (indicated by the CV analysis) were computed as averages over the constituent age classes (Table 5).

Crown category segregation was not possible for the representative variance procedure because of limited data. Average CV changes with age were similar for columnar and spreading crown cultivars. It was assumed that variances between plot means are similar for crown categories.

Statistical Precision

The precision of the experiment depends on the variance of the cultivar mean, and may vary with the cultivar tested and the trait measured (Federer, 1955). The least significant difference (LSD) and the fraction (W) of the representative age class mean (X) detectable between any two cultivars are directly proportional to the representative variance (V). Accordingly, a test of a cultivar whose between-plot variance is greater than V has lower precision (greater LSD and W) than a test of the representative cultivar. W is inversely proportional to the representative means so that a test of a cultivar with an age class mean (X) less than X has lower precision (greater W) than the representative cultivar.

In general, using the representative means and variances to estimate precision is a conservative procedure. Estimates of precision, whether expressed as LSD or W, are "worst case" estimates and useful as guidelines for testing any of the cultivars.

Detectable least significant differences (LSD, WX) were computed for various sizes and arrangements of city performance tests (Table 6). Relationships between the LSD patterns and certain design parameters are evident (Figure 1). LSD decreases (experimental precision increases) as total trees per cultivar (nr) increases. When nr is held constant LSD decreases as trees per plot (n) increases from two to five, but increases as n increases from five to eight (with the consequent decrease in plots or replicates per cultivar (r)). It is useful at age four to compare the LSD's to representative means, which are at minimum values for most traits.

Table 5
Representative variances between plot means, all cultivars

Age Class[a]	Representative Variance Level[b]	TRAIT: Trunk Diameter (inches)	Total Height (feet)	Width of Live Crown (feet)	Width of Crown-to-Height Ratio	Foliage Condition (1 to 5)[c]	Branch and Trunk Condition (1 to 5)[c]	Maintenance Needs (1 to 5)[d]	Injury Incidence(#)[e]
1-2	Low	0.002	0.720						
	Avg.	0.004	0.821						
	High	0.006	0.921						
3-22	Low	0.050	0.020						
	Avg.	0.288	1.327						
	High	0.500	5.492						
1-22	Low			0.020	0.0002	0.001	0.001	0.001	0.001
	Avg.			0.996	0.0029	0.039	0.113	0.101	0.056
	High			4.054	0.0157	0.092	0.369	0.524	0.237

[a]Age class extended to age 22 to accomodate traits with representative means available to age 22 but with no variance estimates beyond age 14.

[b]Representative variance levels reflect estimated minimum (low), mean (avg.), and maximum (high) values expected for all cultivars and cities sampled.

[c]Foliage condition and branch and trunk condition were measured using the continuous scale 1 = excellent health to 5 = tree dying.

[d]Maintenance needs were measured using the continuous scale 1 = very low cost to 5 = very high cost.

[e]Injury incidence equals number of injury types per tree.

Table 6
Trait means at age four and detectable differences for various designs of a performance test

Age Class Four Representative Means[a]			Trunk Diam. (in.)	Total Ht. (ft.)	Width Live Crown (ft.)	Crown Width: Height Ratio	Foliage Condition[b]	Branch & Trunk Condition[b]	Maintenance Needs[c]	Injury incidence[d]
Spreading Cultivars			2.15	13.90	6.70	0.47	1.35	1.85	0.80	0.40
Columnar Cultivars			1.75	12.45	3.30	0.27	1.84	2.25	1.39	0.65
Total Trees Per Cv. (nr)	Plots Per Cv. (r)	Trees Per Plot (n)								
24	12	2	2.78[e]	5.97	5.12	0.279	0.91	1.74	1.65	1.23
24	8	3	1.15	2.48	2.12	0.116	0.38	0.72	0.68	0.51
24	6	4	0.99	2.11	1.81	0.099	0.32	0.62	0.58	0.43
25	5	5	0.94	2.03	1.74	0.095	0.31	0.59	0.56	0.42
24	4	6	0.98	2.09	1.80	0.098	0.32	0.61	0.58	0.43
24	3	8	1.03	2.22	1.90	0.104	4.34	0.65	0.61	0.46
48	24	2	1.97	4.22	3.62	0.197	0.65	1.23	1.17	0.87
48	16	3	0.82	1.75	1.50	0.082	0.27	0.51	0.48	0.36
48	12	4	0.70	1.50	1.28	0.070	0.23	0.44	0.41	0.31
50	10	5	0.67	1.43	1.23	0.067	0.22	0.42	0.40	0.29
48	8	6	0.69	1.48	1.27	0.069	0.23	0.43	0.41	0.30
48	6	8	0.73	1.57	1.35	0.073	0.24	0.46	0.43	0.32

[a]Representative age class means are from Tables 2 and 3.

[b]Condition of foliage, branches, and trunk were measured using the continuous scale 1=excellent health to 5=tree dying.

[c]Maintenance needs were measured using the continuous scale 1=very low cost to 5=very high cost.

[d]Injury incidence equals number of injury types per tree.

[e]LSD ($W\bar{\bar{X}}$) detectable between two cultivars at a five percent level of significance.

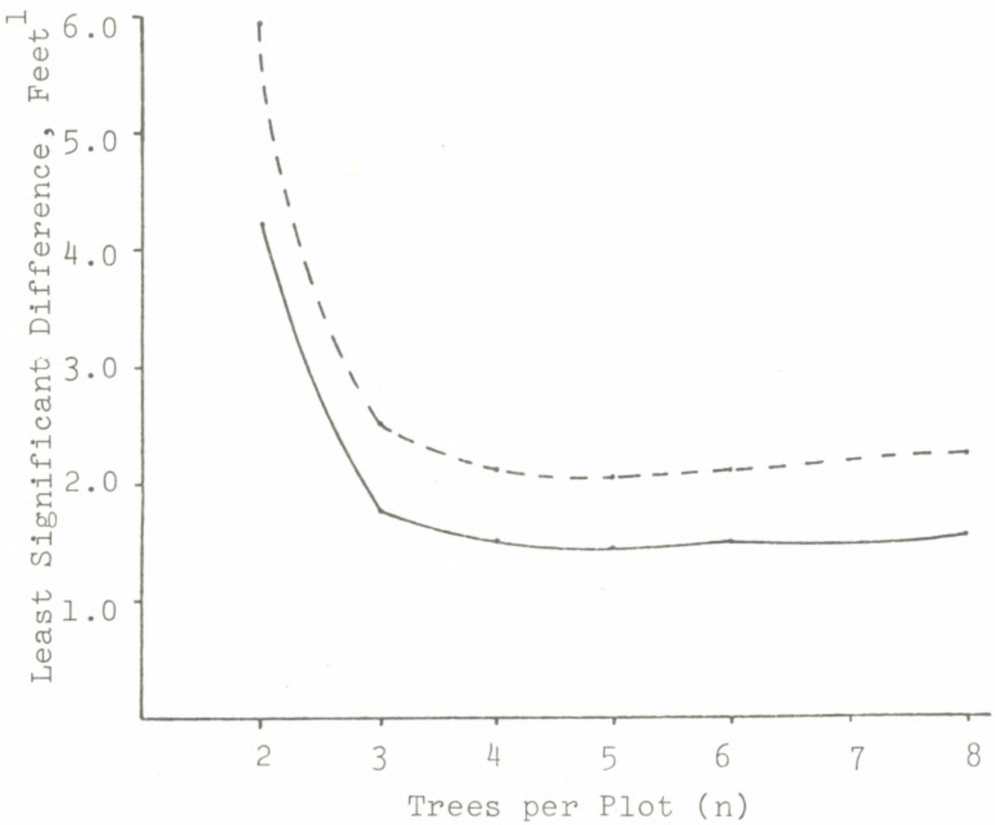

Figure 1. Detectable total height differences for various designs of a city performance test. (Dashed line: total size of experiment (nr is 24 or 25 where r = number of plots per cultivar. Solid line: nr is 48 or 50).

[1] Least significant difference (LSD, WX) detectable between two cultivars at a five percent level of significance. (Total height representative mean at age class four is 13.90 feet for spreading crown cultivars and 12.45 feet for columnar cultivars).

To show how the fraction (W) of the mean that is detectable changes with age, W was computed in each age class for height and foliage condition of spreading crown cultivars (Table 7). LSD's for two city test sizes were used and W was computed by dividing LSD by the age class mean. These W values were compared to similar W values computed for columnar cultivars. W decreases, as does LSD, as nr and r increases (Figure 2). Unlike LSD, W is dependent on the mean and thus age class and crown type. For height, which increases with age, W decreases with age; age class four is then the critical age where W is greatest and experimental precision is poorest. Small differences among cultivars for age classes one to three are presumably not useful for selection, as these may reflect lingering nursery effects. For foliage condition, which does not change systematically with age, age class six is critical because the mean was smallest and W was greatest. Columnar crown cultivar tests were about one-eighth *less* precise (greater W) in detecting total height differences than spreading crown cultivar tests, but were about one-third *more* precise in detecting foliage condition differences.

Practical Precision

The precision of the experiment also varies with the traits measured. The *useful* precision for a trait is the minimum difference that an arborist would consider significant enough for a choice between the cultivars tested. For instance, a height difference of one-half foot at age four (7 = 13.9 feet, W = 0.04 for spreading crown cultivars) is probably insignificant for practical purposes, though it may be indicative of lower health or vigor when selecting planting stock; but a difference of two feet (W = 0.14) would probably warrant selection of one cultivar over another. A useful level of precision for total height might be W = 0.15 which four years after planting denotes a minimum difference, LSD, of 2.1 feet (0.15 x 13.9 feet (age four X)). Determining a useful level of precision for subjectively measured traits is not as straightforward. For instance, foliage condition is measured on the continuous scale of one (excellent health) to five (tree dying). If the level of precision is W = 0.20, then the minimum LSD = 0.22 (0.20 x 1.10 (age five X). It is uncertain whether such a difference is useful to any arborist in making selections. A useful level of precision for foliage condition might be W = 0.40 which yields a minimum difference, LSD = 0.44 (0.40 x 1.10 (age five X)).

Table 7
Detectable differences of spreading crown cultivars related to age and design of city tree performance tests

	TRAIT			
	Total Height		Foliage Condition	
Total Trees per Cultivar (nr)	25	50	25	50
Plots (Replicates) per Cultivar (r)	5	10	5	10
Age Class				
	100 W (%)[a]		100 W (%)	
4	14.6	10.3	23.0	16.3
6	12.6	8.9	25.8	18.3
8	11.0	7.8	22.1	15.7
10	10.0	7.0	19.4	13.8
12	9.3	6.6	17.2	12.2
14	8.9	6.2		
16	8.5	6.0		
18	8.2	5.8		
20	7.9	5.6		
22	7.7	5.4		

[a] 100W equals percent of age class mean detectable as a difference between two cultivars at a five percent level of significance for five-tree plot tests.

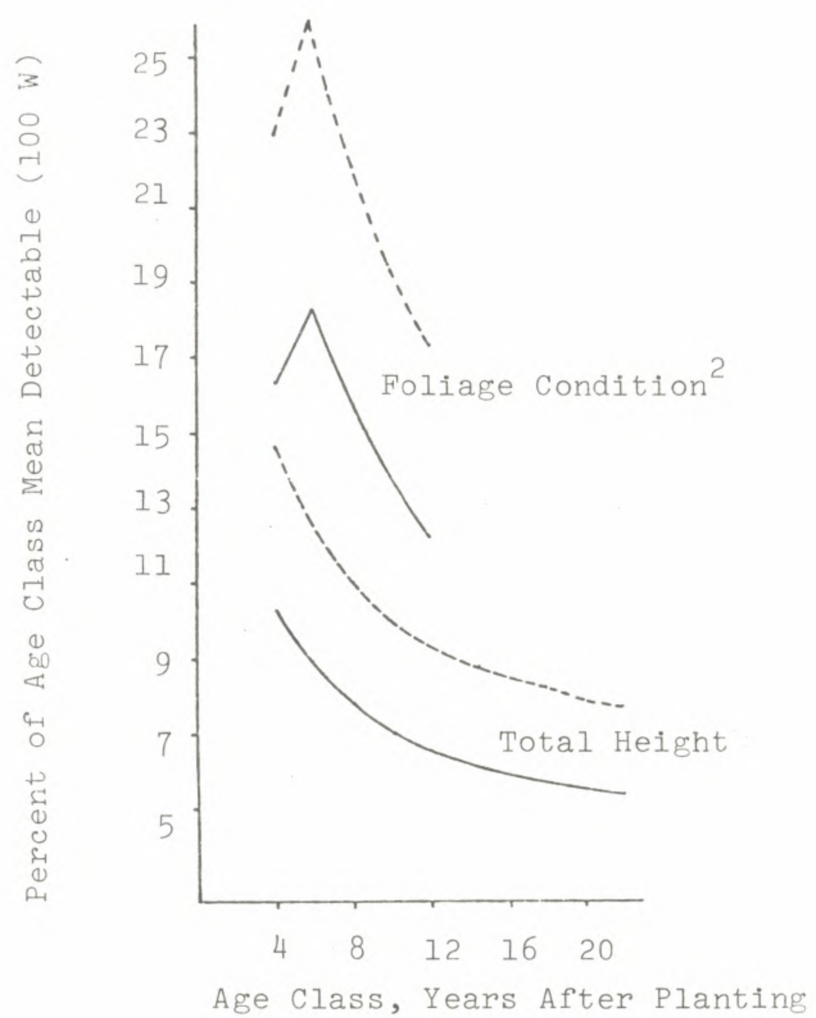

Figure 2. Detectable differences of spreading crown cultivars related to age and design of city tree performance tests. (Dashed lines: total size of experiment (nr) is 25, n = 5 trees per one cultivar plot, r = 5 = plots per cultivar. Solid lines: nr = 50, n = 5, r = 10.)

[1] 100W = percent of age class mean detectable as a difference between two cultivars at a five percent level of significance for five-tree plot tests.

[2] Foliage condition was measured on the continuous scale 1 = excellent health to 5 = tree dying.

By comparing useful levels of precision for total height (e.g. the suggested minimum LSD = 2.1 feet) and foliage condition (0.44) with the LSD's computed from representative means and variances (Table 6), useful experimental design alternatives can be determined. A design is useful if it yields LSD's equal to or less than the LSD's prescribed above. Therefore, designs with 24 or 25 trees per cultivar in a city and four to eight trees per one-cultivar plot may be considered useful. These design alternatives are sufficiently sensitive with respect to all other traits except trunk diameter (W = 0.46 at age four) and injury incidence (W = 1.10). (Analysis of injury incidence is dubious because of the scale of measurement, a count of injury types per tree). The optimal plot size is five, which yields the greatest precision for all traits. Forest geneticists usually opt for plot sizes of four to ten trees (Steiner and Gerhold, 1976). Wright and Freeland (1960) found four trees per plot was optimal for six forest plantation experiments. In general, there is greater precision if a cultivar is tested in a series of scattered small plots than in a few large ones (Wright and Freeland, 1960).

Planting Arrangements

The sizes and numbers of plots determined above apply to a completely randomized design in which equal numbers of plots of each of two cultivars are randomly located in a city's land area corresponding to one site category. Thus, if 50 trees are planted, 25 of each cultivar, a completely randomized design could have ten plots at ten locations, i.e., five plots of five trees per cultivar. The design is especially useful if the area sampled is essentially homogeneous, e.g. small variation among plots in a site category (small contribution to error variance).

The randomized complete block design probably would be more precise than the completely randomized design. City site categories are probably not homogeneous; differences in soil fertility, root space, sunlight, etc. occur among urban sites. By grouping plots into blocks containing both cultivars, variation between plots (cultivars) within blocks presumably would be less than among plots in different blocks (Steel and Torrie 1960). Thus, for 50 trees, 25 per cultivar, a randomized complete block design could have five ten tree blocks and five trees per cultivar in each block.

A useful alternative is to plant the cultivars in pairs within blocks producing a complete block design with nonrandom pairs (split plots). Like the randomized complete block design, there would be five trees per cultivar per block. This type of design is most useful when there is a significant environmental effect on cultivar performance, e.g. if both cultivars grow rapidly in some locations and slowly in others (variation between trees of a pair is less than that between trees of different pairs (Steel and Torrie 1960)).

Constraints on Testing

Constraints on the size, arrangement, and evaluation of city cultivar tests affect the experimental design by diminishing error control, increasing the error variance, and thus decreasing its reliability and the precision of significance tests. Mortality must be considered. Numbers of replications and trees per plot should be greater than required for statistical efficiency.

Costs of the testing program constrain the size of the experiment and determine the program's feasibility. The size and arrangement of the experiment should provide useful but not excessive precision and minimize cost per unit of information. Gerhold and Bartoe (1976) described a labor-sharing system that might provide adequate financing for most costs of the testing program.

Limited spaces available for city test plots also constrain the size of the experiment. Additional information could be gathered from trees planted as replacements among older trees and from larger experiments in fields or parks in or near a city, or in highway sites. Such information would have to be used discriminantly when making inferences about landscape tree performance in cities.

The managerial practices of a city also constrain the size and arrangement of experiments. City planning strategies usually include fixed annual budgets and designated planting sites as well as source and identity of cultivars to be planted. Such factors do not necessarily preclude testing in cities. For instance, if an arborist cannot plant enough trees in one year to achieve the desired statistical precision, he could plant test cultivars in successive years. The total number of trees and replicates would have to be increased to offset the year effect. The planting of some cultivars, e.g. columnar cultivars, may be so limited as to make city testing entirely infeasible. City planting strategies may also provide flexibility in experimental design. For instance, if split plots are used and many sites are available, sites that allow planting whole plots parallel to the site gradient are best (Wright and Freeland 1960).

Many other aspects of performance tests can be controlled to improve experimental precision. Obtaining different cultivars from one nursery avoids a nursery effect. Epstein (1972) and Wareing (1970) found some variability within a cultivar may occur if different rootstocks are used; this effect should be avoided, if possible. Variations in spacing between trees, planting methods, etc., may also impose extraneous effects on cultivar performance. Finally, standardized measurement techniques are needed, especially for subjective traits.

Extending the Test Universe

The experimental design has been viewed primarily with a single city, single site category (residential, business, etc.), two-cultivar test in mind. The design concept can be extended to test many cultivars in "a large number of test sites that sample a geographic region extensively, and represent the range of conditions in which each species is planted, particularly along streets and highways" (Gerhold and Bartoe 1976). Various genotype-environment interactions such as cultivar-site category and cultivar-city (or region) interactions could be determined when many site categories and cities are included in the testing program. Refined analytical methods will be required as well as a centralized coordinating facility. Pooling of information from many cities over many years should provide arborists, nurserymen, and breeders with a comprehensive understanding of landscape tree performance in cities.

LITERATURE CITED

Epstein, E.
1972. MINERAL NUTRITION OF PLANTS: PRINCIPLES AND PERSPECTIVES. John Wiley and Sons, Inc., New York. 412 pp.

Federer, W. T.
1955. EXPERIMENTAL DESIGN. Macmillan Co., New York. 544 pp.

Gerhold, H. D., and W. D. Bartoe, II.
1976. PERFORMANCE TESTING TREE CULTIVARS IN METROPOLITAN ENVIRONMENTS. J. Arbor. 2(12):221-227.

Gerhold, H. D., A. J. Long, and M. E. Demeritt, Jr.
1975. GENETIC INFORMATION NEEDED FOR METROPOLITAN TREES. J. For. 73(3):150-153.

Gerhold, H. D., and K. C. Steiner.
1976. SELECTION PRACTICES OF MUNICIPAL ARBORISTS. U.S.D.A. For. Serv. Gen. Tech. Rep. NE-22:159-166.

Moeller, G. H.
1976. FOREWARD. U.S.D.A. For. Serv. Gen. Tech. Rep. NE-22.

Mower, R. G.
1973. SOME OBSERVATIONS ON STREET TREE PLANTINGS. Proc. International Shade Tree Conf. 49:49-55.

Reisch, K. W., G. Hull, and H. M. Hill.
1971. CASE HISTORIES OF SEVERAL STREET TREE SPECIES AND CULTIVARS AT SELECTED SITES IN FIVE OHIO CITIES. Ohio Agr. Res. and Dev. Center. Hort. Dept. Series 376 (July). 65 pp.

Santamour, F. S., Jr.
1968. THE FOREST AND THE STREET: A COMPARISON OF TREE IMPROVEMENT IN FORESTRY AND HORTICULTURE. Proc. Central States Tree Improv. Conf. 6:1-5.

Santamour, F. S., Jr.
1971. SHADE TREE IMPROVEMENT RESEARCH: GOALS AND LIMITATIONS. Proc. Northeast. For. Tree Improv. Conf. 17:60-64.

Steel, R. G. D. and J. H. Torrie.
1960. PRINCIPLES AND PROCEDURES OF STATISTICS. McGraw-Hill Book Co., Inc., New York. 481 pp.

Steiner, K. C., and H. D. Gerhold.
1976. TESTING AND EVALUATING NEW TREES FOR METROPOLITAN LANDSCAPES. U.S.D.A. For. Serv. Gen. Tech. Rep. NE-22: 227-234.

Wareing, P. F.
1970. GROWTH AND ITS COORDINATION IN TREES. Proc. Second Long Ashton Symposium (University of Bristol):1-21.

Wright, J. W., and F. D. Freeland.
1960. PLOT SIZE AND EXPERIMENTAL EFFICIENCY IN FOREST GENETIC RESEARCH. Mich. State Univ., Agr. Exp. Sta. Tech. Bul. 280. 28 pp.