APPENDIX A
SURVEY DESIGN AND CALCULATION OF NATIONAL ESTIMATES

Appendix A
Survey Design and Calculation of National Estimates

EPA has collected information from aquatic animal production by using a two-phase sample design with a questionnaire in each phase. A two-phase[1] sample design is a standard survey statistic technique (see, for example, Cochran (1977) or Kish (1965)). In the first phase of this design, information is collected from every unit (e.g., facility) in the sample. In the second phase, detailed information is collected from each unit in a second, smaller, sample. Typically, the first phase sample is used to classify the population for the second phase sample and this second sample is selected from the units in the first sample. Statistical inference can be made using the information from the second phase alone or in some combination of the first and second phases.

In the first phase conducted in August 2001, EPA sent a short screener questionnaire, entitled "Screener Questionnaire for the Aquatic Animal Production Industry" ("screener questionnaire," USEPA, 2001) to a list of 5939 possible aquatic animal production (AAP) facilities. This sample frame (list) is discussed in Section A.1 below. The screener questionnaire consisted of eleven questions to solicit general facility information, including confirmation that the facility was engaged in aquatic animal production, species and size category produced, type of production system, wastewater disposal method, and the total production at the facility in the year 2000. Section A.2 describes the census conducted in this first phase and the data analysis of the responses.

In the second phase conducted in June 2002, EPA selected 263 screener questionnaire respondents to receive the detailed questionnaire, "Detailed Questionnaire for the Aquatic Animal Production Industry," ("detailed questionnaire," USEPA, 2002). EPA designed this second questionnaire to collect detailed site-specific technical and financial information. The detailed questionnaire is divided into three parts. The first two parts collect general facility, technical, and cost data. The third part of the detailed questionnaire elicits site-specific financial and economic data. EPA sent each facility only the portions of each part that were relevant to the operations reported in the screener questionnaire. Section A.3 describes the sample selection criteria and estimation procedures from the responses from this second phase.

A.1 SAMPLE FRAME

In 1998, the US Department of Agriculture (USDA) identified 4,028 aquaculture facilities in its Census of Aquaculture ("USDA Census"). Because their database was confidential and thus not available, EPA constructed a sampling frame from alternative sources consisting of data received from Dun & Bradstreet, augmented with supplemental sources of facilities. Attachment A-1 to this appendix summarizes the differences between the sample frames and other aspects of the two questionnaires.

EPA developed its initial list of facilities from the February 2001 version of the Dun & Bradstreet (D&B) database. D&B provided a list of 2,025 facilities whose primary,

[1] Some textbooks and journal articles refer to two-phase sampling as >double sampling.

secondary, or tertiary SIC codes related to AAP. The SIC codes included 0273 (animal aquaculture), 0279 (animal specialties), and 0921 (fish hatcheries and preserves). EPA found that the D&B database only contained half as many facilities as the USDA Census, 2,025 compared to 4,028. Although the size of the industry may have changed between 1998 (USDA Census) and 2001 (D&B), it was more likely that D&B did not include some facilities identified by the USDA. EPA then examined the total revenue of facilities in the D&B database, and found that it exceeded that of the Census by about 10%. Because both estimates of total revenue were about $1.0 billion, EPA concluded that the facilities not included in the D&B database probably were quite small.

In order to identify AAP facilities not identified by the D&B database, a number of secondary sources were identified and utilized. About 4,000 facilities were identified from supplemental sources. These included:

- An initial list of 2,241 facilities supplied by 24 state agencies such as Departments of Agriculture or Environmental Protection. These data varied considerably in quality and utility, including some lists that were incomplete and/or out of date.
- U.S. Fish and Wildlife Service
- The Internet, associations, and trade journals.
- EPA used its own list of 288 farms from which a subset of 121 new listings in 28 states was identified. EPA developed this list of 288 farms from its Permit Compliance System (PCS), Discharge Monitoring Reports (DMR), and other permit information. In addition, some additional facilities were added from a list of 30 facilities on EPA's site visit list.
- The frame was augmented with a list of public aquariums in the United States. These were identified largely through the Internet as well as data supplied by the American Zoo and Aquarium Association.

Identification and deletion of duplicate facilities (i.e., those appearing more than once on the list, perhaps with slightly different addresses or company names) was conducted both prior to and after mailing the questionnaires. In order to ensure that no active AAP facility would be inadvertently removed, only obvious duplicates were deleted prior to the mailing.

A.2 SCREENER QUESTIONNAIRE (PHASE 1)

This section describes the screener questionnaire responses that were collected in Phase 1 of EPA's survey of the AAP industry. Section A.2.1 describes the sample design, which was a census of the industry, and the number of responses. Section A.2.2 describes the data analysis of the responses including the use of conversion factors; development of sample weights that adjust for non-response; and the estimation of national totals, national means, and their standard errors. While EPA has primarily relied on the Phase 2 data in developing the final rule, this Appendix describes Phase 1 because it was used to develop the proposed rule, and to select the Phase 2 sample.

A.2.1 Sample Design: Census

In Phase 1, the screener questionnaires were mailed to all 5934[2] addresses on the frame. Because they were mailed to all facilities on the sample frame, the sample design for this phase is considered to be a 'census.' After the mailing, 53 unsolicited questionnaires were received that were not on the original mailing list. Many of these were from facilities that operated more facilities than the number of questionnaires that they received. In its data analyses and selection of the sample for the second phase, EPA considers these 53 facilities as if they were part of the original sample frame. Thus, the 'final' frame contained 5987 potential AAP facilities.

As of 8/8/02,[3] EPA had received 4199 completed,[4] 58 incomplete, and 75 blank questionnaires. EPA also had identified an additional 161 duplicate questionnaires (i.e., more than one questionnaire was sent to the same facility). For questionnaires returned by the delivery service, EPA attempted various data retrieval and searches to obtain a better mailing address. For 435 addresses, EPA was unsuccessful in finding better addresses, and thus, EPA assumed that these facilities did not exist (e.g., out of business). In addition, although they received a letter reminding them to return the questionnaire, 1064 facilities did not return their questionnaires and are considered to be 'non-respondents' in the statistical analysis presented in this appendix. (Five of the 1064 facilities returned a blank questionnaire and also are considered to be non-respondents.) Response rates can be calculated in various ways. One widely accepted method is to use the ratio of the number of returned questionnaires to the number of valid addresses. EPA was able to determine the number of valid addresses because the delivery service required recipients to sign a manifest. For the screener questionnaire, the number of valid addresses was 5552, that is, the remainder of the 5987 potential AAP facilities after subtracting the 435 addresses without a viable address. The response rate of 75.6 % is the ratio of the 4199 completed questionnaires to the 5552 valid addresses.

From the completed questionnaires, EPA identified 2329 facilities in the AAP industry. These facilities answered 'Yes' to question 1 which asked 'Do you produce (grow) aquatic animals (fish, shellfish, other aquatic animals) at this facility?'

A.2.2 Data Analysis

Weighting the data allows inferences to be made about all eligible facilities, including those that did not respond to the questionnaire. Another advantage is that weighted

---

[2] Elsewhere in this document and other record materials, EPA may have identified the total number of questionnaires as 5939; however, five were replacements of questionnaires with incomplete mailing labels. In some summaries, EPA includes the replacements as five new questionnaires.

[3] After this date, EPA received a small amount of additional screener information, which EPA reviewed and evaluated. However, because EPA primarily relied on data from the detailed questionnaire in developing the final rule, EPA did not incorporate the more recent screener information into the screener estimates presented in this appendix.

[4] The values in this appendix are upon a more recent version (8/8/02) of the screener database than the version used for Chapter 3. Thus, there are slight discrepancies between the values in that chapter and this appendix.

estimates have less biased variances than unweighted estimates (i.e., counts of the responses). Because of time constraints for the proposal, EPA was unable to incorporate these weighted results into its most analyses, such as economic achievability. However, EPA presented weighted results in Appendix A of the proposal TDD. These results also are provided in this appendix. This section presents its methodology for calculating the weighted results presented in Attachment A-3.

This section consists of three subsections. Section A.2.2.1 describes various conversion factors and their application in determining the biomass, predominant species, predominant production method, and total revenue at each facility. Section A.2.2.2 describes the sample weights that adjust for non-response. Section A.2.2.3 describes the application of these sample weights in developing national estimates (e.g., number of facilities with trout as their predominant species) and the standard errors of these estimates.

A.2.2.1 Use of Conversion Factors

To simplify its data analyses, EPA determined the biomass, predominant species, and predominant production method for each facility, using various conversion factors in Attachment A-2. This section describes the use of the conversion factors and these determinations.

Biomass

For each size category, the screener questionnaire collected production in any of six units (pounds (live weight), number or count, live dry bushels, dozens, dollars sold, or other). To estimate the production at a facility, EPA converted all units into pounds using conversion factors from sources such as the USDA Census of Aquaculture, industry experts, internet sites about fish, and calls to aquaculture farms (see DCN 50070 in Section 10.3 of the proposal record). As shown in Tables A2.1 and A2.2 in Attachment A-2, the conversion factors depended on the species, the size category, and the reported units. When specific conversion factors were not available (for a minority of facilities), EPA used approximate conversion factors based on 1) the weight of food-size animals for the species, 2) an approximate weight ratio of food size to other size animals, and 3) approximate conversion factors from the reported unit into pounds. As an example of using the appropriate conversion factor in Table A2.1, if a facility produced 1,000 catfish of foodsize, the biomass of the catfish was calculated as

1,000 catfish x 1.5 pounds/catfish=1,500 pounds.

As another example, if a facility produced 1,000 whitefish of stocker size, the biomass was calculated using the conversion factor for whitefish of foodsize from Table A2.1 and the stocker size conversion factor from Table A2.2, as follows:

1,000 whitefish $_{stocker}$ x 2.5 pounds/whitefish $_{foodsize}$ x 0.1418 whitefish $_{foodsize}$ / whitefish $_{stocker}$ = 354.5 pounds.

The *total biomass*, or total production, for a facility is the total weight in pounds across all size and species categories.

Predominant Species

To determine the predominant species, EPA calculated the biomass for each species reported by a facility. The species biomass was the total weight in pounds across all size categories for that species. EPA then selected the species with the largest biomass as the predominant species.

Predominant Production Method

In response to question 6 on the screener questionnaire, facilities could specify any of six different production methods (ponds, flow through systems, recirculating systems, net pens or cages, floating aquaculture, and other). However, the screener questionnaire requested species and production information separately from the production method. Thus, for facilities with multiple species, it was not possible to determine which production method was used for a particular species. Also, some facilities reported more than one production method. To assign a single production method to a facility's predominant species, EPA ordered the production methods from most common to least common among facilities with the same predominant species. Table A2.3 in Attachment A-2 presents this ordering of production methods. (As noted in the table, EPA used a slightly different ordering sequence for the data analyses presented in Attachment A-3, than it did for the sample selection for the detailed questionnaire.) As an example, *assume* a facility has catfish as the predominant species and uses both recirculating systems and flow through systems. From Table A2.3, the most common production method for facilities with catfish as the predominant species is ponds; however, ponds are not used at this facility. The second most common production method is flow through systems. Because this facility uses flow through systems, EPA would assume that these flow through systems are the predominant production method for catfish production at this facility.

Total Revenue

In response to question 5 of the screener questionnaire, facilities could report production in any of six units: pounds (live weight), number or count; live dry bushels; dozens; dollars sold; and other. Most facilities reported their total production in pounds, counts, or dollars. To convert the production units into dollars, EPA used the conversion factors in Table A2.4, in Attachment A-2, to estimate the number of facilities that would be subject to the proposed rule in three revenue classes: $20,000-$100,000; $100,000-$499,999, and >$500,000.

As explained in the preamble to the proposed rule, in evaluating the screener questionnaire responses to question 5 (production), EPA used six production size categories that correspond with the revenue classifications used in the 1998 Census of Aquaculture (i.e., $1,000-$24,999; $25,000 - $49,999; $50,000 - $99,999; $100,000 - $499,999; $500,000 - $1,000,000; and >$1,000,000). These classifications were used to develop model facilities representing these size ranges for each species evaluated. Because of the small numbers of facilities in some of the species and production method

categories, EPA has not presented these results to protect confidential business information.

A.2.2.2 Sample Weights

This section describes the methodology used to calculate the base weights, non-response adjustments, and the final weights for the screener questionnaire. The sample weights accounted for different response rates and ineligible facilities. In conjunction with the conversion and predominant determinations described in the last section, the sample weights were used to calculate the national estimates presented in Attachment A-3.

The *base weight* is equal to 1.0 for all facilities because the screener questionnaire was sent to the entire sample frame (i.e., a census).

$$base\ weight = 1.0 \tag{A-1}$$

The number of returned questionnaires includes duplicate questionnaires, whether they were completed or not, but does not include questionnaires that were not deliverable. The non-response adjustment in effect spreads the weight associated with the non-responses (questionnaires not returned) across the responses. The non-response adjustment assumes that the fraction of duplicate addresses among those who responded is the same as the fraction among those who did not respond. Because different species tend to be located in different parts of the country, EPA decided to use the facility location as a basis for calculating the non-response rate. For states with 50 or more respondents, EPA defined the location of the facility as its state. For states with less than 50 respondents, EPA grouped the facilities into one stratum. (See Westat, 2002b, for a logistic regression that assessed which factors were significant predictors of non-response.) Within each stratum g, the non-response weight adjustment is the ratio of the number of facilities with valid addresses to the number that responded.

$$w_g = \left(non-response\ adjustment\right)_g = \frac{Number\ of\ valid\ addresses\ in\ stratum\ g}{Number\ of\ returned\ questionnaires\ in\ stratum\ g} \tag{A-2}$$

The final screener weight w_i for facility i in non-response stratum g can be written as:

$$w_i = \left(base\ weight\right) \times \left(Non-response\ adjustment\right) = 1.0 \times w_{g(i)} \tag{A-3}$$

Where $w_{g(i)}$ is the non-response adjustment corresponding to the non-response stratum (g) associated with facility i.

Although the weight is applicable to all responding facilities, EPA is interested in only those facilities in AAP. For each non-response strata g, Table A.1 shows the number of valid addresses (excluding any duplicate addresses), the number of returned

questionnaires, the screener weight, and the number of responding AAP facilities. The weights for the screener respondents ranged from 1.14 to 1.55.

As an example of the application of the screener weights, consider strata 1 which had 124 valid addresses and 93 returned questionnaires. The sample weight for all facilities in stratum 1 is:

$$w_1 = 1.0 \times \left(\frac{124}{93} \right) = 1.33$$

As shown in the last column, 56 of the 93 returned questionnaires are from AAP facilities. Then, using the sample weight, the estimated number of AAP facilities is 1.33 x 56 = 75 (rounded to an integer).

Using a non-response adjustment assumes that the fraction of facilities doing AAP is the same among the respondent and non-respondents. In its data analyses of the screener questionnaire responses, EPA has assumed that non-respondents have the same characteristics, proportionally, as the respondents. This is a common assumption used in survey estimation. However, if the non-respondents within a non-response adjustment stratum are different from the respondents, the survey estimates may be biased. There is considerable research into the area of non-response estimation (see, for example, Groves and Couper (1998)).

Table A.1. Screener Weights and Number of Facilities by Non-Response (Location) Strata

Non-Response (Location) Stratum	*Number of Valid Addresses*	*Number of Returned Questionnaires*	*Screener Weight* w_g	*Number of Responding AAP Facilities in the Stratum*
1 (AK)	124	93	1.333	56
2 (AL)	162	111	1.459	74
3 (AR)	450	323	1.393	164
4 (CA)	316	249	1.269	144
5 (CO)	65	52	1.250	30
6 (FL)	524	410	1.278	125
7 (GA)	155	118	1.314	69
8 (HI)	163	105	1.552	50
9 (IA)	67	57	1.175	31
10 (ID)	109	92	1.185	59
11 (IN)	68	55	1.236	29
12 (LA)	246	182	1.352	119
13 (MA)	323	218	1.482	114
14 (ME)	100	73	1.370	50
15 (MI)	107	85	1.259	51
16 (MO)	74	65	1.138	44
17 (MS)	220	163	1.350	121

Non-Response (Location) Stratum	Number of Valid Addresses	Number of Returned Questionnaires	Screener Weight w_g	Number of Responding AAP Facilities in the Stratum
18 (NC)	261	194	1.345	123
19 (NE)	117	86	1.360	35
20 (NY)	116	93	1.247	53
21 (OH)	70	58	1.207	35
22 (OK)	68	55	1.236	31
23 (OR)	99	74	1.338	55
24 (PA)	75	64	1.172	44
25 (TX)	308	254	1.213	122
26 (VA)	114	90	1.267	40
27 (WA)	217	162	1.340	102
28 (WI)	226	171	1.322	98
29 (Other States)	615	462	1.331	261
Total	5559	4214		2329

A.2.2.3 National Estimates and Standard Errors

This section presents the general methodology and equations for estimating national totals, national means, and their standard errors, from the responses to the screener questionnaire.

Estimates of national totals were obtained for each characteristic and domain of interest by multiplying the reported value by the screener weight and by summing all weighted values for the facilities that belong to the domain of interest *k*:

$$\hat{y}_k = \sum_i w_i y_i I_{i \in k} \qquad \text{(A-4)}$$

Where $I_{i \in k}$ is one if facility *i* is in domain *k* and zero otherwise. For example, if the domain of interest was 'Facilities in Western USDA Region,' y_i was the trout production at each facility *i*, and w_i was the screener weight for that facility, then $\hat{y}_k$ was the estimate of trout production for facilities in the Western USDA region.

Similarly, ratio estimates (for example, means and percentages) in a given domain *k* were obtained as a ratio of estimates of two total values. For example, the average trout production in the Western USDA region was the ratio of the estimate of trout production, $\hat{y}_k$ in that region, and the estimate of the number of facilities in that region producing trout, n_k:

$$\bar{y}_k = \frac{\hat{y}_k}{n_k} = \frac{\sum_i w_i y_i I_{i \in k}}{\sum_i w_i I_{i \in k}} \qquad \text{(A-5)}$$

After calculating the national estimates, EPA calculated standard errors (s.e.) of its estimates using a jackknife replication method (Wolter, 1985). Under the jackknife replication method, a series of samples (called jackknife replicates) are selected from all responses (n). EPA created 100 replicates to obtain 99 degrees of freedom which EPA considered to be adequate for the statistical estimates while resulting in reasonably sized data files for the replicates. Each facility response was randomly assigned a number between 1 and 100. The first replicate used the responses from all facilities except those assigned to group 1. The other replicates were derived in a similar way by excluding the values for a different group each time. The replicate weights were used to adjust the replicate sample size for the missing group. That is, if there were 100 responses and 10 responses were randomly assigned to group r, then the replicate weight adjustment, $w_{(r)}$, was the ratio, 1.11, of the 100 responses in the full sample to the 90 responses (n(r)=90) in the replicate sample. In this way, a series of replicate weights were generated for each facility response, which together with the screener weight were used to calculate national estimates and averages:

$$\hat{y}_{k(r)} = \sum_i w_i w_{(r)} y_i I_{i \in k} \qquad \text{(A-6)}$$

$$\bar{y}_{k(r)} = \frac{\hat{y}_{k(r)}}{n_{k(r)}} = \frac{\sum_i w_i w_{(r)} y_i I_{i \in k}}{\sum_i w_i w_{(r)} I_{i \in k}} \qquad \text{(A-7)}$$

In order to illustrate how the sampling errors are calculated, let $\bar{y}$ be the weighted national average estimate of a characteristic y (e.g., average trout production at facilities that produce trout). If $\bar{y}_{(r)}$ is the corresponding estimate calculated using the facility responses for all groups except group r, then the estimated variance of $\bar{y}$ is given by the following formula:

$$\text{var}(\bar{y}) = \frac{99}{100} \sum_{r=1}^{100} \left(\bar{y}_{(r)} - \bar{y} \right)^2 \qquad \text{(A-8)}$$

where the summation extends over all 100 jackknife replicates that were formed from the screener responses. The standard error is then the square root of the variance:

$$s.e. = \sqrt{\text{var}(\bar{y})} \qquad \text{(A-9)}$$

In Attachment A.3, the tables provide various estimates and their standard errors. These standard errors can be used to compute 95 % confidence intervals around the estimate. These intervals are given by:

$$confidence\ \text{int}erval = \bar{y} \pm (1.96 \times s.e.) \qquad \text{(A-10)}$$

A.3 DETAILED QUESTIONNAIRE (PHASE 2)

This section describes the detailed questionnaire that was distributed in Phase 2 of EPA's survey of the AAP industry. Section A.3.1 describes the sample design and sample selection for the detailed questionnaire based upon the responses to the screener questionnaire in Phase 1. Section A.3.2 describes the methods that EPA used in developing national estimates from the responses to the detailed questionnaire.

A.3.1 Sample Design: Stratified Random Sample

After reviewing the results from the screener questionnaire, EPA decided that the information from the detailed questionnaire was needed for only a subset of the AAP facilities. Because the proposed rule is applicable only to concentrated aquatic animal production (CAAP) facilities, EPA was particularly interested in facilities, classified as either Commercial, Government, Research, or Tribal, and subject to the current NPDES regulations. (40 CFR 122.24 and Appendix C to Part 122.) According the NPDES regulations, CAAP facilities can be in either of two categories: cold water or warm water. The cold water species category includes ponds, raceways, or other similar structures which discharge at least 30 days per year but does not include: facilities which produce less than 9,090 harvest weight kilograms (approximately 20,000 pounds) per year of trout or salmon; and facilities which feed less than 2,272 kilograms (approximately 5,000 pounds) during the calendar month of maximum feeding. The warm water category includes ponds, raceways, or other similar structures which discharge at least 30 days per year but does not include: closed ponds which discharge only during periods of excess runoff; or facilities which produce less than 45,454 harvest weight kilograms (approximately 100,000 pounds) per year of any species except trout and salmon. Although EPA excluded ponds from the proposed rule, EPA determined that it needed additional information from facilities with ponds and large production volumes to evaluate whether EPA had appropriately excluded such facilities from the proposed rule. EPA also considered aquariums to assess concerns from interested parties, particularly with respect to drug and chemical use. EPA selected these based upon the facility name, responses to questions 4 and 5, and additional information from an industry trade association.

After considering these factors, EPA determined that it should sample facilities meeting one of the following six criteria:

1. Aquariums.

2. Production includes alligators and total biomass exceeds 100,000 pounds.

3. Production includes trout or salmon and total biomass exceeds 20,000 pounds.

4. Predominant production method is ponds; predominant species is catfish; and total biomass exceeds 2,200,000 pounds.

5. Predominant production method is ponds; predominant species is shrimp, tilapia, other finfish, or hybrid striped bass; and total biomass exceeds 360,000 pounds.

6. Predominant production method is any except ponds; and total biomass exceeds 100,000 pounds.

By applying these criteria, EPA identified 539 facilities with these characteristics from the screener questionnaire responses. In developing the sample selection criteria, EPA determined each facility's predominant species and predominant production method as explained in Section A.2.2.1, except that it excluded molluscan shellfish from its determination of the predominant species.[5] EPA then classified the 539 facilities into 44 strata which were defined by facility type (commercial, government, research, or tribal),[6] the predominant species, and predominant production.

In calculating the sample sizes, EPA used a common method for estimating sample sizes that is based upon the binomial distribution (see, for example, Cochran (1977)). The binomial distribution applies to situations where there are only two possible outcomes. For example, there are only two outcomes (yes or no) to a dichotomous question such as 'Does any of this water go to a publicly owned treatment works.' Because the assumption that 50 % of the respondents would answer "Yes" results in the largest possible variance for the binomial distribution and the largest possible sample size, EPA assumed that the probability of one outcome would be 0.5 (i.e., 50 % would select 'Yes' and 50 % select 'No.') This probability is written as 'p=0.5.' EPA used this probability (p=0.5) and its precision targets to derive the sample sizes. EPA's criteria for its sample can be summarized as follows:

1. For estimates for each stratum: a 95% confidence interval for p=0.5 is (0.2, 0.8); *and*

[5] Before selecting the sample for the detailed questionnaire, EPA evaluated the impact of its 'approximate' conversion factors in the total biomass calculations described in Section A.2.2.1. Because it had identified facilities with production close to the cutoff for inclusion into the selected strata and expended additional effort to obtain more precise conversion factors, the use of approximate conversion factors had relatively little effect.

[6] Facility type was determined by the facility's response to question 4 of the screener questionnaire. If the facility type was missing (7 cases) or indicated as being 'Other,' EPA excluded these facilities from consideration for the detailed questionnaire.

2. For overall estimates (i.e., of the entire population meeting the criteria above): a 95% confidence interval for p=0.5 is (0.45, 0.55); *and*

3. No one facility unduly influences the overall estimate.

To achieve the desired precision, EPA determined that information should be collected from 263 of the 539 facilities in the 44 strata. For 34 strata with five or fewer facilities, EPA determined that a census was appropriate because of the relatively small sample sizes, and thus, selected the 163 facilities in those strata. (Of these 34 strata, 20 strata contained only one facility.) For the other 10 strata, EPA selected 200 of the 376 facilities. Table A.2 lists the variables defining each stratum, the number of facilities in the stratum (N_h), the number of facilities in the sample (n_h), and the sampling weight. The number of facilities are based on the responses to the screener questionnaire, without adjusting for non-response. As shown in Table A.2, the sampling weights are fairly consistent, ranging from 1.0 to 2.6. (Although aquariums and alligators are not listed in Table A.2, facilities selected for the sample included facilities that were aquariums and alligator farms.)

In selecting the sample for each of the 10 strata, EPA selected the first n_h facilities in alphabetical order. Assuming that the information collected in the detailed questionnaire is not correlated with the alphabetical ordering of the facilities, the sample can be treated as a random statistical sample. By examining the production levels calculated from the screener questionnaire responses in each stratum, the sample appears to be representative of the population in each of the 10 strata (Westat, 2002c). After selecting the sample, EPA identified 8 of the 539 facilities as being duplicates of other facilities; however, they either were not selected for the sample or were only selected once. EPA also identified another facility that should have been excluded from consideration for the detailed questionnaire, because it did not meet the selection criteria. Although the facility was one of the 263 selected to receive the detailed questionnaire, it has been removed from the sample. EPA has concluded that the 262 remaining facilities in the sample will provide acceptable precision estimates for the 530 facilities.

Table A.2 Sampling Strata for Detailed Questionnaire

Facility Type	*Predominant Species*	*Predominant Production Method*	*Number of Facilities (based on Screener Responses)* N_h	*Number of Sampled Facilities* n_h	*Sampling Weight* $DQ_h=N_h/n_h$
Commercial	Catfish	Flow through	<5	all	1.0
		Ponds	50	20	2.5
	Other	Flow through	<5	all	1.0
		Other	<5	all	1.0
		Ponds	<5	all	1.0
		Recirculating	<5	all	1.0
	Trout	Flow through	135	52	2.596
		Net pens	<5	all	1.0
	Salmon	Flow through	16	8	2.0
		Net pens	10	7	1.429

Facility Type	*Predominant Species*	*Predominant Production Method*	*Number of Facilities (based on Screener Responses)* N_h	*Number of Sampled Facilities* n_h	*Sampling Weight* $DQ_h=N_h/n_h$
		Recirculating	<5	all	1.0
	Striped Bass	Flow through	<5	all	1.0
		Ponds	<5	all	1.0
		Recirculating	<5	all	1.0
	Tilapia	Flow through	<5	all	1.0
		Recirculating	12	7	1.714
	Other Finfish	Flow through	<5	all	1.0
		Net pens	<5	all	1.0
		Ponds	8	6	1.333
		Recirculating	<5	all	1.0
	Baitfish	Recirculating	<5	all	1.0
	Ornamentals	Flow through	<5	all	1.0
		Recirculating	<5	all	1.0
	Shrimp	Flow through	<5	all	1.0
		Ponds	<5	all	1.0
		Recirculating	<5	all	1.0
Government	Catfish	Flow through	<5	all	1.0
		Ponds	<5	all	1.0
	Trout	Flow through	157	61	2.574
		Other	<5	all	1.0
		Ponds	<5	all	1.0
	Salmon	Flow through	64	25	2.560
		Net pens	<5	all	1.0
		Recirculating	<5	all	1.0
	Striped Bass	Ponds	<5	all	1.0
	Other Finfish	Flow through	<5	all	1.0
		Ponds	12	7	1.714
Research	Catfish	Ponds	<5	all	1.0
	Other	Recirculating	<5	all	1.0
	Trout	Flow through	<5	all	1.0
	Other Finfish	Recirculating	<5	all	1.0
Tribal	Trout	Flow through	<5	all	1.0
	Salmon	Flow through	10	7	1.429
	Other Finfish	Ponds	<5	all	1.0
Totals			537	263	

A.3.2 Final Survey Weights

EPA used the information collected by the detailed questionnaires to re-estimate the costs and benefits associated with the proposed regulatory options and the NODA options. This section provides an overview of EPA's development of survey weights that were used in the final analyses. These final analyses are described elsewhere in this document and the EIA.

Weighting the data allows inferences to be made about all eligible facilities, not just those included in the sample, but also those not included in the sample or those that did not respond to the either the screener or detailed questionnaire.

For the final rule, EPA has applied survey weights that are slightly different than those used for the NODA which were adjusted for non-response within each sampling strata.[7] As a first cut, adjusting by sampling strata is a reasonable approach given the stratified sample design. Dividing the sample into many strata has the advantage that the true response probabilities are likely to be relatively constant within a strata, because the facilities have some of the same characteristics which might lead to similar types of behavior such as responding or not responding to the questionnaire. As a result, the non-response adjustment within each strata is likely to be close to the correct adjustment for all facilities in the strata. However, when the number of facilities in a strata is small, the calculated non-response adjustment factor is variable (or imprecise). The imprecision of the non-response adjustment will contribute variability to the estimates. If strata are combined (or "collapsed" as statisticians generally describe it) to create fewer strata with more facilities, the non-response adjustment factor in each strata will be more precise (the objective is to collapse strata which have similar probabilities of response). At the same time, the within-strata true response probabilities may differ more from the estimated value because the collapsed strata now include more different kinds of facilities. This difference contributes to bias in the survey estimates. Thus, there is a trade-off. After examining the sample sizes in the strata, EPA noted that many of the sampling strata have less than 10 facilities. For this reason, EPA determined that using collapsed strata, containing a larger number of facilities, to determine the non-response adjustments would be more appropriate than adjusting by strata as it had for the NODA analyses. Further, EPA used a stepwise logistic regression to determine which factors or 2-way interactions of factors were significant predictors of non-response. After combining Other, Baitfish, and Ornamentals into a general Other category, ownership (with Tribal and Research facilities collapsed into one group) was the only significant predicator of non-response. (Westat, 2004). This finding provided additional support for EPA's determination that ownership, rather than strata, should be used to adjust for non-response. The final weights and the NODA weights have a correlation of about 0.96, and thus, the results are similar, regardless of which set of weights are used.

Table A.3 shows the number of surveys sent and the number received by ownership category, excluding nine facilities that returned incomplete responses. As explained in the following paragraphs, EPA adjusted the response status in two situations.

First, for a few cases, the sampling frame listed facilities twice, and EPA has excluded the duplicate entry of each pair from the totals in this table. Generally, EPA did not send the duplicate questionnaire to the facility. For the one exception, for purposes of

[7] The NODA survey weights also included an adjustment for strata without any respondents. While such an adjustment is not necessary or correct, it has no effect on estimates of means and proportions and a relatively small effect on the estimate of totals, such as numbers of facilities in the population that potentially would be affected by a regulation. For the final survey weights, EPA has chosen a method that does not include this adjustment.

developing the final survey weights, EPA simplified the calculations by treating this single exception as a non-respondent.

Second, a few completed questionnaires contained information on more than one facility. Subsequently, EPA separated that information into several questionnaires so that a single questionnaire represented an individual facility. These questionnaires with multiple facility data resulted in eight additional facilities contributing relevant data to the detailed survey. However, for purposes of the costing analyses, for each original questionnaire, EPA combined the data values from the multiple facilities into a single value. EPA then applied the survey weight associated with the original questionnaire to that single, combined, value in using the data to calculate national estimates.

Table A.3 Detailed Questionnaire: Response Rates

Ownership Category	*Number Sent*	*Number Received*	*Response Rate* ($R_{ownership}$)
Commercial	134	107	0.79851
Government	102	97	0.95098
Tribal or Research	15	11	0.73333

The final detailed questionnaire weight W_i for facility i in stratum h and ownership category g can be written as:

$$W_i = \frac{DQ_h}{R_{ownership}} \times w_i \quad \text{(A-11)}$$

where DQ_h is the sampling weight from Table A.2; $R_{ownership}$ is the response rate from Table A.3; and w_i is the screener weight from Table A.1 adjusted for non-response as in equation (A-3). EPA then used these adjusted survey weights to calculate national estimates as described in Section A.2.2.3. These estimates are presented elsewhere in this document and the record for the final rule.

REFERENCES

Brick, J.M., D. Morganstein, R. Valliant. (2000). "Analysis of Complex Sample Data Using Replication." Westat (www.westat.com).

Cochran, W.G. (1977). *Sampling Techniques*. New York: Wiley.

Groves, R.M. and M.P. Couper. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

USDA (2000). 1998 USDA Census of Aquaculture. Located at DCN 60605 and http://www.nass.usda.gov/census/census97/aquaculture/aquaculture.htm.

USEPA (2001). "Screener Questionnaire for the Aquatic Animal Production Industry." DCN 10001 in the proposal record. Also at http://www.epa.gov/waterscience/guide/aquaculture/screenersurvey.pdf.

USEPA (2002). "Detailed Questionnaire for the Aquatic Animal Production Industry." DCN 10002 in the proposal record.

Westat (2002a). "Alternative weighting plan." DCN 50050 in Section 10.3.

Westat (2002b). "Logistic Regression Results." DCN 50058 in Section 10.3.

Westat (2002c). "Analysis of Variance Results" DCN 50057 in Section 10.3.

Westat (2004). "Sampling Weights for the Detailed Survey." DCN 51400 in Section 24.3.

Wolter, K. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.

ATTACHMENT A-1.
COMPARISON OF USDA CENSUS OF AQUACULTURE AND EPA SCREENER QUESTIONNAIRE

	USDA Census of Aquaculture	*EPA Screener Questionnaire*	*Difference*
Primary Objective	Economic description of the industry	Data for regulatory analysis	
Year	1998	2000	Two years
Target Population	All “aquaculture farms” from which aquaculture products were sold, or produced for restoration or conservation purposes during the census year (1998).	All "facilities" in the Aquatic Animal Production Industry which answer "Yes" to the question "Do you produce (grow) aquatic animals (fish, shellfish, or other aquatic animals) at this facility?"	EPA did not require that any products be sold for the farm to be included in its population. The USDA Census generally excluded farms that did not sell its products (e.g., state hatcheries). The USDA Census included "other aquaculture products" including algae. The EPA Screener excluded algae and non-animal products. While the USDA includes such farms, only 20 farms report algae and sea vegetables production (Table 19 in USDA (appears to be only about 20 farms in the USDA count.
Frame Source	Answered positively a 1997 Census of Agriculture question on whether there were "fish and other aquaculture products" in 1997. This list was supplemented by other USDA information and lists of State and Federal fish hatcheries.	A mailing list of 5988 facilities was constructed from Dun & Bradstreet, state lists, tribal information, aquaculture journals, various associations, the internet, and aquaculture facilities identified by respondents. The D&B SIC codes included: 0273 (animal aquaculture), 0279 (animal specialties), and 0921 (fish hatcheries and preserves).	Although constructed differently, it is difficult to say how the resulting lists of farms/facilities might differ. Both frames may miss some aquaculture farms or facilities. Total revenue of facilities in the D&B database exceeded that of the Census by about 10%. Both estimates of total revenue were about $1.0 billion. It was concluded that the facilities missed by D&B probably were quite small.
Survey Design	Census	Census	None.
Frame Size	Not available, assumed to be 4028 or more addresses based upon the reported number of farms.	5988 addresses	Differences are due to how the frames were constructed, differences in the target population, or changes in the industry over time.

	USDA Census of Aquaculture	*EPA Screener Questionnaire*	*Difference*
Instrument	Mailed questionnaire augmented with telephone and personal interviews. Telephone calls and personal interviews were used to collect data from non-respondents.	Mailed questionnaire augmented with follow-up phone calls to clarify data. Reminder letters were sent to non-respondents. A limited effort was made to correct invalid addresses.	USDA had little non-response due to intensive follow-up of non-respondents. Screener results were weighted to adjust for non-response to calculate national estimates.
Number of Respondents doing Aquaculture	Apparently 4028, assuming no non-response due to non-response follow-up. Responses to some questions were imputed.	2329	Differences are primarily due to screener non-response as well as frame under-coverage for either questionnaire, changes in the industry over time, and inclusion of non-animal production in the USDA Census.
Estimated Number of Aquaculture facilities nationally	4028 aquaculture farms	3075 AAP facilities	Differences may be due to frame under-coverage for either questionnaire, changes in the industry over time, and inclusion of non-animal production in the USDA Census.
Scope	Collected detailed information relating to on-farm aquaculture practices, size of operation based on water area, production, sales, method of production, sources of water, point of first sale outlets, cooperative agreements and contracts, and aquaculture distributed for restoration or conservation purposes.	Collected information on the type of facility (commercial, Government, Tribal, etc.), quantities of animals produced in 2000 by species and size category, production methods used, whether water from the facility left the property and whether to a POTW and/or with an NPDES permit, and a description of pollution control practices.	Comparable values include production methods, species produced and some production information.
Production Totals for Selected Species	National estimate: Catfish: 593 million pounds Trout: 63 million pounds	Weighted national estimate: Catfish: 637 million pounds Trout: 121 million pounds	For catfish and trout, total screener production is somewhat larger than from the USDA Census. Screener estimates are based on unit conversion assumptions. Comparisons for other species would require additional assumptions. Differences may be due to changes over time and under coverage in the two frames.

ATTACHMENT A-2.
CONVERSION FACTORS FOR SCREENER QUESTIONNAIRE RESPONSES

Table A2.1. Biomass Calculations for Predominant Species: Pounds-to-Count Conversion Factors

Species Code1	*SPECIES*	*SIZE (Size Category from Question 5 in the Screener)2*					
		Foodsize (1)	*Stockers (2)*	*Fingerlings (3)*	*Seed Stock (6)*	*Brood-stock (7)*	*Fry (4)*
1	Catfish	1.5	0.18	0.0334		4.31	
2	Trout	1	0.32	0.035		2.5	
3	Salmon	5	0.32	0.035		10	
4	Striped Bass	1.75	0.33	0.06		5	
5	Tilapia	1.75	0.32	0.035		2.5	
6	Other Finfish (except as listed)	1	0.32	0.035		2.5	
6-15	bass - smallmouth and largemouth	2.00					
6-19	Crappie	1.13					
6-20	Eel	4.62					
6-24	Paddlefish	2.00					
6-26	Perch	0.59					
6-27	Saugeye	1.00					
6-29	Sturgeon	45.00					
6-30	Sucker	2.19					
6-31	Sunfish (including bluegill and panfish)	0.25					
6-33	Walleye	3.00					
6-34	Whitefish	2.50					
6-35	Pike	4.63					
6-69	Shad (including threadfin)	2.50					
6-71	Charr	2.00					
6-73	Amberjack	75.00					
6-74	Bream	0.33					
6-75	Shell cracker	0.50					
7	Baitfish (except smelt)	0.01					
7-48	Smelt	0.19					
8	Ornamentals (except carp)	0.01					
8-17	Carp (includes koi, white amur)	4.00					
9	Shrimp	0.0444			6.6E-06	0.1	
10	Crawfish	0.0444				0.08	
11	Other Crustaceans	0.10					
12	Molluscan shellfish	0.10					
13	Other (except as listed)	1.00					
13-14	Alligators (and caimen)	13.00					
13-21	Frogs and tadpoles	0.13					
13-32	Turtles					3.5	.03

Table A2.2. Total Biomass Calculations: Foodsize-to-Other Sizes Conversion Factors (when not specified in Table A2.1)

Size Code from Question 5 in the Screener	*Size*	*Food Size Multiplier*
1	Foodsize	1.0000
2	Stockers	0.1418
3	Fingerlings	0.0214
4	Fry	0.0014
5	Eggs	0.00001
6	Seed stock	0.0001
7	Brood size	3.4247
8	Other	0.1000

Table A2.3. Determination of Predominant Production Method: EPA's *Assumed* Hierarchy of Most to Least Common Production Method

Purpose	*Predominant Species*	*Most Common to Least Common Production Method*[1]
Screener Questionnaire Data Analysis[2] (See Attachment A-3)	Catfish	Ponds, Flow through, Recirculating, Other
	Trout	Flow through, Recirculating, Ponds, Net Pens
	Salmon	Net pens, Flow through, Recirculating
	Striped Bass	Ponds, Recirculating, Flow through
	Tilapia	Recirculating, Flow through, Ponds
Detailed Questionnaire Sample Selection[3]	Catfish	Ponds, Flow through, Net pens, Recirculating, Other
	Trout	Flow through, Ponds, Recirculating, Other, Net Pens, Floating aquaculture
	Salmon	Flow through, Net pens, Recirculating, Ponds, Other, Floating aquaculture
	Striped Bass	Ponds, Flow through, Recirculating, Net pens, Other
	Tilapia	Recirculating, Flow through, Ponds, Net pens, Other, Floating aquaculture
	Other Finfish	Ponds, Flow through, Recirculating, Net pens, Other, Floating aquaculture
	Baitfish	Ponds, Flow through, Recirculating
	Ornamentals	Ponds, Recirculating, Flow through, Other, Net pens, Floating aquaculture
	Shrimp	Ponds, Recirculating, Flow through, Other, Net pens, Floating aquaculture
	Crawfish	Ponds, Net pens, Flow through, Recirculating, Other
	Other crustaceans	Recirculating, Flow through, Floating aquaculture, Ponds
	Other	Ponds, Recirculating, Other, Flow through, Net pens, Floating aquaculture

[1] The production methods (e.g., Other) are from the choices provided in question 6 of the screener questionnaire.
[2] This hierarchy was based upon sources other than the screener questionnaire responses.
[3] This hierarchy is based upon a data analysis of the screener questionnaire responses. EPA acknowledges that floating aquaculture is unlikely to be used as a production method for certain species, and EPA plans additional review of these questionnaire responses.

Table A2.4. Revenue Calculations: Prices for Species by Size[1]

Species	*Size*	*Prices*	*USDA Table (page)* [2]
Catfish	Foodsize	$0.74/lb	8 (20)
	Stockers	$1.03/lb	8 (21)
	Fingerlings/Fry	$1.66/lb	8 (22)
	Brood Stock	$0.91/lb	8 (19)
Trout	Foodsize	$1.06/lb	9 (24)
	Stockers	$2.29/lb	9 (25)
	Fingerlings	$162.16/1000 fish eggs	9 (26)
Salmon	Foodsize (except Alaska)	$2.00/lb	[3]
	Foodsize (Alaska)	$0.23/lb	12 (39)
	Fingerlings/Fry	$0.17/lb	12 (40)
Striped Bass	Foodsize	$2.44/lb	12 (34)
	Fingerlings/Fry	$0.26/lb	12 (35)
Tilapia	Foodsize	$1.70/lb	12 (41)
	Fingerlings	$0.11/fish	12 (42)

[1] EPA included *only* the listed species/size categories in its revenue calculations. Of those categories, EPA included only those responses that were reported in dollars sold, in pounds (applying the above conversion factors), or counts that could be converted to pounds using the conversion factors in Table A2.1.

[2] See USDA (2000).

[3] EPA adjusted the national average provided in Table 12 (p. 39) to obtain a value that did not include Alaska as follows:

(National total sales - Alaska sales)/(National total quantity - Alaska quantity)
= ($103,583,000 - $16,340,000)/(110,588,000 lbs - 70,129,000 lbs)
= $2.16/lb which EPA rounded to $2.00/lb

ATTACHMENT A-3.
NATIONAL ESTIMATES BASED ON SCREENER QUESTIONNAIRE

The following tables provide national estimates (i.e., adjusted for non-response) of the responses to the screener questionnaires. Each table presents estimates for different types ('domains') of facilities, such as facilities in each USDA region or facilities using each production method. The facility domains are shown in the left column. Within each domain, Tables A3.1 through A3.8 show the number of facilities, percent of facilities, and total aquatic animal production. The total aquatic animal production is the total production of all species across all facilities in the domain. In contrast, Table A3.9 shows the total production of only the species used to define the domain rather than all species.

In some tables in this attachment, EPA has not presented the totals, because some facilities were placed in more than one category. For example, Table A3.7 provides the number of facilities and their production for each production method. Thus, if a facility has ponds and flow-through systems, the facility and its production would be counted under both production methods.

Table A3.1. USDA Region

	Number of Facilities		*Percent of Facilities*		*Production (thousands of pounds)*	
Region	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
NORTHEASTERN	452	18	14.7%	0.5%	74,673	15,890
SOUTHERN	1393	30	45.3%	0.7%	820,946	112,800
NORTH CENTRAL	485	18	15.8%	0.5%	27,138	5,978
WESTERN	664	20	21.6%	0.6%	258,830	96,884
TROPICAL	80	9	2.6%	0.3%	7,382	4,088
ALL	3075	46	100.0%		1,190,000	150,200

Table A3.2. Facility Type

	Number of Facilities		*Percent of Facilities*		*Production (thousands of pounds)*	
Facility Type	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Commercial	2384	44	77.5%	0.9%	1,060,000	146,700
Government	447	23	14.5%	0.7%	102,046	18,743
Research	67	9	2.2%	0.3%	1,738	724
Tribal	29	6	1.0%	0.2%	2,356	782
Other	147	14	4.8%	0.4%	20,762	5,266
ALL	3075	46	100.0%		1,190,000	150,200

Table A3.3. Predominant Species[1]

	Number of Facilities		*Percent of Facilities*		*Production, All Species (thousands of pounds)*	
Predominant Species	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Catfish	739	29	24.0%	0.9%	613,627	103,700
Trout	707	30	23.0%	0.9%	98,373	19,398
Salmon	197	13	6.4%	0.4%	111,756	21,466
Striped Bass	91	11	3.0%	0.3%	17,788	5,538
Tilapia	129	14	4.2%	0.4%	12,599	3,843
Other Finfish	376	21	12.2%	0.7%	31,542	9,313
Baitfish	116	13	3.8%	0.4%	8,371	2,220
Ornamentals	173	13	5.6%	0.4%	8,800	2,465
Shrimp	54	8	1.7%	0.3%	11,702	4,620
Crawfish	38	7	1.2%	0.2%	629	310
Other crustaceans	15	5	0.5%	0.1%	160	129
Molluscan shellfish	274	17	8.9%	0.5%	139,231	97,493
Other	168	13	5.5%	0.4%	134,390	53,166
ALL	3075	46	100.0%	0.0%	1,190,000	150,200

[1] The predominant species is the species with the largest production at a facility. Each facility has only one predominant species.

Table A3.4. Predominant Production Method

	Number of Facilities		*Percent of Facilities*		*Production (thousands of pounds)*	
Predominant Production Method	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Ponds	1561	38	50.8%	1.0%	763,380	109,500
Flow through raceways, ponds, or tanks	960	33	31.2%	0.9%	278,181	98,100
Recirculating systems	228	18	7.4%	0.6%	61,256	24,797
Net pens or cages	40	7	1.3%	0.2%	45,455	17,670
Floating aquaculture or bottom culture	233	17	7.6%	0.5%	37,564	11,463
Other	53	8	1.7%	0.3%	3,134	2,003
ALL	3075	46	100.0%	0.0%	1,190,000	150,200

Table A3.5. Water Discharge Status to POTW[1]

	Number of Facilities		*Percent of Facilities*		*Production (thousands of pounds)*	
Does water go to a POTW?	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Water leaves to POTW	127	13	4.1%	0.4%	9,242	3,583
Water leaves, not to POTW	1981	39	64.5%	1.0%	1,030,000	142,400
Water does not leave	954	35	31.0%	1.0%	147,904	39,775
No answer	13	4	0.4%	0.1%	474	324
ALL	3075	46	100.0%	0.0%	1,190,000	150,200

[1] The responses in the table combine the answers to questions 7 and 8 in the questionnaire.

Table A3.6. Water Discharge Status and NPDES Permits[1]

	Number of Facilities		*Percent of Facilities*		*Production (thousands of pounds)*	
Does water go to a POTW?	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Water leaves, facility has NPDES permit	541	27	17.6%	0.8%	278,103	98,129
Water leaves, No NPDES permit	1565	35	50.9%	1.0%	762,451	110,600
Water does not leave	954	35	31.0%	1.0%	147,904	39,775
No answer	14	5	0.5%	0.1%	511	324
ALL	3075	46	100.0%	0.0%	1,190,000	150,200

[1] The responses in the table combine the answers to questions 7 and 9 in the questionnaire.

Table A3.7. Production Method[1]

	Number of Facilities		*Percent of Facilities*		*Production (thousands of pounds)*	
Production Method	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Ponds	1860	43	60.7%	1.0%	786,298	104,400
Flow through raceways, ponds, or tanks	1358	36	44.3%	1.0%	394,321	101,100
Recirculating systems	610	26	19.9%	0.8%	129,575	29,385
Net pens or cages	262	17	8.6%	0.6%	71,454	19,388
Floating aquaculture or bottom culture	248	16	8.1%	0.5%	38,315	11,296
Other	155	13	5.1%	0.4%	19,432	9,026

[1] If a facility reports using more than one production method, the facility is included in the table totals for each production method used. Therefore the sum of the column for the number of facilities is greater than the number of facilities represented by the data, and the same is true for the production numbers. Thus, the totals are not presented.

Table A3.8. Species[1]

	Number of Facilities		*Percent of Facilities*		*Production, All Species (thousands of pounds)*	
Species	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Catfish	901	30	29.3%	0.9%	637,211	99,896
Trout	818	28	26.6%	0.8%	120,600	23,065
Salmon	277	16	9.0%	0.5%	128,305	23,860
Striped Bass	155	13	5.1%	0.4%	24,817	6,168
Tilapia	178	15	5.8%	0.5%	24,005	7,236
Other Finfish	644	28	21.0%	0.8%	75,781	14,802
Baitfish	259	18	8.4%	0.6%	30,044	8,485
Ornamentals	267	19	8.7%	0.6%	24,031	7,881
Shrimp	73	10	2.4%	0.3%	12,957	4,623
Crawfish	83	9	2.7%	0.3%	12,353	6,430
Other crustaceans	24	6	0.8%	0.2%	293	170
Molluscan shellfish	303	18	9.9%	0.6%	140,308	96,204
Other	156	11	5.1%	0.4%	135,762	45,566

[1] If a facility produces more than one species, the facility is included in the table totals for each species produced. Therefore, the sum of the column for the number of facilities is greater than the number of facilities represented by the data, and the same is true for the production numbers. Each row provides the total production for all species at those facilities having the individual species in the left-hand column. See Table A3.9 for total production of just the individual species at those facilities.

Table A3.9. Species[1]

	Number of Facilities		*Percent of Facilities*		*Production of Listed Species (thousands of pounds)*	
Species	***Estimate***	***Standard error***	***Estimate***	***Standard error***	***Estimate***	***Standard error***
Catfish	901	30	29.3%	0.9%	613,569	99,705
Trout	818	28	26.6%	0.8%	97,381	20,420
Salmon	277	16	9.0%	0.5%	112,514	23,443
Striped Bass	155	13	5.1%	0.4%	17,848	5,228
Tilapia	178	15	5.8%	0.5%	13,771	3,870
Other Finfish	644	28	21.0%	0.8%	26,888	6,317
Baitfish	259	18	8.4%	0.6%	10,781	2,975
Ornamentals	267	19	8.7%	0.6%	10,054	2,828
Shrimp	73	10	2.4%	0.3%	11,634	4,501
Crawfish	83	9	2.7%	0.3%	754	309
Other crustaceans	24	6	0.8%	0.2%	131	98
Molluscan shellfish	303	18	9.9%	0.6%	139,321	96,225
Other	156	11	5.1%	0.4%	134,324	45,584

[1] The total production is the production for the species listed in the left column. See Table A3.8 for the total facility production which includes production of all other species at the facilities producing that species in the left column.

Table A3.10. For Selected Species, Number of Facilities by Predominant Production Method

Species	*Predominant Production Method*	*National Estimate*[1]	*Responses*
Catfish	**Ponds**	861.13	649
	Flow Through & Not(Ponds)	26.87	20
	Recirculating & Not(Ponds or Flow Through)	ND	ND
	Other, Not(Ponds, Flow Through, or Recirculating)	ND	ND
	All systems	900.91	679
Trout	Flow Through	735.07	569
	Recirculating & Not(Flow Through)	17.22	13
	Ponds & Not(Flow Through or Recirculating)	60.94	47
	Net Pens & Not(Flow Through, Recirculating, or Ponds)	ND	ND
	Missing Production Information	ND	ND
	All systems	818.40	633
Salmon	Net Pens	46.98	35
	Flow Through & Not(Net Pens)	219.25	166
	Recirculating & Not(Net Pens or Flow Through)	ND	ND
	Other, Not(Net Pens, Flow Through, or Recirculating)	ND	ND
	All systems	276.65	201
Shrimp	Ponds	55.57	42
	Recirculating & Not(Ponds)	ND	ND
	Flow through & Not(Ponds or Recirculating)	ND	ND
	All systems	72.53	55
Tilapia	Recirculating	119.42	90
	Flow Through & Not(Recirculating)	35.61	26
	Ponds & Not(Recirculating or Flow Through)	ND	ND
	Missing Production Information	ND	ND
	All systems	176.08	132
Sportfish (other Finfish)	Ponds	557.95	432
	Flow Through & Not(Ponds)	59.28	45
	Recirculating & Not(Ponds or Flow Through)	20.68	16
	Other, Not(Ponds, Flow Through, or Recirculating)	6.33	5
	All systems	644.24	498
Striped Bass/ Hybrid Striped Bass	Ponds	129.33	99
	Recirculating & Not(Ponds)	19.59	15
	Flow through & Not(Ponds or Recirculating)	ND	ND
	Other & Not(Ponds, Recirculating, or Flow Through)	ND	ND
	All systems	155.22	119
Alligator	All systems	41.12	31

[1] Sample sizes masked by 'ND' ('Not Disclosed') indicate there are five or fewer facilities for one or more of the production methods for that specie.

Table A3.11. Estimated Number of Facilities Covered by the Proposed Rule[1]

Predominant Production Method	***Species***	***Size***	***Revenue Classes***			
			Class 1 ≥ $20,000 and <$100,000	***Class 2 ≥ $100,000 and <$500,000***	***Class 3 ≥ $500,000***	***Total ≥ $20,0004***
Flow-through	Trout	Foodsize	92	44	13	149
		Stockers	139	131	39	309
	Salmon	All with $ value	44	52	38	133
	Striped Bass	All with $ value	n/a[2]	ND[3]	ND	ND
	Tilapia	All with $ value	n/a	ND	ND	9
Recirculating	Striped Bass	All with $ value	n/a	ND	ND	ND
	Tilapia	All with $ value	n/a	13	12	26
Net Pens	Salmon	All with $ value	ND	ND	19	32

[1] In the preamble to the proposed rule, EPA discusses six production size categories that correspond with the revenue classifications used in the 1998 USDA Census of Aquaculture (i.e., $1,000-$24,999; $25,000 - $49,999; $50,000 - $99,999; $100,000 - $499,999; $500,000 - $1,000,000; and >$1,000,000) to develop model facilities representing these size ranges for each species evaluated. Because small sample sizes for some revenue categories have small sample sizes, the national estimates are presented here. They are included in the non-public record as DCN50066CBI in Section 10.3.

[2] n/a: not applicable in the proposed rule

[3] ND: Sample sizes masked by 'ND' ('Not Disclosed') indicate there are five or fewer facilities for one or more of the classes in the production method/specie/size category

[4] Due to rounding, totals in this column may differ slightly from the sum of the numbers for the Classes.