Appendix A.
Statistical Methodology

THE SURVEY POPULATION

The 2009 Census of Horticultural Specialties (CHS) was designed as a follow-on to the 2007 Census of Agriculture, and includes all operations that reported horticultural crop sales of $10,000 or more, or the presence of sod, nursery products, short rotation woody crops or Christmas trees on the 2007 Census of Agriculture. Horticultural crops include bedding plants, potted flowering plants, cut flowers, cut cultivated florist greens, trees, shrubs, ground covers, vines, fruit and nut trees, sod, dry bulbs, greenhouse produced vegetables, commercial vegetable transplants, vegetable and flower seeds, Christmas trees, short term woody crops, aquatic plants, unfinished or prefinished plants, propagation materials, and other nursery or greenhouse plants.

Data collection for the 2009 Annual Floriculture Survey was conducted in conjunction with the 2009 Census of Horticultural Specialties. Supplemental questions, not summarized in the 2009 Census of Horticultural Specialties, were included in the data collection to meet the requirements needed for the 2009 Annual Floriculture Survey.

DATA COLLECTION

Method of Enumeration

The 2009 CHS was conducted primarily by mail. It was supplemented with Electronic Data Reporting (EDR) via the Internet, telephone enumeration, and personal enumeration.

Report Form

A 24-page report form was designed to capture the number of horticultural products sold and the value of sales for both retail and wholesale sales. The various types of plants sold were grouped by sections in the report form.

Additionally, information was obtained for area in production for several types of crops; marketing channels; estimated value of land, buildings, machinery, and equipment; production expenses; and the number of hired workers employed by the operation in 2009. See Appendix B for facsimiles of the report form and instruction sheet.

Report Form Mailings and Respondent Follow-up

The 2009 CHS report form was mailed from the Census Bureau's National Processing Center (NPC) at Jeffersonville, IN on December 14, 2009. The mail packet included a labeled report form, an instruction sheet, a letter requesting a prompt response that included electronic data reporting instructions, and a return envelope to NPC for data capture. The report form carried a return due date of February 5, 2010. A second mailing to non-respondents took place from NPC on February 22, 2010. Telephone follow-up interviews to non-respondents took place from March 1 to July 10, 2010 from a NASS Data Collection Center.

Data collection for the 2009 CHS was coordinated with other NASS surveys. In some cases, if a horticulture operation was also selected for a survey, NPC mailed the census materials to NASS field offices. Field office personnel were responsible for collecting the census data and completing other survey report forms in the most efficient way to reduce the number of contacts and minimize respondent burden.

REPORT FORM PROCESSING

Data Capture

All report forms returned to NPC were immediately checked in using bar codes printed on the mail label and were removed from follow-up mailings. All forms were reviewed prior to data keying to identify inconsistencies and ensure the data could be keyed. Major inconsistencies, respondent remarks, and blank forms were reviewed by analysts and adjusted prior to keying. In some cases, report forms were mailed to field offices for further editing. All forms with any data were scanned and an image was created for each page of the report form. After images were created, the data were keyed directly from the report form.

Data Editing and Analysis

Data from each report form were processed through a computer edit that flagged inconsistent entries. Each report form with a flagged entry was reviewed by an analyst. Action was required for any record with reported data that were obviously incorrect. In some cases, respondents may have failed to provide all of the information requested, only indicating the presence of an item but not the amount. These items were coded for computer imputation.

After the initial edit, an automated imputation program supplied missing data based on state or national averages. A post-imputation computer edit was performed to ensure imputation actions provided acceptable results. Instances where imputed data failed edit checks were referred to analysts for corrective action.

The computer edit ensured the data on a report form were internally consistent. An analysis tool was provided to examine the data across records to check for distributional irregularities and data outliers. Analysts corrected suspect data when necessary and re-edited the record.

ESTIMATION

Nonresponse Weighting

The 2009 Census of Horticulture is a census in the sense that an attempt was made to obtain data from every known horticulture operation on the NASS horticulture mail list. Although much effort was expended to obtain a response from each operation, it is not possible to obtain a complete set of responses. Nonresponse can lead to biases in published estimates because the information concerning horticulture production on the nonresponding farms could not be factored into the estimates. Such estimates of totals will be biased low. To reduce this bias, NASS made nonresponse weight adjustments. The nonresponse weight adjustment increases the weight of responding operations to account for the data that would have been reported by the nonresponding operations. This increased the estimates of totals obtained by the respondents and reduced this bias.

Conceptually, each operation on the mail list began the census with a weight of one. In other words, if each operation on the list provided the requested data, the data could be simply added up to the total. In the presence of nonresponse, nonresponse adjustments are computed and applied to the initial weights of the responding operations resulting in a nonresponse-adjusted weight greater than one for these operations. The initial weight of each nonresponding operation is then adjusted to zero. The adjustments are computed in a manner that requires the sum of the nonresponse-adjusted weights across the responding operations on the mail list to equal the sum of the initial weights across all operations on the mail list.

If the total number of operations on the mail list is N, the sum of the initial weights across all operations on the mail list equals N, because the initial weight for each operation on the list is one. The sum of the nonresponse-adjusted weights across all responding operations on the mail list is also required to equal N. In fact, the sum of the nonresponse-adjusted weights across all operations on the list would sum to N as well because the nonresponse-adjusted weight of nonresponding operations is set to zero.

Weight-Adjustment Groups

To compute nonresponse adjustments, each record on the mail list was placed in a weight-adjustment

group. Each operation was assigned to a group based on the characteristics used to define the group. It was necessary that the characteristics that defined the weight-adjustment groups were available for responding and nonresponding operations alike. Therefore, it was not possible to define weight-adjustment groups using data collected via the CHS. Information used to define the groups was obtained from historical information maintained on the mail list and was, therefore, available for each operation.

The information used to create the weight-adjustment groups was a measure of the horticultural economic size (HES) that was available for all operations on the mail list. The basic definition of the weight-adjustment groups is given below:

Definition	Weight Adjustment Group ID
HES<=$50,000	101
$50,000<HES<=$250,000	102
$250,000<HES<=$500,000	103
$500,000<HES<=$1,000,000	104
$1,000,000<HES<=$5,000,000	105
$5,000,000<HES	106
Extreme Operator	900

Each operation on the mail list was placed in a group based on the operation's HES. One additional weight-adjustment group was composed of influential horticulture operations (extreme operators) whose data were obtained from the operation itself, or the data were imputed manually. These operations were economically large and/or possibly unusual in some characteristic, making them difficult to adjust for through nonresponse weighting adjustments. The operations that were placed in the extreme operator group were placed there without regard to the HES. Weight-adjustment groups and weight adjustments were created and carried out separately for each State.

To ensure sufficient numbers of responding operations in each group, some collapsing of weighting groups occurred, resulting in some states having fewer weight adjustment groups than others.

Nonresponse-Adjustment Computation

A separate nonresponse adjustment was calculated within each weight-adjustment group. All responding records within each group received the same nonresponse adjusted weight. The nonresponse-adjustment was obtained by dividing the total number of operations contained in a group by the number of responding operations in the group. If the total number of operations in the group was 50 and the number of responding operations in the group was 40, the nonresponse-adjustment for the responding operations was 50/40 or 1.25. The nonresponse-adjusted weight for all responding operations in the group was the product of the census weight and the nonresponse adjustment of 1.25. This was simply (1 x 1.25) or 1.25. Note that 1.25*40=50, the total number of operations in the group.

The assumption was that, within each weight-adjustment group, the data that the nonrespondents would have provided had they responded was collectively similar to the data provided by the respondents. This assumption was made somewhat more plausible because operations in the same group shared similar characteristics with respect to the information used to define the group - the HES.

Coverage Weighting Adjustments

The target population for the 2009 CHS was all operations that had at least $10,000 of commercial horticulture production in 2009. Unfortunately, it is impossible to compose a list of operations that is complete. Due to this incompleteness of the mail list, data produced from it, even if perfectly corrected to account for nonresponse, will still have a tendency to be biased downwards because operations not on the list would not have any representation. This bias due to list incompleteness is called coverage bias, or more specifically, bias due to under coverage of the list.

To reduce the amount of this bias, an additional adjustment was calculated and applied to the nonresponse-adjusted weight for each responding operation. This was called the coverage adjustment. The coverage adjustment was calculated using the same weight-adjustment groups defined above.

Coverage Adjustment Computation

Each operation on the horticulture mail list was a respondent to the 2007 Census of Agriculture. All data produced from the agricultural census were adjusted for coverage error. Coverage adjustment was made possible using data that were based on a list (or area frame) of geographic land segments. In theory, every segment of land in the U.S. is contained in one of these segments. This implies that a census or survey based on such a frame will have complete coverage and represent all farms in the U.S. The CHS was used to derive estimates of the magnitude of the coverage error associated with the census mail list. Each 2007 census respondent's nonresponse-adjusted weight was also adjusted for coverage based on the estimated coverage error obtained from the area frame survey. This resulted in a fully coverage-adjusted weight for each 2007 census respondent.

These weights were pulled forward for every operation on the horticulture mail list. The information contained in the 2007 census weights was used to create a coverage adjustment for the CHS. This information was somewhat dated and imperfect, but it helped account for horticulture operations that either did not respond to the 2007 census or were not on the census mail list.

The coverage adjustment for responding operations to the CHS was calculated by summing the 2007census fully adjusted weight across all operations that resided in the weight-adjustment group. This provided an estimate of the total number of horticulture operations in a State that fell into that group, whether they were contained on the horticulture mail list or not. This number was then divided by the sum of the nonresponse-adjusted weights for all responding operations in the group. This resulted in the CHS coverage adjustment for that group. If the sum of the fully adjusted census weights in a group for all operations in the group was 60 and the sum of the nonresponse-adjusted weight across all responding operations in the group was 50, the CHS coverage-adjustment was 60/50 or 1.2. Multiplying the coverage adjustment by the nonresponse adjusted weight resulted in the fully-adjusted CHS weight. In the given example with 40 responding operations, this would be 1.25 x 1.2=1.5. Note that 1.5 x 40=60, the estimated total number of horticulture operations that would fall into that group, whether on the list or not. All responding records in a group would have the same fully-adjusted CHS weight.

Summary Weights

Many of the fully adjusted weights for the 2009 Census of Horticultural Specialties were not whole numbers (integers). Using these weights to create the estimates published in the tables would result in fractional values. These would be difficult to read and cause consistency problems between related tables. To avoid some of these problems, summary weights were created by randomly moving the fully adjusted weights up or down to an integer, which preserved the total census weight. This process is called weight integerization. The resulting summary weights were used to produce the numbers published in the tables.

MEASURES OF PRECISION AND ACCURACY OF THE ESTIMATES

All numbers published in the tables are merely estimates of particular characteristics of the entire population of horticulture operations. The true values of these characteristics are unknown and unknowable. Even though an attempt was made to contact every operation on the mail list, the data produced by the census will not exactly attain the true values. This is due to a number of factors, such as survey nonresponse, mail list incompleteness, and the weight integerization process. Hypothetically, if the entire survey process was repeated over and over again, each replication of the survey would almost certainly produce a different estimate for the same population value every time. This is because each time the survey is carried out, a different set of respondents would be obtained, response rates would fluctuate, and coverage rates of the mail list could change.

It is possible to obtain an idea of how much this variation would be on average by calculating the estimate's variance. The variance of an estimate gives a measure of the average squared random fluctuation that would be seen in an estimate if the survey was carried out multiple times. Because the

variance measures random fluctuation in squared units, the square root of the variance is computed to obtain a random fluctuation measure that is in the same units as the original estimate. This is called the standard error (se) of the estimate. The standard error can then be divided by the estimate itself to show the relative size of the standard error to the estimate. If this ratio is small, the estimate is quite precise. If this ratio is large, the estimate is imprecise. An estimate of 100 with a standard error of 2 would result in a relative standard error of .02 or 2 percent. This would be a very precise estimate. An estimate of 100 with a standard error of 20 would result in a relative standard error of 20 percent. This might be considered to be an imprecise estimate. The idea of precision can be made a little more clear by stating that if the estimate is 100 with a standard error of 2, you could be quite confident that the true population value would be in the interval 96 to 104 (within two standard errors of the estimate).

Unbiased estimates are generally accurate. This is to say that if the survey is hypothetically repeated over and over, the average of the estimates obtained would be very close to the true value being estimated. This does not mean that any particular realization of the estimate will be "close" to the true value. An accurate estimate that is not precise has a good chance of missing the true value of the characteristic being estimated by a significant amount.

If the estimate contains some bias, both precision and accuracy are measured by computing the mean square error (MSE) of the estimate. Bias is systematic error that would be about the same for every hypothetical replication of the survey. Bias is not random fluctuation. The weight adjustments described earlier are used to decrease biases in the estimates. However, the weight integerization process introduces some bias. Ideally, the amount of bias contained in an estimate should be small or non-existent but, in conducting actual surveys, some biases may be hard to avoid. Biased estimates can be precise, but in hypothetical replications of the survey, will tend to be systematically lower or higher than the true population value being estimated.

The mean square error is computed by adding a term to the variance called the estimated squared bias. The mean square error can be used to measure the effects of both random variation and bias contained in an estimate. The mean square error measures the estimate's precision and accuracy.

Like the variance, the mean square error is measured in squared units, so the square root of the mean square error is often taken. This results in what is called the root mean square error (RMSE). Like the standard error, the ratio of the root mean square error to the estimated value can be created. This ratio gives a measure of the relative root mean square error (relative RMSE) of the estimate. When this ratio is small, the estimate is both precise and accurate. If this ratio is large, the estimate may be precise but not accurate, it might be accurate but not precise, or it might be neither accurate nor precise.

Table A. **Reliability Estimates of Operations and Value of Sales of All Horticultural Specialty Crops – United States and State: 2009**

[For meaning of abbreviations and symbols, see introductory text]

Geographic area	Operations		Total sales		Wholesale sales		Retail sales	
	RMSE	Relative RMSE	RMSE	Relative RMSE	RMSE	Relative RMSE	RMSE	Relative RMSE
United States	113	0.52	84,840	0.73	55,975	0.57	35,312	1.97
Alabama	10	2.95	3,459	1.57	3,106	1.52	466	3.07
Alaska	2	5.12	302	2.96	146	5.33	247	3.30
Arizona	2	1.93	3,849	1.39	4,588	1.77	2,405	13.36
Arkansas	5	3.83	2,075	6.03	1,241	5.68	1,080	8.61
California	35	2.16	66,904	2.93	48,537	2.37	23,324	9.94
Colorado	10	3.84	4,326	2.11	2,845	1.67	2,711	7.86
Connecticut	12	3.62	2,249	1.24	1,954	1.33	1,413	4.09
Delaware	4	6.02	917	5.18	893	11.71	330	3.28
Florida	38	2.00	11,411	0.89	9,222	0.75	5,727	9.72
Georgia	11	2.45	5,601	2.45	5,645	2.85	2,617	8.71
Hawaii	16	3.36	1,431	1.73	1,515	2.05	628	6.94
Idaho	9	3.42	3,359	4.42	3,390	5.17	424	4.08
Illinois	17	3.09	4,055	1.53	2,999	1.45	1,724	2.95
Indiana	14	3.78	3,740	3.38	4,631	5.55	1,776	6.50
Iowa	8	3.84	1,623	2.07	1,504	2.62	1,403	6.74
Kansas	9	4.68	1,509	2.79	1,060	3.22	1,460	6.86
Kentucky	16	3.86	1,304	1.73	1,006	1.95	637	2.68
Louisiana	8	3.37	3,281	4.32	3,232	5.10	684	5.43
Maine	10	3.93	817	1.66	562	1.67	509	3.28
Maryland	11	2.94	21,766	9.67	22,333	11.71	1,406	4.11
Massachusetts	20	5.14	6,606	5.81	4,462	6.41	2,803	6.34
Michigan	29	2.70	6,731	1.19	4,346	0.95	3,944	3.66
Minnesota	14	3.11	10,091	3.75	4,150	2.42	6,613	6.80
Mississippi	7	4.58	2,731	9.38	1,556	8.75	1,194	10.53
Missouri	13	3.97	1,561	1.49	1,333	1.59	792	3.75
Montana	8	5.88	1,085	5.89	783	7.77	455	5.45
Nebraska	6	3.87	663	1.93	463	3.89	721	3.22
Nevada	1	7.92	546	13.27	119	5.88	435	20.80
New Hampshire	8	4.77	3,700	9.06	3,413	12.49	943	6.97
New Jersey	15	2.46	3,773	1.09	3,474	1.14	1,221	2.87
New Mexico	3	4.79	2,324	6.49	2,248	7.14	429	9.91
New York	34	3.25	6,854	2.26	5,948	2.71	2,246	2.69
North Carolina	23	2.24	6,280	1.25	5,525	1.21	1,355	3.10
North Dakota	2	10.10	1,101	23.76	704	48.57	447	14.03
Ohio	24	3.03	13,908	3.96	13,306	4.86	2,876	3.72
Oklahoma	6	3.32	2,277	1.39	1,854	1.34	1,220	4.80
Oregon	24	2.05	10,802	1.29	9,877	1.25	2,045	4.59
Pennsylvania	40	3.13	4,929	1.67	3,098	1.49	3,480	3.94
Rhode Island	6	4.74	1,323	4.92	1,475	7.50	1,101	15.24
South Carolina	6	2.96	5,024	3.70	4,437	3.66	1,457	10.04
South Dakota	3	6.16	372	3.48	278	4.80	194	3.96
Tennessee	15	2.75	2,527	0.97	2,628	1.10	1,463	7.48
Texas	24	3.24	9,738	1.57	8,677	1.63	4,115	4.74
Utah	3	2.29	1,707	2.10	2,126	3.79	1,269	5.04
Vermont	9	4.57	925	4.86	622	8.72	738	6.19
Virginia	13	3.09	9,603	5.16	3,369	2.32	6,910	16.76
Washington	17	2.73	3,113	1.11	2,742	1.26	1,372	2.19
West Virginia	5	4.61	328	1.47	293	1.81	185	3.04
Wisconsin	30	3.48	2,763	1.51	1,685	1.68	2,420	2.92
Wyoming	1	4.30	239	5.85	244	14.63	363	14.99

Table B. **Relative Standard Errors of Operations and Value of Sales for Selected Horticultural Specialty Items – United States: 2009**

[For meaning of abbreviations and symbols, see introductory text]

Item		Total	RMSE	Relative RMSE
TOTAL VALUE OF SALES BY SIZE				
$10,000 - $19,999	operations	3,180	30	0.96
	$1,000	45,521	446	0.98
$20,000 - $24,999	operations	1,249	18	1.43
	$1,000	27,619	412	1.49
$25,000 - $39,999	operations	2,641	27	1.02
	$1,000	83,615	857	1.03
$40,000 - $49,999	operations	1,261	16	1.30
	$1,000	56,083	732	1.30
$50,000 - $99,999	operations	3,538	39	1.10
	$1,000	251,593	2,534	1.01
$100,000 - $249,999	operations	3,991	35	0.87
	$1,000	631,503	5,331	0.84
$250,000 - $499,999	operations	2,258	20	0.89
	$1,000	791,029	6,731	0.85
$500,000 - $999,999	operations	1,530	15	1.01
	$1,000	1,074,815	10,841	1.01
$1,000,000 - $2,499,999	operations	1,133	12	1.08
	$1,000	1,766,102	19,876	1.13
$2,500,000 or more	operations	804	13	1.61
	$1,000	6,959,443	81,266	1.17
OPERATIONS BY TYPE OF CROP				
Annual bedding/garden plants	operations	7,989	67	0.84
	$1,000	2,305,913	17,123	0.74
Potted herbaceous perennial plants:	operations	6,416	48	0.75
	$1,000	843,788	5,278	0.63
Potted flowering plants for indoor or patio use	operations	4,043	33	0.81
	$1,000	871,474	5,418	0.62
Foliage plants for indoor or patio use	operations	2,728	27	0.99
	$1,000	509,873	8,832	1.73
Cut flowers and cut lei flowers	operations	1,703	21	1.25
	$1,000	403,254	4,238	1.05
Cut cultivated greens	operations	634	14	2.22
	$1,000	84,148	1,347	1.60
Nursery stock sold (see text)	operations	8,441	46	0.54
	$1,000	3,850,363	28,919	0.75
Cuttings, plug seedlings, liners, tissue cultured plantlets, and prefinished plants	operations	1,178	16	1.35
	$1,000	601,657	17,376	2.89
Sod, sprigs or plugs sold (see text)	operations	1,403	17	1.21
	$1,000	876,847	13,885	1.58
Dried bulbs, corms, rhizomes and tubers	operations	223	7	3.32
	$1,000	48,512	1,981	4.08
Food crops grown under protection	operations	1,476	22	1.50
	$1,000	553,270	20,457	3.70
Transplants for commercial vegetable and strawberry production	operations	502	9	1.71
	$1,000	330,647	26,442	8.00
Vegetable seeds	operations	340	10	3.08
	$1,000	89,031	7,856	8.82
Flower seeds	operations	141	5	3.85
	$1,000	30,825	7,070	22.94
Aquatic plants	operations	375	12	3.13
	$1,000	26,000	1,770	6.81
Cut Christmas trees sold (see text)	operations	2,699	27	1.02
	$1,000	249,821	7,088	2.84
Short rotation woody crops sold (see text)	operations	23	2	8.27
	$1,000	2,173	1,096	50.42
Tobacco transplants	operations	189	7	3.82
	$1,000	9,728	462	4.75