Improving the NEFSC clam survey for Atlantic surfclams and ocean quahogs[1]

Larry Jacobson, Dan Hennen (NEFSC, Woods Hole, MA) and

The Clam Survey Working Group[2]

1. Introduction

A working group (WG) consisting of Northeast Fisheries Science Center (NEFSC) and Mid-Atlantic Fishery Management Council (MAFMC) staff, academic partners and interested persons met five times during 2017 to discuss ideas for improving the NEFSC clam survey (Appendix 1). The goals were to develop ideas for improving the precision and utility of survey data used in stock assessments and for using survey resources more efficiently. Preliminary ideas were presented to the Mid-Atlantic Fishery Management Council's Science and Statistical Committee on May 18, 2017 with a final presentation and review of this document scheduled for March 13, 2018. In addition, there were several meetings within NEFSC between Survey Branch and Population Dynamics staff with participation of leadership. This report presents data, analyses and recommendations for consideration by NEFSC.

The survey targets Atlantic surfclams (*Spisula solidissima*) and ocean quahogs (*Arctica islandica*) in federal waters (US Exclusive Economic Zone or EEZ, 3-200 nautical miles from the coast) along the northeastern coast of the US between North Carolina and the US-Canada border on Georges Bank (Figure 1.1). The primary purpose is collection of abundance, biomass, shell length and meat weight data for both species as well as age data for surfclams which are all used in stock assessments for the EEZ stocks. The hydraulic clam dredge sampling gear is designed to reduce bycatch and little useful data are collected for other species. Data are routinely collected for additional studies based on sampling requests submitted by NEFSC and external partners prior to each survey. Beginning in 1997, surveys included experimental field work to test sampling gear and estimate capture efficiency and size-selectivity for survey gear. Substantial environmental data (depth and bottom temperatures) have been collected since 1997 but have been little used to date. The fishing industry uses survey results to locate new fishing grounds.

Surfclams are also found in state waters (< 3 nm from shore) and in federal waters of the Gulf of Maine (NEFSC 2017a). However, the NEFSC clam survey does not currently and is not expected to operate in state or Gulf of Maine waters. New Jersey and New York conduct surveys in state waters but

---

[1] Dedicated to Dr. Stephen Smith (Canada Department of Fisheries and Oceans, emeritus and WG member-deceased) with gratitude and in recognition of a lifetime of service to fisheries science.

[2] See Appendix 1 for working group members and other contributors.

these programs are independent. Federal stock assessments do not include state waters but information about state resources and fisheries is typically provided in appendices.

Surfclams and ocean quahogs in the US EEZ are managed as single stocks. Their distributions are divided into the northern Georges Bank (GBK) and southern assessment *areas* for stock assessment purposes (Figure 1.1). The southern area for both species is divided into five *regions*. The regions (from north to south) are: Southern New England (SNE), Long Island (LI), New Jersey (NJ), Delmarva (DMV) and Southern Virginia/North Carolina (SVA). Area and regional boundaries lie along current NEFSC shellfish strata and are based on historical survey, fishery and stock distribution patterns (Figure 1.1). This report recommends changes to current survey strata and area/regional boundaries.

Stock size is high (>1 million mt) and exploitation rates are low (<0.01 per year) for the surfclam and ocean quahog stocks as a whole (NEFSC 2016; 2017b). However, stock size is lower and exploitation rates are higher in southern regions, particularly on small productive grounds where the fisheries operate. Fishing grounds tend to be smaller than the 10' square resolution of mandatory logbook reports and are difficult to monitor using the NEFSC clam survey, which is designed to track stock size at larger area or regional scales. Survey and fishery logbook data show that the distributions of surfclams and ocean quahogs and their fisheries shifted northward over the last thirty years. Based on surveys, the proportion of total swept area biomass for surfclams on GBK changed from about 10% during the 1980's to almost 60% in 2010 (NEFSC 2017b). During the same time period, the proportion of ocean quahog biomass on GBK changed from about 35% during the 1980's to almost 50% (NEFSC 2017c).

Depth distributions depend on latitude, depth, currents and temperature for both species. Distributions in southern regions DMV and SVA regions changed dramatically starting in the 1990s due to relatively warm water conditions and thermal stress (Weinberg 2005). Surfclam habitat begins in the intertidal zone. About 95% of individuals are currently found at depths shallower than 60 m in the north on GBK and shallower than 40 m in the southern DMV region (NEFSC 2017b). Ocean quahogs use deeper habitat with 95% of individuals currently at depths of at least 55 m on GBK and 30-60 m in more southern regions. The distributions of NJ and DMV surfclams shifted into ocean quahog habitat over the last two decades making survey and fishery catches with both species more common. The probability of catching both in the same tow has clearly increased in the NJ region. Habitats for the two species are more stable and clearly partitioned by depth in the northern GBK, SNE and LI regions.

Surfclams may live to at least 30 y compared to ocean quahogs which may live to at least 200 y (Cargnelli et al. 1999a,b). Stock size is more variable for surfclams because they are shorter lived, grow faster, and live in relatively shallow water where environmental variability more strongly effects recruitment, mortality and growth.

The clam survey follows a random stratified design based on NEFSC shellfish strata developed in the late 1970s (Figure 1.1). The shellfish strata were not designed for clams and are also used for Atlantic sea scallop (*Plactopecten magellanicus*) surveys. Sampling gear and protocols were generally consistent during 1982-2011 with surveys covering both surfclam and ocean quahog habitat every 1-3 years (see Table C6 in NEFSC 2003). In some years, either GBK or the southern area (but not both) were surveyed (NEFSC 2017b; NEFSC 2017c).

Prior to 2012, a small lined hydraulic dredge with a 1.82 m blade was deployed from the NOAA Research Vessel Delaware II. Original protocols specified tows at 1.5 knots for 5 minutes (nominal tow distance 232 m). Sensors were used beginning in 1997 to measure tow distance directly during each dredge tow. The sensor data showed that actual tow distances were much longer than expected (up to 875 m), depth dependent and varied from survey to survey (Weinberg et al. 2002; NEFSC 2009).

In 2012, the survey was moved to a commercial fishing vessel (*F/V Pursuit*) that carries a more efficient hydraulic dredge with a wider 4 m blade and automatic sorting equipment. Current protocols specify tows at 3 knots for 5 minutes (nominal tow distance 154 m, area swept 617 m^2). The heavy dredge, faster free spooling winches and improved sensor equipment reduce uncertainty about tow distance. The commercial dredge and sorting equipment are configured to increase retention of relatively small animals not targeted in the fishery. However, size selectivity is lower for small clams in the current survey than historically (NEFSC 2017b; NEFSC 2017c).

It is important to identify and avoid untowable grounds in the new survey. The commercial dredge used for sampling is highly efficient and robust but difficult to repair at sea. Significant damage due to boulders or rough ground would require a trip to port and loss of at least 2-3 days of sampling. A survey might have to be terminated if a dredge was destroyed. A replacement dredge would be expensive.

Beginning in 1997 with the introduction of sensor equipment, surveys included field studies to estimate capture efficiency and size selectivity. Based on these studies, the original survey dredge captured about 23% of large fully selected surfclams in the path of the dredge while the new commercial dredge captures about 59% (NEFSC 2017b). The original dredge captured about 19% of large fully selected ocean quahogs while the new commercial dredge captures about 61% (NEFSC 2017c). These types of experiments are expensive and will be less important in future because capture efficiency and size-selectivity estimates are available for both the original and new surveys. They will be necessary, however, if and when the survey gear changes. The present survey schedule allows for gear and other types of experiments every third year but there was no need for experimental work during 2017. This report recommends using the time for regular survey sampling.

Decisions about the relative numbers of tows in each stratum during a survey were compromises based on optimal allocations for surfclams and optimal allocations for quahogs with *ad-hoc* adjustments to obtain a minimum number of tows in each stratum. Extra tows were added to important regions, for special studies and for other purposes. The compromise allocations were far from optimal for either species (see below).

Key questions

Discussions centered on the following key and interrelated questions:

1) *Should clam surveys target surfclams and ocean quahogs separately rather than simultaneously?*
2) *Should sampling in poor habitat areas cease, particularly if the two species are surveyed separately?*
3) *Should new species-specific stratification schemes be used if the two species are surveyed separately?*
4) *Is it feasible to survey the entire stock (GBK plus south) for surfclams or ocean quahogs in one survey year if the species are separated and sampling area reduced?*
5) *What scheduling options (number, location and frequency of surveys for both species) should be considered if surveys for the two species are separated?*
6) *Can rough ground with risk to equipment damage be avoided?*
7) *Should new strata be constructed from current strata or built from scratch using smaller building blocks?*
8) *How heavily should location and depth information vs. survey catch data be weighted in developing new strata?*
9) *Should new strata schemes with discontiguous strata be considered or should strata be defined traditionally as single contiguous areas?*
10) *What are the recommended stratification options (method, location, shape and number of strata) for each species and area?*
11) *Will the recommended changes affect observation and estimation of biological characteristics, such as shell length-weight relationships and growth rates?*
12) *How would potential changes in the clam survey (e.g. lower survey frequency and increased precision) affect management advice and stock assessment modeling?*
13) *How often should future changes in stratification be considered?*
14) *What types of additional research would benefit the clam survey?*

NEFSC staff and the WG addressed each of these questions using methods described below.

2. Materials and methods

Two sets of data decisions are described before analytical approaches are described below. The first set was implemented in the preliminary survey data used by the WG to make recommendations at its final meeting. The remaining decisions were made during or just after the WG's final meeting (see "Data decisions at/after last WG meeting" and "Survey catch density calculations" for details). The latter consist mainly of adjustments to current stratum or regional boundaries and do not materially affect analytical results or recommendations. Where relevant, the data used (preliminary vs. final) are specified below or in the Results section.

DMVSVA

Clam densities have been very low in the southern portion of SVA since the 1980s due to warming water and sampling has been sporadic there (Figure 1.1). Therefore, strata 1-4 and 80 in SVA were omitted while stratum 5 was combined with DMV to form the new DMVSVA region (Figures 2.1-2.2).

Depth range

NEFSC clam survey data used in these analyses were collected at depths of 9-80 m, which is the depth range for new recommended survey areas. The innermost stratum boundary in the current survey at 9 m is near the boundary between state and federal waters.

The outermost strata in the current strata set covers 73-110 m but has never been fully surveyed. The maximum depth sampled during 1997-2011 was 104 m, the maximum depth sampled by the commercial vessel during 2012-2016 was 75 m and the maximum depth practical is currently about 80 m. Fishable concentrations may extend out to about 80 (Dave Wallace, pers. comm.) but quahog catches are low in survey tows at 80 m or deeper. We therefore extended the current strata at 55-73 m out to 80 m and eliminated portions of current strata in deeper water from the survey area (Figures 2.1-2.2). For example, the stratum building block identified as “304” is the union of current stratum 3 and the shallow portion of current stratum 4, “708” is used for the union of strata 7 and 8, “1516” and so on. Differences in sample density were ignored when deep water strata were combined.

Spatial scale

It is important to define portions of the survey area corresponding to the smallest spatial scale of particular interest (the spatial scale at which it is most important to track abundance) because stratification and other design options are evaluated at this level. The GBK and southern areas were selected because current assessment models allow for separate population dynamics in the two areas and because the availability of survey data over time is different for GBK and south.

Stock assessments were done historically at smaller regional levels in the south (SNE, LI, NJ, DMV and SVA, Figure 1.1) and some analyses at regional or smaller spatial scales are still carried out. Survey designs meant to track regional trends in the south were examined but did not perform well for the southern area as a whole because stations that could be allocated to precisely track abundance in the south as a whole had to be shifted to other strata to adequately track abundance in each region. In other words, the more important goal of tracking the southern assessment area as a whole conflicted with the goal of tracking individual southern regions. Effects of recommendations on assessment area and regional survey abundance time series were addressed by post-stratification and domain analyses described below.

Survey years used

The survey data in most analyses were collected during 1997-2016 because sensor data were available to calculate tow distance and because recent data are most relevant to current and future conditions if water temperatures increase as expected. Most of GBK was covered in the new survey during 2013 with the remainder sampled during 2014. Data from both years was combined to form a complete survey for 2013. Similarly, most of SNE was covered during 2013 and 2015 with the rest covered during 2014 and 2016. The SNE data for 2013 and 2014 were combined to form a complete set for 2013. Data for 2015 and 2016 were combined to form a complete set for 2015. Years with complete coverage for GBK were 1986, 1992, 1994, 1997, 1999, 2008, 2011, 2013 and 2016. Years with complete or nearly complete coverage in the south were 1982, 1983, 1986, 1989, 1992, 1994, 1997, 1999, 2002, 2005, 2008, 2011, 2012 and 2015.

Water temperatures increased in the south and surfclam depth distributions shifted abruptly into deeper water off DMV during the 1990's. Therefore, catch density and station data collected during 1997 and 1999 were omitted in analyses involving only DMVSVA.

Survey data collected before during 1992-1994 were used for "out of sample" (external cross-validation, Picard and Cook 1984) testing as described below. The years 1992-1994 provide a reasonable sample size (two complete surveys) collected under conditions as similar as possible to 1997-2016. Years with incomplete sampling for an area were omitted.

Building blocks for new strata

Fifteen-minute squares (FMSQ) and current strata (both omitting portions outside 8-80 m and with other modifications described above) were used as "building blocks" in developing new stratification schemes. The decision to use FMSQ was a pragmatic compromise balancing biological and statistical considerations. Larger twenty-minute squares were not used because clam densities and ecological characteristics seem to change substantially over distances as large as 20 nm. Ten-minute squares were not used because numbers of tows during 1997-2016 in these smaller areas were usually too small to characterize mean catch with sufficient accuracy. Many FMSQ and some current strata also had low sample sizes, but to a lesser extent. Lack of precision due to low sample size was an important consideration in choosing and grouping building blocks to form new strata, choice of methods and in interpreting results.

Regional and area boundaries were modified when using FMSQ because the latter do not fall evenly on current boundary lines. FMSQ entirely within an original region or area were assigned to the original region or area. A square falling on one of original regional boundaries was assigned based on the amount of area on either side of the original boundary. For example, if a FMSQ lay across the boundary between NJ and LI with 55% of its total area on the NJ side, then the FMSQ would be assigned to the NJ region.

Data decisions at/after last WG meeting

Stratum 73 "floater"

The original NEFSC shellfish strata includes a small portion of stratum 73 separated from the rest of stratum 73 and surrounded by strata 72 and 74 (Figure 1.1). For simplicity, and to use contiguous strata building blocks consistently, the floater was assigned to stratum 74 (Figures 2.1-2.2).

Inconsistent strata designations

In a small number of cases (1.7%), the survey stratum identifier recorded on station records did not match the stratum assigned based on the GIS shapefile for shellfish strata currently used by NEFSC. The discrepancies were caused by small differences in the location of stratum boundaries used historically and in the shapefile to define strata or errors in the recorded stratum. The original stratum designations were replaced by their GIS designation. A small number of tows were omitted entirely because they were outside the survey area based on the GIS designation. This correction had nil impact on results and recommendations but simplified data presentation (Tables 2.1-2.5). Adjustments to survey database software should be considered prior to the next clam assessment.

GBK/SNE and LI/NJ regional boundaries

The boundary lines for regions and areas lie along stratum boundaries. Based on original NEFSC shellfish strata, the GBK/SNE boundary falls across the southern part of GBK instead of through the Great South Channel as might be expected based on geographic and oceanographic patterns (Figure 1.1). Similarly the current LI/NJ regional boundary is slightly north of Hudson Canyon even though Hudson Canyon is a major feature that strongly effects clam distributions. The original shellfish stratum lines were adjusted so that the GBK/SNE boundary ran through the Great South Channel and the LI/NJ boundary ran through the Hudson Canyon (Figure 2.1-2.2). Strata were renumbered where necessary to reflect these changes and to ensure that stratum ID numbers for the GBK and SNE areas were unique. Current stratum 47, in particular, occurs on both sides of the original GBK/SNE boundaries and was renamed such that 471 is in SNE and 472 is on GBK. Any differences in sample density were ignored in redrawing region and area boundary lines.

Division of large southern strata by depth for ocean quahogs

Current strata 5, 9, 13, 17 cover the 27-46 m depth range in the southern SVA, DMV and NJ regions (Figure 1.1). Based on 1997-2017 survey data, surfclams tend to occur across the entire depth range in these strata but there is a clear tendency for ocean quahogs to occur in the east at depths > 32. We therefore divided strata 5, 9, 13 and 17 at about 32 m in such a way that all or nearly all of the ocean quahog catches occurred east of the dividing lines (Figures 2.1-2.2). Strata 21 might have been divided as well but there were small quahog catches near the relatively shallow western border. The decision to divide strata 5, 9, 13 and 17 may increase the precision of quahog surveys because no time will be spent surveying these relatively large areas without any ocean quahog catch and because the within-building block variance was reduced.

Identifiers for surfclams and ocean quahog building blocks

The building blocks for surfclams and ocean quahogs were different after strata 5, 9, 13 and 17 were divided for ocean quahogs (Figures 2.1-2.2). We used the prefix 7 for surfclam and 8 for ocean quahogs to distinguish the two sets. For example, current stratum 39 (coded 6039 in the current survey database) becomes building block 739 for surfclams and 839 for ocean quahogs. Current stratum 9 for surfclams becomes 709. The combined 5960 building block in deep water on GBK becomes 75960 for surfclams and 85960 for ocean quahogs. A special convention was used for the quahogs and current strata 5, 9, 13 and 17. For example, the shallow portion of stratum 5 for ocean quahogs becomes building block 89905 while the deep portion becomes building block 805. These conventions had no effect on results or recommendations but the information is useful in interpreting Tables 2.1-2.5 and as general documentation.

Nantucket shoals

The Nantucket shoals fishing grounds in federal waters south of Cape Cod currently supplies a large fraction of the total surfclam catch. The area is not routinely surveyed because of dangerous currents and shallow water. However, it was sampled during 2017 by the Science Center for Marine Fisheries (SCeMFiS) using a small commercial vessel (Powell et al. 2017). The polygon used in designing the 2017 survey may serve as a new stratum if the area is surveyed again (Figure 2.3).

Survey catch density calculations

Survey data in all analyses were catch numbers rather than catch weight per unit swept area. Catch numbers and weight were highly correlated for surfclams and ocean quahogs and the same results, decisions and recommendation would result using either.

Survey catches were expressed as densities (N m^{-2}) to account for variation in tow distance, dredge width and capture efficiency that occurred over time and with replacement of the original survey equipment. Survey catch densities (D_t) were computed:

$$D_t = \frac{N_t}{w d_t r}$$

where N_t is the total catch of clams 40+ mm shell length (SL), w is the width of the dredge (which changed in 2012), d_t is the tow distance for each tow based on sensors and r is relative capture efficiency (probability of capture for a clam above, below or in front of the path the dredge). The parameters r= 0.59/0.23=2.57 for surfclams and 0.608/0.194= 3.13 for ocean quahogs are ratios of capture efficiency estimates and account for fishing power differences in the original and current surveys. Differences in size selectivity for gear used during the two time periods were ignored but the calculations were restricted to clams 40+ mm shell length (SL).

Survey data for 1997-2016 were pooled in bootstrap and other analyses to increase sample size because spatial patterns were important while interannual variability was not. Interannual differences are clear in the original data and the variance of the pooled data might greatly exceed the variance among tows in a single survey. Catch data for each year were therefore rescaled to obtain a more realistic level of between tow variance in the pooled data set:

$$d_t = \frac{D_t}{\overline{D}}$$

where D_t was the catch from tow t and $\overline{D}$ was the simple (not stratified) average density in the same spatial domain and year as the tow. Preliminary data analyzed before the final WG meeting were scaled before applying the 1% rule. Final analyses used data standardized correctly after applying the 1% rule. These changes affected results but the overall effect on conclusions and recommendations was nil.

1% rule

The 1% rule was used to reduce the survey area by omitting building blocks in poor habitat areas. A 5% rule was also considered but it seemed to exclude too much area. The 1% rule guaranties that any building block (FMSQ or current strata) with appreciable densities or abundance are sampled and limits bias in overall swept-area biomass estimates due to reductions in survey area to less than 1%.

The first step was to calculate the mean density and swept-area abundance (density x area) in each building block for the region or area of interest. The second step was to sort the building blocks by mean density, calculate the cumulative sum of the sorted densities and then divide by the sum of the densities to compute cumulative proportions. The third step was to repeat these calculations using swept-area abundance. Building blocks with cumulative proportion ≥ 1% for *either* density or abundance were retained. The swept-area criterion is nearly identical to the density criterion for FMSQ because they are of similar size. The swept-area criterion is more important when using current strata as building blocks because a large low density stratum with high abundance could be excluded using the density criterion alone. The reductions in survey area were smaller using current strata building blocks.

Optimal allocation

Random stations were assigned in bootstrap analyses to potential new strata based on a compromise version of Neyman optimal allocation. Given a specified total number of random stations (e.g. determined by budget), the Neyman procedure estimates the proportion and number of random stations in each stratum that would minimize the variance of the stratified mean. The calculations are based on stratum area and within-stratum variances (both assumed known precisely). The compromise ensures that at least two tows (the minimum number for calculating variance) are assigned to each stratum (stations are removed from strata with the highest allocation so that the total number is unchanged). In practice, optimal allocation calculations are degraded by uncertainty in the stratum variance, particularly if the number of strata is large and number of tows in each stratum is small. Neyman allocation does not include any adjustments for factors other than stratum size and variance that may be important in the survey (e.g. interest in a small area of heavy recruitment).

Numbers of stations

Assumptions about the total number of stations in each area were based on survey performance during 2012-2016. During 2012-2016, both GBK and the south were covered by the clam survey twice (two “complete” surveys) with a total of 270 random stations on GBK and 425 random stations in the south. Thus 270/2=135 random stations per year on GBK and 425/2=212 random stations per year in the south appear feasible. In contrast, the maximum number of random stations per year was 159 for GBK and 186 for the south. Based on these figures, bootstrap and other analyses assumed 150 (about (135+159)/2) random stations on GBK and 200 (about (212+186)/2) random stations in the south. These assumptions had little effect on results and recommendations but sample size is important in estimating the expected variance in either relative or absolute terms (Appendix 2).

The assumed numbers of stations may be optimistic, particularly if funding for clam surveys is reduced. On the other hand the number of station could be higher if future surveys target one species at a time and cover a smaller area. Surveys during 2012-2016 included extra days for weather and special field experiments which could also be used for normal survey work but are not included in calculations above. We estimated precision and accuracy for surveys based on a range of sample sizes to account for potential changes in funding (Appendix 2).

Design effects

Design effect (DEF) statistics were used to evaluate performance of survey designs. DEF measures the success of a design in reducing the variance of the stratified random mean relative to the variance of a random survey with the same number of tows. The DEF statistic is $100 \times (1-\sigma_S^2/\sigma_R^2)$ where σ_S^2 is the variance of the stratified random mean under the design being tested and σ_R^2 is the variance of a random survey with no stratification. DEF can be decomposed into parts due to stratification and allocation so that $DEF = DEF_{Allocation} + DEF_{Stratification}$. It is important to remember that DEF statistics are estimates that include within-stratum variance estimates. They tend to be highly variable because variances tend to be highly variable. DEF statistics were calculated using the BIOSurvey2 library in R provided by Dr. Stephen Smith (Canada Department of Fisheries and Oceans, emeritus and WG member).

Historical comparisons

Effects of separating surveys for the two species were evaluated by comparing CVs for stratified random means during 1986-2016 (GBK) or 1982-2016 (south) using current strata, raw survey data and BIOSurvey2. The 1% rule was not used but optimal allocation in these tests used a minimum value of zero stations so that marginal strata could be eliminated entirely. In effect, the test measured effects of both optimal allocation and not sampling marginal habitat. As with other methods used, these calculations understate uncertainty because they assume variance estimates for each stratum in each survey are known.

Cluster analysis

Univariate and multivariate cluster analyses were used to group building blocks (FMSQs or current strata) into 2-10 clusters that could potentially be used as new strata. The two types of analyses represent different assumptions about accuracy of the mean survey density estimates for building blocks. The univariate method assumes that mean catch estimates for building blocks are accurate while the multivariate analysis accommodates imprecise mean catch estimates by using location and depth data also. Location and depth can be measured accurately although their relationship to spatial patterns in abundance may be inconsistent and likely to change over time.

The univariate procedure tends to form spatially discontiguous clusters (e.g. clusters consisting of building blocks that are scattered across the survey area). In contrast, multivariate procedure tends to group building blocks at adjacent locations and similar depths (e.g. adjacent FMSQ). The univariate and multivariate analyses are the same when the weights for location and depth in the multivariate case are set to zero.

The univariate procedure was carried out using the Ckmeans.1d.dp library in R with the number of samples in each building block as weights (Wang and Song 2011). It is guaranteed to form clusters that minimize within cluster sums of squares:

$$SSQ = \sum_{c=1}^{N} \sum_{j=1}^{n_c} n_{c,j} \left(\bar{d}_{c,j} - \bar{\bar{d}}_c \right)^2$$

where N is the number of clusters specified by the user, n_c is the number of building blocks in cluster c, $n_{c,j}$ is the number of tows for the j^{th} building block, $\bar{d}_{c,j}$ the total mean catch, and $\bar{\bar{d}}_c = \left(\sum_j \bar{d}_{scj}\right)/n_c$ is the mean catch in the cluster. This formula is similar to the formula for the variance of a stratified random mean so that minimizing it nearly minimizes the variance of the stratified random mean. There is nothing in the formula to reward or penalize proximity in terms of space or depth so that building blocks grouped into the same cluster can be discontiguous and scattered throughout the region.

Multivariate cluster analyses grouped building blocks based on expected catches, position (latitude and longitude) and depth using the kmeans routine in the R stats library (Hartigan and Wong 1979). Multivariate cluster analysis minimized the sum-of-squares (*SSQ*):

$$SSQ = \sum_{c=1}^{N} \sum_{j=1}^{n_{c,j}} \left[{d'_{s.j}}^2 + {lat'_{s.j}}^2 + {lon'_{s.j}}^2 + {depth'_{s.j}}^2 \right]$$

where *lat* and *lon* are positions at the center of a building block and depth is an average calculated over a fine scale grid in each building block. Data on the right hand side of this expression were weighted z-scores. For example,

$$d' = w_d \frac{\left(d - \bar{d}\right)}{\sigma_d}$$

where, w_d (default = 1) is a weight that defines the importance of catch density in clustering and the mean $\bar{d}$ and standard deviation σ_d for density were computed based on all tows in the analysis. The data were transformed to z-scores because natural differences in scale (e.g. density < 1 and longitude > 100) would overemphasize the variables with larger values. Mean depth, position and depth are straight-forward for FMSQ but harder to interpret for building blocks based on current strata with complex shapes. The k-means algorithm is a standard approach but not guaranteed to find the best combination of subunits because its starting points and selection process are random. We restarted the algorithm 50 times and kept the best cluster with lowest within-cluster sums of squares.

GAM and Tree models in multivariate cluster analysis

Cluster analysis was also used to group building blocks based on estimated density from GAM and random forest tree models into clusters (Breiman 2001; Wood 2006). Six GAM models were fit to observed survey catch densities based on location and depth (see below) using the log link function and assuming the data were from Tweedie distributions to accommodate stations with zero catch. The linear predictors for each model specified in the R programming language were:

```
~ te(L,k=7) + te(depth,k=7)
~ te(L,k=7) + te(depth,k=7) + as.factor(yr)
~ te(L,k=7) + te(depth,k=7) + te(yr,k=x)
~ te(L,depth,k=14)
~ te(L,depth,k=14) + as.factor(yr)
~ te(L,depth,k=14) + te(yr,k=x)
```

where *L* is location (latitude or longitude), as.factor(*yr*) means different intercept parameters for each year, and *k* limits the curvature of univariate [te(*L*,*k*) te(*depth*,*k*)] and bivariate [te(*L*,*depth*,k)] nonlinear terms. Nonlinear terms for year used *k*=min(7, number of years-1). Models with year were meant to account for temporal variation in preliminary data induced by applying the 1% rule before rescaling. *L* was either latitude or longitude depending on which had the lowest correlation with depth because lack of correlation improves model predictions. The best model was selected based on lowest AIC and used to make predictions at each point on a fine scale grid (0.01° spacing) within the study area.

Tree models incorporating additional data were used in a similar way. Random forest regression models were fit to observed survey catches based on a depth, latitude, longitude and environmental and climatological variables (Table 17 in NEFSC 2017c). The fitted tree models were then used to make predictions at each point on the fine scale grid. Missing values in the predictor data were replaced by an imputation routine in the randomForest library in R.

Multivariate cluster analyses were used to assign the predicted catch densities from GAM or tree models at each point in the grid into 2-10 clusters ignoring building block boundaries (univariate cluster analysis may have performed better). Entire building blocks were assigned to a new stratum based on the most common cluster assigned to the points within it. Eventually, as the number of potential strata increased, one or more strata were lost in the final assignment because the corresponding cluster was not the majority assignment in any of the building blocks. Such cases were abandoned.

Stratification options for bootstrap tests

A wide range of options with 2-10 new strata based on FMSQ and current strata were compared for both species and regions. Options are identified based on whether current strata or FMSQ were used, the number of new strata and method used to form building blocks). For example, option opt2.3 is based on FMSQ (“opt” means FMSQ building blocks), has two new strata and was developed using method type 3 (univariate cluster analysis, see table below). Option svd2.3 is analogous but based on current survey strata building blocks (“svd” means current strata as building blocks). The scenarios tested included the current survey design (after applying the 1% rule to current strata) and random assignments of TNMS or current strata to new strata. The current stratification scheme with optimal allocation (option opt.quo, 1% rule not applied) was also included.

ID number	Method
1-2	Random
3	Univariate cluster analysis
4	Multivariate cluster analysis
5	GAM model
6	Tree model

Bootstrapping to evaluate stratification options

Bootstrap techniques were used to evaluate relative performance of stratification options in several types of analyses. Each bootstrap (usually 3000 iterations) involved a simulated survey from an area or region represented by the survey data for 1997-2016 (or earlier years for cross-validation, see below). Tows in future surveys will probably be allocated to strata optimally based on stratum areas and possibly imprecise variances estimated using previous survey data. To mimic this in bootstrap analysis, variances were estimated during each bootstrap iteration for each new stratum using data from a simulated pre-survey with the total number of tows allocated to new strata based on relative stratum size (both surveys were simulated in each bootstrap iteration). The total number of tows in pre-surveys was pN where N is the total number of tows to be carried out in the real survey and p is a proportion used to reduce the sample size in the pre-survey and subsequent accuracy of the variances and allocation (p=1 in model runs reported here).

It was important in bootstraps to obtain spatially unbiased samples from new strata that were usually comprised of portions of two or more building blocks originally sampled with different intensities. For example, a new stratum might be composed of two building blocks with areas of 10 vs 20 km^2 and with 15 vs. 5 historical tows. Simple random samples from the pooled data gives biased bootstrap results because the first building block would be oversampled due to its smaller size and the large number of tows there. To achieve spatially unbiased bootstrap sampling, we first sampled tows from building blocks within a new stratum in proportion to their area based on a multinomial distribution:

$$\vec{n} \sim Multinomial(N, \vec{p})$$

where N is the total number of tows in the new stratum after optimal allocation, $\vec{p}$ is a vector with the proportions of the total area in each old stratum, and $\vec{n}$ is a random vector holding the number of tows (zero to N-1) in each building block ($\sum \vec{n} = N$). The second step was to randomly sample n_j tows with equal probability from the pooled data for block j. The stratified random mean, its CV, design effects (DEF) and other statistics were calculated and stored at the end of each iteration.

Median DEF estimates from each bootstrap iteration were saved and used as the primary performance measure in evaluating stratification options. Distributions of relative errors $\left(\frac{(\bar{x}-x_{true})}{x_{true}}\right)$ where x_{true} is the "true" value were sometimes used as well. Other statistics including the interquartile range (IQR) for relative errors, median and IQR for $DEF_{Allocation}$ and $DEF_{Stratification}$ and the proportion of DEF statistics > 0 were saved and examined but DEF was convenient, correlated with the other measures and seemed to work well. Bias was not considered because all stratification options gave unbiased results if sampling was spatially unbiased. The truth was calculated by calculating the stratified random mean using all available data and building blocks as strata (thus the true value would vary among bootstrap runs with different sets of building blocks).

Bootstrapping was useful but it is important to point out shortcomings related to sample size and underestimation of variance. In some cases, the number of historical tows in a new stratum or portion of an original stratum was usually low so that the same tows were sampled repeatedly with less variance than would be expected in a real survey. Sample sizes (particularly for FMSQ) were generally low enough that rare large catches were not included. Finally, we used the same data to develop and test stratification options (see exception described below). These shortcomings tend to exaggerate the bootstrap estimates of precision, potential benefits of optimal allocation, and the apparent benefits of increasing numbers of strata. In a real survey, stratum variances are likely to be larger and the number of optimal strata is likely to be smaller than suggested by the bootstrap.

Out of sample bootstrap testing

The performance of a stratification option developed using 1997-2016 data was compared to its performance based on 1992-1994 data. A high correlation between in- and out-of-sample DEF would provide information about how well the option might work in future with new data relative to what might expected base on the original analysis and data. The test was imperfect because stock conditions during 1992-1994 differ from current and future conditions (particularly in the south). Nevertheless, the 1992-1994 data were the only available independent data for testing.

Stability of clusters

It is possible that cluster results were strongly affected by noise in mean catch density estimates. A bootstrap-type procedure was used to investigate the stability of clusters and to choose weights for catch density in multivariate analysis.

Stability analysis used station catch densities and building block summary data for each region and species, which are the information actually used in clustering. They include the number of tows ($n_{s,j}$), latitude, longitude and mean depth for each block. The original catch density data (tow-by-tow station records) were pooled into groups defined by their original cluster assignment. For example, if the existing analysis assumed two clusters, then all of the catch densities for each station were assigned to one stratum or the other based on the original cluster analysis. In each iteration and for each building block, $n_{s,j}$ catch densities were sampled with replacement from the pool for the stratum and the mean of the sample data for each building block was added to a data file with the original latitude, longitude and depth. The cluster analysis was rerun using these data to assign each block to a bootstrap cluster, which was recorded. This approach did not fully replicate sampling in actual surveys because catch densities for building blocks within the same original cluster probably have different means and variances. However, there were not enough data to bootstrap catch densities at the individual block level. This procedure was carried out using the univariate algorithm (effective weight on catch data very large) and the multivariate algorithm with catch weights of 1, 0.5 or 0.25. Weights for location and depth were all set to one.

It is challenging to define and quantify stability in multivariate cluster analysis. To see why, say FMSQ 1&2 were assigned to the cluster 1 with the lowest mean catch density while FMSQ 3&4 were assigned to cluster 2 with the highest mean catch density in the original and first bootstrap analysis. In the second bootstrap analysis, FMSQ 1&2 were assigned to the cluster with highest mean catch density and FMSQ 2&3 were assigned to the cluster with the lowest catch density due to sampling error. The cluster assignments for the four FMSQ were 1,1,2,2 in the original and first bootstrap case and 2,2,1,1 in the second bootstrap case. These assignments are different but functionally identical because the first and second FMSQ were always grouped together as were the third and fourth FMSQ. These issues were surmounted by visual examination of original and bootstrap strata maps (all building blocks assigned to a new stratum plotted in the same color) and using a root mean square error statistic.

The root mean sum of squares statistic (RMSE) was:

$$RMSE = \sqrt{\frac{\sum_{b=1}^{N}\sum_{i=1}^{n_R}\sum_{j=1}^{n_R}\left(\boldsymbol{O}_{i,j} - \boldsymbol{P}_{i,j}\right)^2}{2Nn_R}}$$

where N=1000 is the number of bootstrap iterations, n_R is the number of building blocks in the area based on the 1% rule, and $\boldsymbol{O}$ and $\boldsymbol{P}_\mathbf{b}$ are square (n_R x n_R) symmetrical matrices that record the number of times the r^{th} and c^{th} building block occurred in the same cluster based on either the original or b^{th} bootstrap cluster analysis. Cases where a cluster consisted of a single building block are recorded along the diagonals where r=c. If the cluster analysis were perfectly stable, then $\boldsymbol{O}$ and $\mathbf{P}$ would always be the same so that RMSE=0. The statistic increases as stability is reduced. The statistic is divided by 2 to account approximately for the symmetry of the matrices, N to account for the number of bootstrap iterations and n_R to account for the number of building blocks. Thus, stability is measured as the root mean square number of times that building blocks in the same original cluster were not assigned to the same cluster, per bootstrap iteration and building block. If an error is defined as failure to assign pairs in the original cluster together, than the statistic is the root mean squared error.

The RMSE statistic measures the number of mistakes, not their magnitude or severity. Two stratification schemes in the same region might have the same RMSE statistic even if one scheme involved assignments among adjacent building blocks while the other involved assignments among distant building blocks (the statistic measures numbers but not severity of errors).

The RMSQ statistic is useful for comparing stability of stratification options from cluster analyses with different catch weights within a region, the same species and for the same number of new strata. RMSQ results for different numbers of new strata or different areas and species are not comparable.

Long-term trends using historical data in new strata

It was important to evaluate the precision and trends of post-stratified means using new stratification options and historical data. New strata are likely to be fewer, larger, have different boundaries and cover a smaller total area than the current design. Therefore, current strata that intersect the new survey area will fall entirely or partially within one or more new strata. The conventional formula for stratified means using overlapping sections of the current and new strata is unbiased for post-stratification in this case. However, the variance formula for stratified means is biased (variances too small) if sampling rates varied among original (current) strata in the same domain (new strata, Sarndal et al. 1992). Variances tend to be underestimated because the number of observations in each overlapping section is itself a random variable with variance that is not included in the conventional formula.

We used the Domain.est and post.stratify functions in BIOSurvey2 to calculate post-stratified means and domain or post-stratified variances for surfclams and ocean quahogs in each area and region. Results were compared to stratified random mean catch densities calculated using the clam survey database used for stock assessment work with current strata. Survey data used in assessments include adjustments not made in the data used elsewhere. Therefore, we used tow data actually employed in the most recent assessments.

Effects on stock assessments and management advice

These analyses focus on ocean quahogs because recommended changes will certainly improve survey data and management advice for surfclam (see Discussion). Surfclam data and advice would improve for surfclams because survey frequency would not change, sample density on surfclam habitat would increase and optimal allocation would be used. The 1% rule will result in negligible negative bias (≤ 1%) in swept-area stock size estimates that can be ignored. Mean survey density will increase but with little or no effect on trends in relative abundance indices unless a substantial portion of the surfclam stock shifts into deep water quahog habitat and is not sampled (see Discussion). Prior distributions for catchability in stock assessment models for surfclams are based on capture efficiency estimates and will be unaffected.

The survey frequency for quahogs might change from one survey every three years to one survey every six years (see below). Thus, the net effect on survey data and management advice will depend on the tradeoff between gains in precision and losses due to reduced survey frequency. This problem was addressed by comparing the relative accuracy of estimated trends in stock size in a simple simulation analysis.

Given the nature of their population dynamics, it is reasonable to approximate ocean quahog stock dynamics with a linear population dynamics model and to approximate the stock assessment model using linear regression. Any changes in the ocean population are likely to be small from year to year. Ocean quahogs grow slowly and experience very low total mortality. Large increases in year class strength may be possible but have probably not occurred during the last century based on preliminary age data (Roger Mann, Virginia Institute of Marine Sciences, Gloucester Point, Virginia). Any strong year classes that may occur will recruit to the fishable stock slowly over decades so that their effect on stock biomass will be slow and muted. Ocean quahog show essentially linear abundance trends over the last 40 years based on a variety of actual stock assessments (Figure 2.4).

Simulated stock size was $\mu_{t+1} = \mu_t + \boldsymbol{\delta}\, \mu_1$, where the annual increment ($\boldsymbol{\delta}\, \mu_1$) was constant, δ was a parameter and *t* was year. Simulated abundance data included lognormal survey errors $\mu_t' \sim lognormal(\mu_t, \sigma_s)$ where μ_t' is a stock size estimate and σ_s was the standard deviation of stratified means from the survey design being considered (σ_c for the current design and σ_p for the proposed design). The standard deviations $\sigma_s = \sqrt{1 + CV_s^2}$ were based on the *CV* calculated from errors in stratified mean catch density during bootstrap analyses using the recommended (new strata and optimal allocation) or current (current strata and mean historical allocation) survey designs with 150 stations on GBK and 200 in the south (Table 2.6).

The simulated population was “surveyed” by drawing a random μ_t' value every three years under the previous survey schedule or every six years under the recommended survey schedule. After a maximum 50 years of surveys (16 total with one every three years for surfclams or 10 total with one every six years for ocean quahogs), a linear regression assessment model was fit to the time series of simulated survey data (without log transforming the data). Estimated trends from the linear models (slope parameters, $\hat{\delta}$) were saved (10,000 iterations) and compared to the true population δ value to compute a relative error $RE = (\hat{\delta} - \delta)/\delta$ for the estimated rate of population change.

The performance of the survey and assessment model for a scenario was calculated as the standard deviation of the distribution of relative errors in each year. This statistic is a better indicator of survey and assessment model performance than measures of statistical significance because it directly reflects the distance of the estimate from the truth. A lower standard deviation indicates higher precision and accuracy in estimating the true trend in abundance, which is an important objective of stock assessment. Bias was ignored because both survey designs were unbiased and the mean of the relative errors was always zero.

Sensitivity analyses were used to explore the sensitivity of results to assumptions about the ocean quahog population and survey/assessment assumptions. Sensitivity to the length of the survey time series was tested by varying the length of the simulated survey from 10 to 50 years. Sensitivity to the interval of time between surveys under the proposed design was tested by varying the survey interval from 1 to 12 years, while holding the interval for the current design constant at 3 years. Sensitivity to the variance of estimated abundance in the proposed survey was tested by varying the CV for simulated survey estimates between 0.01 and 1.01 in the proposed survey design, while holding the CV of abundance in the original design constant at the original value. Finally, sensitivity to the direction and magnitude of the true abundance trend was tested by varying δ between -0.05 and 0.05.

Untowable ground

Untowable ground that might damage the survey dredge was identified using clam survey station records that include location (coordinates at the start of the tow) and information about substrate and gear damage. We focused on Georges Bank where gear damage is most common. Three methods were used.

The first method used substrate data from survey catches during 1980-2011. Data from early surveys were from original paper logs and data from later surveys were from survey database records. During surveys, data are collected for substrate categories that include "rock" and "boulder". A score (0, 1 or 2) is assigned to each category based on the watch chief's estimate of volume. The code 0 meant that no material in the category was observed in the dredge, 1 meant up to 29 bushels and 2 meant 30 or more bushels. Stations less than one tow distance apart were both assigned the higher score. We designated the location of a tow as untowable ground if the score for "rocks" or "boulders" was a 2.

The second method was nearest neighbor analysis using data collected on Georges Bank during 1982-2011. The actual process was more complex but, in effect, each tow was categorized as "towable" or "untowable". The four nearest neighbors in the northeast, southeast, southwest, and northwest quadrants around each "parent" station were identified using a using a modified bishop's move algorithm. Tows near boundaries were often missing at least one of the four neighbors. These tows were excluded from further analysis as parents but retained as neighbors. Parent sites with three out or four neighbors designated "untowable" were deemed to be on untowable bottom.

The last method was based on watch chief comments during 1983-2016 and the SHG (station type-haul type-gear condition) value assigned after each tow. If SHG was greater than 144 (station type random; haul either unrepresentative or aborted due to bad bottom; gear moderately damaged including blade assembly damage, bent bars, knife blade damage and multiple broken nipples) and the damage was determined to be a result of contact with rough bottom by the comments, then the tow location was designated untowable. Tows in which the watch chief's comments referenced dredge damage due to substrate, large rocks in the tow or abandonment of the station due to bad ground the tow location were also designated untowable.

The WG decided to exclude circular areas one nautical mile in diameter centered on the coordinates of each problematic station from the survey area. A one nautical mile is (1,852 m) about twice the maximum distance of tows with the old survey dredge (875 m, Weinberg et al. 2002). The area excluded around each station amounts to about 0.785 nm^2 or 2.7 km^2.

3. Results

Surveys areas identified using the 1% rule were smaller than the current survey area by 26-44% for surfclams and 35%-42% for ocean quahogs (Table 3.1 and Figures 3.1-3.8).

Based on historical analyses, CVs for stratified mean densities in the current survey design were usually similar to what might be expected under random sampling (Figure 3.9). In contrast, CVs with optimum allocation were always about 50% smaller. DEF calculations for historical surveys (not shown) confirmed these results. $DEF_{Stratification}$ was usually positive but $DEF_{Allocation}$ and DEF were usually small or negative. The benefits of the current survey design were reduced or eliminated by the compromise allocation approach used historically.

Based on bootstrap analyses and DEF with preliminary data, strata options based on FMSQ from univariate cluster performed best for both species in both areas (Figures 3.10-3.13). DEF generally increased with the number of strata. The current survey design (with optimal allocation) performed relatively well based on bootstrap results, probably because the numbers of original strata (19 for GBK and 53 in the south) were relatively large. As expected, random assignments of building blocks to new strata performed poorly with median DEF near zero (results not shown). Performance of the current survey design varied.

RMSQ scores show that clusters based on FMSQ were relatively stable with catch weights of 0.25 on 0.5 and that univariate clusters, particularly with FMSQ, were unstable (Table 3.2). The range of RMSQ scores for each species and area were wide using FMSQ as building block indicating that weights had substantial effects on stability. Results for current strata building blocks were mixed with relatively narrow ranges indicating that weights were less important. Univariate clusters with current strata were relatively stable for GBK surfclams and unstable for GBK ocean quahogs. Strata maps (not shown) confirmed these patterns.

DEF scores from in- and out-of-sample bootstraps with FMSQ were correlated in two out of four cases (GBK surfclams and GBK ocean quahogs, Figure 3.14). DEF scores with current strata building blocks were correlated in three out of four cases suggesting that current strata results may be more stable (GBK surfclams, GBK ocean quahogs, and southern quahogs, Figure 3.15).

The out-of-sample and stability results with FMSQ and univariate cluster analysis may seem contradictory because performance of FMSQ options was good in some cases based on in- and out-of-sample bootstrap tests, but always poor in stability tests. The seeming contraction is due to FMSQ with persistent high densities. The univariate procedure tends to consistently (and correctly) cluster the persistent high density FMSQ together in a way that improves precision. Most of the instability, however, was among FMSQ where densities were lower and less certain due to limited data or variability over time. In the lower density cases, random differences in mean density due to sampling resulted in substantially different mean catch densities and cluster assignments. Building blocks based on current strata were more stable because they were larger, involved fewer potential combinations, were based on more data, and had relatively certain average catch densities with larger sample sizes in bootstrap analyses.

Effects on stock assessments and management advice

Based on simulations, the proposed survey is expected to better detect a trend in ocean quahog abundance than the current survey under likely future conditions (Figures 3.16-3.19). Considering sensitivity results, the standard deviation for relative errors in trend estimates in the proposed survey was less than in the current survey in most cases. The proposed survey was better when the trend in true population was gradual ($-0.02 < \delta < 0.02$). Both surveys performed equally well when the trend in true population was steeper ($-0.5 < \delta < 0.5$). The proposed survey was better over short (10 years) to long (50 years) time periods, when the interval between survey years was less than 9 years, and when the proposed survey was more precise than the current survey.

Untowable ground

Method 1 identified 93 problematic locations, method 2 identified 58 locations and method 3 identified 163. There were 314 in problematic locations in of which 208 were unique (Table 3.3 and Figure 3.20). The total area excluded from the GBK survey area would be 561 km^2 (163 nm^2). Most of the excluded area are on the northern half of GBK where surfclams are common.

4. Discussion

Based on WG decisions during and after its last meeting, multivariate cluster analyses (catch weights of 0.5 and current strata building blocks) and bootstraps were rerun using finalized data. These results together with results based on preliminary data were used to make final recommendations described below.

Key questions and recommendations

1) Should clam surveys target surfclams and ocean quahogs separately rather than simultaneously?
2) Should sampling of poor habitat areas cease, particularly if the two species are surveyed separately?
3) Should new species-specific stratification schemes be considered if the two species are surveyed separately?

Recommendation: Surfclams and ocean quahogs should be surveyed separately, poor habitat areas should be omitted based on the 1% rule with current strata building blocks, and species-specific stratification schemes should be used to improve precision of data used in stock assessments.

Rationale: Historical analyses show that survey precision would improve substantially if stations are allocated optimally for one species at a time with more stations per area of habitat for the target species habitat. Nearly all positive stations for both species were included in new recommended surveys areas that were 26%-44% smaller based on the 1% rule (Table 3.1 and Figures 3.1-3.8). The distribution of surfclams will probably continue moving into deeper water but there is room for expansion because the recommended survey area for surfclams extends out to 55 m off NJ and DMVSVA where densities are still low. If surfclams do expand to habitats deeper than 55 m, then they will be caught in surveys targeting ocean quahogs and the survey can be adjusted. Dropping poor habitat areas will generate a negligible negative bias (<1%) but future swept-area biomass estimates will be more accurate because improved precision will offset the bias. Ocean quahogs occur but at very low densities along the 80 m limit of their recommended survey area which is the limit of sampling using the current survey gear. Quahogs recruit and grow slowly and are not likely to establish large dense beds beyond 80 m over the next two decades.

4) Is it feasible to survey the entire stock (GBK plus south) for surfclams or ocean quahogs in one survey year if the species are separated and sampling area reduced?
5) What scheduling options (number, location and frequency of surveys for both species) should be considered if surveys for the two species are separated?

Recommendation: It is not feasible to survey the entire stock of either species given available resources according to Survey Branch personnel, particularly considering surveys days lost to weather and equipment problems. Six scheduling options were developed Table 4.1). The recommended alternative (Option 5) maintains the current survey frequency for surfclams (two complete surveys every six years) and reduces survey frequency for ocean quahogs from two surveys to one complete survey every six years.

Rationale: Both species are at high biomass levels and overall fishing mortality rates are low due to market limitations. Surfclams older than 30 y and ocean quahogs 200 y are common so there are no demographic problems likely to result in rapid stock changes. Simulation results suggest that assessment accuracy depends on having precise survey data and reliable information about survey capture efficiency, rather than more frequent but less precise survey data points. It is likely that rapid surfclam declines would be detected at recommended survey frequencies or using mandatory logbook data which are spatially detailed (reported by trip and ten-minute square) and evaluated annually during routine management. Survey frequency could be adjusted if rapid declines, increased fishing mortality or any other problems occur.

6) Can rough ground with risk to equipment damage be avoided?

Recommendation: The areas identified in this analysis which should be incorporated into station selection algorithms and omitted from the survey area. Strong protocols for dealing with new rough ground during future surveys should be maintained. The database of untowable grounds should be expanded based on new data from the clam survey, other surveys and any other data or method available. Untowable ground in the southern region should be identified.

Rational: Every effort should be made to avoid grounds where gear damage may occur. The current survey uses a large and expensive commercial dredge mounted on a steep ramp at the end of the vessel and it is not feasible to make major repairs at sea. Any significant damage to gear at sea will require a return to port, substantial loss of sampling opportunities and reduced data precision. The database describing untowable grounds should be useful in surveys for species and projects.

7) Should new strata be constructed from current strata or built from scratch using smaller building blocks?

8) How heavily should location and depth information vs. survey catch data be weighted in developing new strata?
9) Should new strata schemes with discontiguous strata be considered or should strata be defined traditionally as single contiguous areas?
10) What are the recommended stratification options (method, location, shape and number of strata) for each species and area?

Recommendations: Based on analyses with final data and all of the WG decisions, the options with six strata identified by multivariate cluster analyses using current strata building blocks and catch weights of 0.5 are recommended for both species on GBK and for surfclams in the south (Table 4.2 and Figures 4.1-4.3). The option with six strata could also be recommended based on analytical considerations for ocean quahogs in the south (Figure 4.4). However, one potential stratum for southern ocean quahogs consists of two discontiguous groups of building blocks (838 + 83940 in the west and 846 + 84748 in the east. The final recommendation for ocean quahogs in the south separates the discontiguous parts for a total of seven new strata to reduce the potential impact of operational problems that might reduce the total number of tows and leave the original large stratum partially sampled. These recommendations can be adjusted by Survey Branch to accommodate logistical or other issues not considered here. Modest deviations from these recommendations would not require reanalysis.

Rationale

Based on median DEF statistics from bootstrap analyses, these recommendations will reduce the variance of stratified random means by 10% for surfclams in the southern area relative to the variance from a simple random design and 25%-35% otherwise (Table 4.2). In contrast, historical calculations indicate that DEF for the current design is near zero and often negative for both species and areas, although DEF estimates from different types of calculations may be difficult to compare.

The average number of tows allocated to each stratum during bootstrap analysis ranged 2-78 and 6-94 for surfclams in the GBK and in the southern areas (Table 4.3). The stratum with two tows for GBK surfclams could be eliminated entirely if needed. The average number of tows allocated to each stratum ranged 6-56 and 10-55 for ocean quahogs in the GBK and southern areas.

Rationale: Judgement is required in making survey design recommendations because uncertainties and limited data make it impossible to identify the best options with certainty. Decisions about contiguity of new strata, building blocks, cluster analysis method, use of survey catch data and numbers of strata were interrelated and all were affected by limitations in the available survey data. Stability analysis with preliminary data showed that multivariate clusters with catch weights of 0.5 and current strata building blocks were relatively stable. Corresponding bootstrap analysis with preliminary data indicated stratification schemes based on these decisions performed reasonably well based on DEF once univariate results were discounted due to stability. Options based on tree and GAM models did not perform particularly well. Out-of-sample bootstrap tests showed that recommended sets worked as expected with "new" data not used to develop the options.

Recommending numbers of strata was a challenging problem. More than five strata are seldom required in survey design (Cochran 1977) and the number of strata should be correlated with the amount of information available and inherent variability. The recommended number (6-7 strata) for surfclams and quahogs are larger but more strata may be better given the constraints on stratum design (contiguity and old stratum building blocks) and given the size and complexity of the clam populations. There was relatively little improvement in DEF or increases in relative errors with more than six new strata based on bootstrap analysis with final data (Figures 4.5-4.8).

All analytical procedures in this analysis exaggerate the benefits of increasing numbers of strata, because the assumed or implicit variance within building blocks and strata were probably too small. Precision (DEF) in survey design depends on trading off benefits from increasing the number of strata (achieving high $DEF_{Stratification}$) and knowing stratum variances well enough to allocate stations to strata accurately (maintaining high $DEF_{Allocation}$). Consider a hypothetically ideal situation where it is easy to develop “good” strata that reflect the spatial distribution of either clam species and reduce within stratum variances. If the variances were known and sample size unlimited, precision could always be improved by increasing the number of strata because the variance in catch density between strata could always be increased. However, stratum variance estimates degrade in real surveys as the number of strata increase because the sample size in each stratum decreases and allocation becomes less accurate. At some point, the gains in $DEF_{Stratification}$ as strata number increases are offset by loss of $DEF_{Allocation}$. The best choice for number of strata is where the gain in $DEF_{Stratification}$ and loss in $DEF_{Allocation}$ cancel and precision is actually lost if more strata are added. Analytical methods used here understate variances and the losses due to inaccurate allocation so that the point where $DEF_{Stratification}$ and *DEFAllocation* cancel appears shifted in the direction of more strata. The situation is even more complex for clam surveys because the relatively large building blocks based on current strata may make it difficult to match patterns in spatial distributions with stratum boundaries in some cases. Given the misleading nature of such analysis, analysts can choose the number of strata such that the apparent gains in survey performance as additional strata are added is small, as was done here (Figures 4.5-4.8). The current stratification scheme performed well in preliminary bootstrap analyses but this was probably an artifact due to the large number of strata in the current design.

Small samples tend to more often underestimate than overestimate mean catch within a stratum even though the sampling process is unbiased because the distribution of stratified random means for survey catch are right-skewed (Powell et al. 2017). Variances are harder to estimate than means and they also tend to be underestimated with small samples. It may be reasonable to assume that at least 25 stations per stratum are required to estimate variances with adequate precision (Figure 4.9). If 150 stations can be handled on GBK during one survey, then the number of strata should not exceed 150/25=6. Similarly, the number of strata in the south should not exceed 200/25=8. If resources for clam surveys were reduced and station numbers fell by 25% then the maximum number of strata would be 150*0.75/25=4.5 on GBK and 200*0.75/25=6 in the south. Future reductions in sample size are conceivable if survey funding is limited.

All current NEFSC survey strata are contiguous and based on location and depth without reference to expected catch. This conventional approach is probably necessary in multi-species bottom trawl surveys where spatial patterns in catch vary among species but not in single species NEFSC surveys for invertebrates including clams, sea scallops and northern shrimp. Discontiguous strata are currently used in a relatively small Canadian sea scallop fishery with good results, however, the scallop strata are based largely on habitat information rather than catch catches (Smith et al. 2017). Discontiguous strata make less sense in very large survey areas such and GBK and the southern area for surfclams and ocean quahogs because distant portions of the same strata are not likely driven by the same environmental patterns.

We expected to use FMSQ rather than current strata building blocks to construct new strata that precisely reflected underlying stock distributions but there were too few survey observations to reliably assign FMSQ to clusters so that cluster results were unstable. The use of current strata building blocks is an inherently conservative approach that has the advantage of helping main comparability of survey data collected historically and in future.

We expected to use univariate cluster analysis based on catch data alone or multivariate cluster analysis with a higher weight on catch but there was too little catch data (particularly for FMSQ) to weight it heavily without causing instability and discontiguous stratification patterns that were widely scattered and potentially problematic.

The decision to use fewer larger strata enhances potential gains in DEF for allocation because within-stratum variances should be substantially more accurate. It will be important to allocate stations optimally in future surveys, particularly if funding and the total number of tows is reduced. The variability in DEF for recommended options in bootstrap analysis suggests that there may be substantial variability in survey performance and DEF from year to year (Figures 4.5-4.8).

Minimizing the variance of the stratified random mean is just one goal in carrying out a survey and it is possible that future clam surveys will allocate extra tows to important strata (or vice-versa) and make other adjustments where necessary. Similarly, increases to the minimum number of tows per stratum could be considered if searching for new concentrations or monitoring changes in marginal habitat are deemed important.

11) *Will the recommended changes affect observation and estimation of biological characteristics, such as, length to weight relationships and growth rates?*

The recommended changes will improve collection of biological data and estimation of biological characteristics because a higher percentage of tows will catch the target species for sampling (more sampling during each survey). Thus, the density of potential samples within the species habitat area will increase. Sampling protocols that call for a fixed number of samples or measurements per stratum will need to be changed, however, because the number of recommended strata is much smaller than the number of current strata.

12) How would potential changes in the clam survey (e.g. lower survey frequency and increased precision) affect management advice and stock assessment modeling?

Recommendation: The proposed changes in the clam survey are expected to improve management advice and stock assessment modeling for both species in both areas.

Rationale*:* Post stratification and domain variance statistics indicate that historical trends estimated based on the current and recommended survey designs were similar and that variances under the recommended design were similar or smaller. Post-stratified estimates of mean survey catch density during 1982-2016 were often noticeably higher than estimates from the current survey design because of the 1% rule which excludes poor habitat areas with low catch (Figures 4.10-4.17). Swept area stock size estimates would be unaffected because the proportional decrease in area was about the same as the increase in mean density.

A simulation analysis shows that more precise survey data will make up for the reduction in survey frequency. Higher survey precision increased a simulated assessment model's ability to measure small changes in stock size for both species and areas. The longer period between surveys for ocean quahog results in larger changes between surveys that are easier to estimate. Rapid declines in ocean quahog stock size are unlikely but might be detected in mandatory logbook data which are spatially detailed (reported by trip and ten-minute square), considered accurate and evaluated annually. Survey frequency can be adjusted if stock size declines rapidly, fishing mortality increases dramatically or any other problems occur.

13) How often should future changes in stratification be considered?

Recommendations: Stations should be allocated to strata based on the best available information, optimal allocation calculations and other considerations prior to every survey.

Changes might be considered after ten years (three surveys for surfclams and two for ocean quahogs). However, changes should be avoided unless there is a serious need and substantially more data or new analytical methods for identifying strata. Most of the benefits from recommendations in this report will stem from omitting poor habitat areas from the survey and optimal allocation. Decisions about stratum boundaries were probably less important. Similar to other cases (Smith et al. 2017), a wide range of stratification options appear to work reasonably well for surfclams and ocean quahogs if optimal allocation is employed. The opportunities to reduce survey areas and separate species so that optimal allocation is possible are "low hanging fruit" that will disappear after the recommendations in this report are implemented making substantial future improvements more difficult. A great deal of additional survey data may be required to improve the survey design but such data will accumulate slowly because clam surveys will be infrequent, particularly for ocean quahogs.

Finally, the expense in making substantial changes to the clam survey should be considered. The recommendations in this report involved a WG consisting of industry representatives, academics from three universities, staff from the Greater Atlantic Regional Fisheries Office and Mid-Atlantic Fishery Management Council, several Northeast Fisheries Science Center scientists and two outside experts with statistical and survey expertise funded by the clam industry (Appendix 2). Five meetings were required over more than a year of time. External review will require funding as well.

14) What types of additional research would benefit the clam survey?

Recommendations:

a) Augment and refine the database used to avoid locations likely to damage sampling gear using data from other surveys bottom type data (e.g. from multi-beam sonar) and any other data or methodology available. Consider using the data to avoid gear damage in other surveys.
b) Carry out simulation studies to shorten cruise tracks and increase the number of stations that can be occupied in the available time at sea.
c) Data limitations precluded use of small building blocks (FMSQ) to identify strata. Spatial methods might make better use of available data. Options based on tree and GAM models with location, environmental and climatological data did not perform well because model estimates of habitat within small areas were patchy and heterogeneous but the general approaches see promising.

5. References

Breiman, L. 2001. Random forests. Machine Learning 45: 5-32.

Cargnelli L.M., Griesbach S.J., Packer D.B., Weissberger E. 1999. Essential fish habitat source document: Atlantic surfclam, *Spisula solidissima*, life history and habitat characteristics. NOAA Tech Memo NMFS NE 142; 13 p.

Cargnelli L.M., Griesbach S.J., Packer D.B., Weissberger E. 1999. Essential fish habitat source document: ocean quahog, *Arctica islandica*, life History and habitat characteristics. NOAA Tech Memo NMFS NE 142; 13 p.

Cochran, W.G. 1977. Sampling techniques, 3rd edition. John Wiley & Sons. NY.

Hartigan, J.A., and Wong, M.A. 1979. A K-means clustering algorithm. Applied Statistics 28: 100-108.

Northeast Fisheries Science Center. 2003. Report of the 37th Northeast Regional Stock Assessment Workshop (37th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish. Sci. Cent. Ref. Doc. 03-16; 597 p.

Northeast Fisheries Science Center. 2009. Ocean quahog assessment for 2009. *In:* 48th Northeast Regional Stock Assessment Workshop (48th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-15; 834 p.

Northeast Fisheries Science Center. 2016. Atlantic surfclam assessment summary for 2016. *In*: 61st Northeast Regional Stock Assessment Workshop (61st SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 16-13; 26 p. http://www.nefsc.noaa.gov/publications/

Northeast Fisheries Science Center. 2017a. Ocean quahog assessment summary for 2017. *In*: 63rd Northeast Regional Stock Assessment Workshop (63rd SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-09; 28 p. http://www.nefsc.noaa.gov/publications/

Northeast Fisheries Science Center. 2017b. Atlantic surfclams. *In*: 61st Northeast Regional Stock Assessment Workshop (61st SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-05; 466 p. http://www.nefsc.noaa.gov/publications/

Northeast Fisheries Science Center. 2017c. Ocean quahog assessment. *In:* 63rd Northeast Regional Stock Assessment Workshop (63rd SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-10; 409 p. www.nefsc.noaa.gov/publications/

Picard, R.R., and Cook, R.D. 1984. Cross-validation of regression models. J. Am. Stat. Assoc. 79(387): 575–583.

Powell, E.N., Mann, R., Ashton-Alcox, K.A., Kelsey M., Kuykendall, K.M. and Long, C. 2017. Can we estimate molluscan abundance and biomass on the continental shelf? Estuarine and Coastal Science 198: 213-224.

Powell, E.N., Mann, R., Kuykendall, K.M., Long, M.C., and Timbs, J. 2018. The “East of Nantucket” survey. Report for the New England Fishery Management Council. Newburyport, MA.

Sarndal, C. E., Swenson, B., and Wretman, J. 1992. Model assisted survey sampling. Springer, New York.

Wang, H. and Song, M. (2011) Ckmeans.1d.dp: optimal k-means clustering in one dimension by dynamic programming. The R Journal 3: 29–33.

Weinberg, J.R., Rago. P.J., Wakefield, W.W., and Keith, C. 2002. Estimation of tow distance and spatial heterogeneity using data from inclinometer sensors: an example using a clam survey dredge. Fisheries Research 55: 49-61.

Weinberg, J. R. 2005. Bathymetric shift in the distribution of Atlantic surfclams: response to warmer ocean temperature. ICES Journal of Marine Science 62: 1444-1453.

Wood, S.N. 2006. Generalized additive models. An introduction with R. Chapman and Hall/CRC, New York.

Appendix 1. Working group members, contributors and attendees.

Individual	Affiliation
Tom Alspach	Sea Watch International
Mike Bergman	NMFS, Northeast Fisheries Science Center
Russ Brown	NMFS, Northeast Fisheries Science Center
Jui-Han Chang	NMFS, Northeast Fisheries Science Center
Nicole Charriere	NMFS, Northeast Fisheries Science Center
Peter Chase	NMFS, Northeast Fisheries Science Center
Toni Chute	NMFS, Northeast Fisheries Science Center
Jessica Coakley	Mid-Atlantic Fishery Management Council
Steve Corriea	Massachusetts Department of Marine Fisheries (retired)
Jonathan Duquette	NMFS, Northeast Fisheries Science Center
Kevin Friedland	NMFS, Northeast Fisheries Science Center
Wendy Gabriel	NMFS, Northeast Fisheries Science Center, Mid-Atlantic Council SSC
Joe Godlewski	NMFS, Northeast Fisheries Science Center
Dvora Hart	NMFS, Northeast Fisheries Science Center
Dan Hennen	NMFS, Northeast Fisheries Science Center
Ed Houde	Center for Environmental Science, University of Maryland
Larry Jacobson	NMFS, Northeast Fisheries Science Center (Chair)
Rob Johnston	NMFS, Northeast Fisheries Science Center
Kelsey Kuykendall	University of Southern Mississippi, SCeMFiS
Paul Kostovick	NMFS, Northeast Fisheries Science Center
Chris Legault	NMFS, Northeast Fisheries Science Center
Chase Long	NMFS, Northeast Fisheries Science Center
Roger Mann	Virginia Institute of Marine Science, SCeMFiS
Nancy McHugh	NMFS, Northeast Fisheries Science Center
Jose Montanez	Mid-Atlantic Fishery Management Council
Daphne Munroe	Haskin Shellfish Research Laboratory, Rutgers University
Vic Nordahl	NMFS, Northeast Fisheries Science Center
Sara Pace	University of Southern Mississippi, SCeMFiS
Doug Potts	NMFS, Greater Atlantic Regional Office
Eric Powell	University of Southern Mississippi, SCeMFiS
Paul Rago	NMFS, Northeast Fisheries Science Center (retired), SCeMFiS
Eric Robillard	NMFS, Northeast Fisheries Science Center
Burton Shank	NMFS, Northeast Fisheries Science Center
Michael Simpkins	NMFS, Northeast Fisheries Science Center
Dave Wallace	Wallace and Associates, Ltd.

Appendix 2. Effects of sample size (total number of random tows) on survey accuracy for recommended options.

Recommendation: If possible, relative RMSE statistics should be used to evaluate relationships between accuracy and sample size in NEFSC clam surveys because CVs for stratified means may give an overly optimistic impression about survey accuracy (particularly for GBK surfclams).

Methods

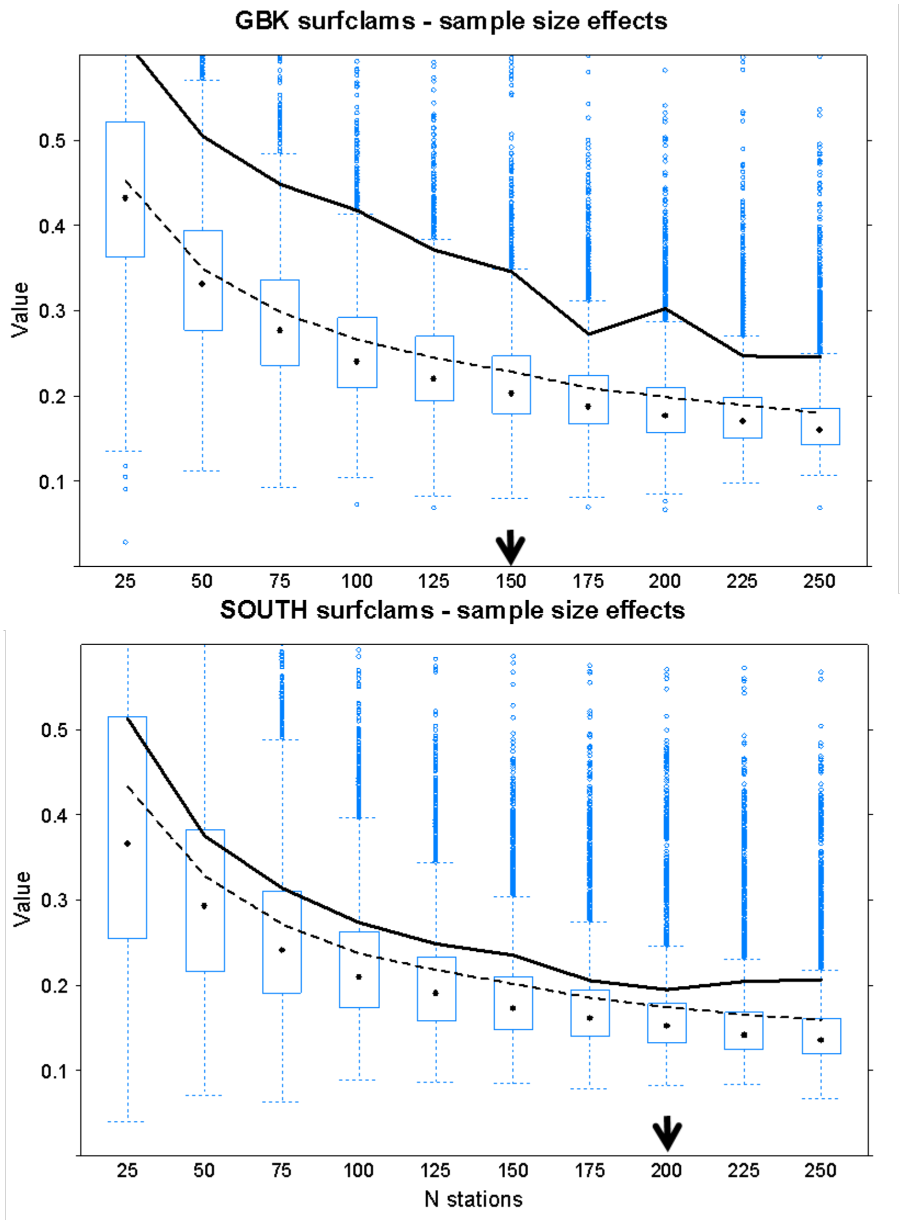
Bootstrap analyses using the recommended stratification options (2000 iterations for GBK surfclams and 1000 iterations otherwise) were carried out for both species and areas assuming that the total number of random tows was 25, 50, 225, or 250 and with optimal allocation. A relative root mean squared error was used to quantify accuracy:

$$RMSE_N = \frac{\sum_{j=1}^{N}\left(x_j - X\right)^2}{N} \Big/ X$$

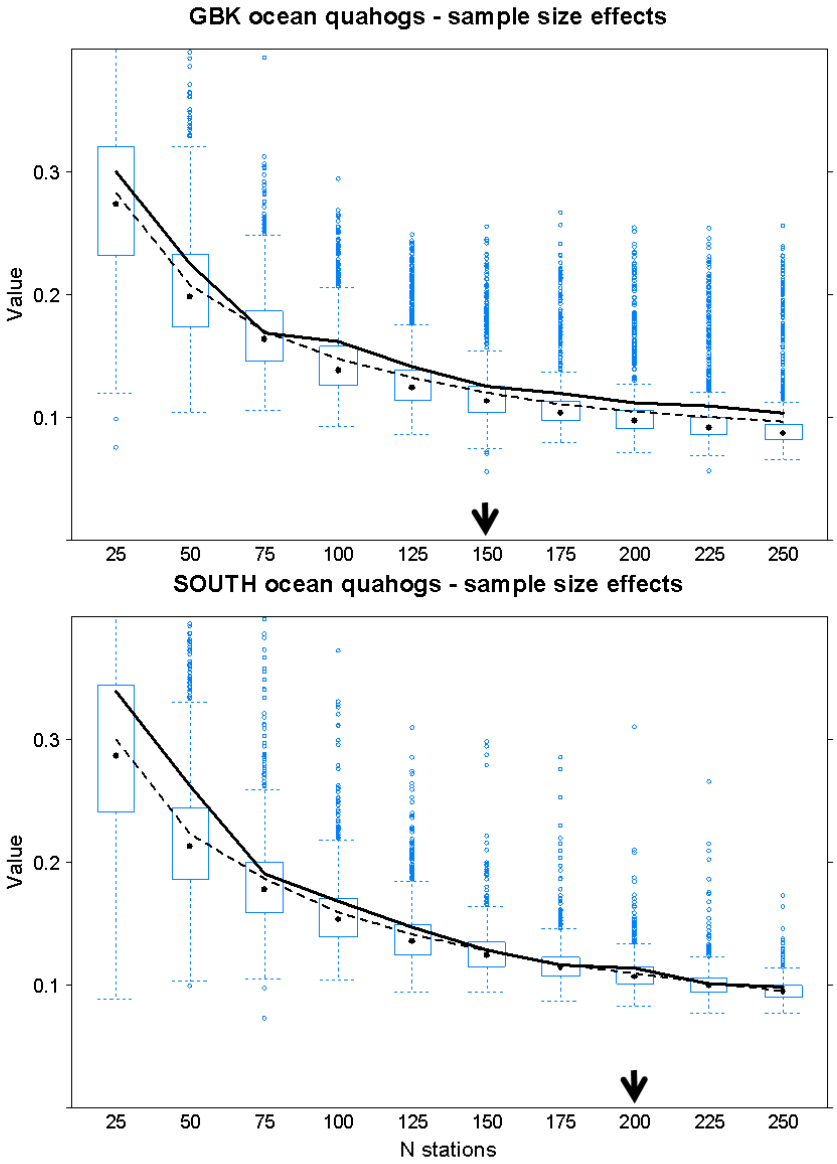
where *N* is the number of random tows, *X* is the true survey density and $\hat{x}_j$ is the estimated density based on the simulated survey in the j^{th} bootstrap iteration. We plotted the distribution of CVs $\left[\,SE\left(x_j\right)/\bar{x}\right]$, the mean CV and relative RMSE statistics for each case. The distribution central tendencies (median and mean) for CVs are of interest because they values encountered in real surveys.

Results

Results show how survey accuracy improves as sample size increases (Figures Appendix 2.1-2.2). Mean CVs and relative RMSE statistics were always similar, as would be expected in an unbiased survey. Median CVs were noticeably smaller than relative RMSE for surfclams in the GBK and southern areas. These differences reflect the skewed distribution of CV statistics and survey density estimates for surfclams which have more patchy distributions and more variable survey catches. The difference between the median CV and relative RMSE tends to become smaller as sample size increases and the within stratum variance estimates become more accurate.



Appendix Figure A2-1. Bootstrap results (2000 iterations for GBK and 1000 iterations for the southern area) showing relationships between accuracy (relative root mean squared error, solid line), average survey CV (dash line) and the distribution of survey CVs for surfclams in the GBK (top) and southern assessment areas (bottom). The symbol in the middle of the boxplots is at the median and the block shows the underlying spread. The number of tows assumed in evaluating stratification options (based on historical performance and indicated by arrows) was 150 for GBK and 200 for the southern assessment area. Note y-axis scale is different in the next figure.



Appendix Figure A2-2. Bootstrap results (2000 iterations for GBK and 1000 iterations for the southern area) showing relationships between accuracy (relative root mean squared error, solid line), average survey CV (dash line) and the distribution of survey CVs for ocean quahogs in the GBK (top) and southern assessment areas (bottom). The symbol in the middle of the boxplots is at the mean and the block shows the underlying spread. The number of tows assumed in evaluating stratification options (based on historical performance and indicated by arrows) was 150 for GBK and 200 for the southern assessment area. Note y-axis scale is different in the previous figure.

Table 2.1. Numbers of positive, zero and total stations for surfclams in the GBK area during 1997-2017 that were used in this analysis, by building block and original stratum number (final data set reflecting all decisions). Building blocks omitted based on the 1% rule are shaded.[3]

Area building block, orig. stratum		N positive tows	N zero tows	Total
GBK		330	300	630
763		0	1	1
	74	0	1	1
767		16	13	29
	54	1	0	1
	67	15	13	28
768		11	21	32
	68	11	21	32
769		13	24	37
	69	12	24	36
	70	1	0	1
770		27	37	64
	69	1	0	1
	70	26	37	63
771		6	16	22
	71	6	16	22
772		12	40	52
	72	12	40	52
773		7	34	41
	69	0	1	1
	73	7	33	40
774		15	27	42
	73	2	3	5
	74	13	24	37
75354		16	7	23
	54	15	7	22
	55	1	0	1
75556		31	5	36
	55	24	5	29
	56	6	0	6
	70	1	0	1
75758		33	12	45
	57	25	12	37
	58	8	0	8
75960		53	43	96
	59	41	42	83
	60	12	1	13
76162		73	7	80
	61	57	7	64
	62	16	0	16
76566		8	12	20
	65	8	12	20
747256		9	1	10
	47	8	1	9
	55	1	0	1

---

[3] Note: small numbers of tows from unexpected strata (e.g. one tow from stratum 69 in building block 770) occur because of errors in the original strata designation or because the tow was close to a stratum boundary and small differences in boundaries used originally and currently to define strata.

Table 2.2. Numbers of positive, zero and total stations for ocean quahogs in the GBK area during 1997-2017 that were used in this analysis, by building block and original stratum number (final data set reflecting all decisions). Building blocks omitted based on the 1% rule are shaded. See footnote for Table 2.1.

Area building block, orig. stratum		N positive tows	N zero tows	Total	Area building block, orig. stratum		N positive tows	N zero tows	Total
GBK		244	386	630	**85354**		6	17	23
863		1	0	1		54	6	16	22
	74	1	0	1		55	0	1	1
867		17	12	29	**85556**		6	30	36
	54	1	0	1		55	5	24	29
	67	16	12	28		56	0	6	6
868		24	8	32		70	1	0	1
	68	24	8	32	**85758**		2	43	45
869		32	5	37		57	2	35	37
	69	32	4	36		58	0	8	8
	70	0	1	1	**85960**		1	95	96
870		11	53	64		59	1	82	83
	69	1	0	1		60	0	13	13
	70	10	53	63	**86162**		1	79	80
871		20	2	22		61	1	63	64
	71	20	2	22		62	0	16	16
872		50	2	52	**86566**		5	15	20
	72	50	2	52		65	5	15	20
873		38	3	41	**847256**		2	8	10
	69	1	0	1		47	1	8	9
	73	37	3	40		55	1	0	1
874		28	14	42					
	73	5	0	5					
	74	23	14	37					

Table 2.3. Numbers of positive, zero and total stations for surfclams in the southern area during 1997-2017 that were used in this analysis, by building block and original stratum number (final data set reflecting all decisions). Building blocks omitted based on the 1% rule are shaded. See footnote for Table 2.1.

Area building block, orig. stratum		N positive tows	N zero tows	Total
SOUTH		**1386**	**1595**	**2981**
705		**8**	**41**	**49**
	5	8	40	48
	9	0	1	1
709		**127**	**237**	**364**
	5	0	1	1
	9	126	234	360
	10	0	1	1
	83	0	1	1
	84	1	0	1
710		**12**	**19**	**31**
	6	4	0	4
	9	0	1	1
	10	8	18	26
713		**30**	**116**	**146**
	13	30	116	146
714		**20**	**29**	**49**
	14	20	29	49
733		**51**	**25**	**76**
	33	51	25	76
734		**87**	**4**	**91**
	34	87	4	91
737		**18**	**4**	**22**
	37	18	4	22
738		**44**	**5**	**49**
	38	43	5	48
	39	1	0	1
741		**45**	**8**	**53**
	41	45	8	53
745		**18**	**18**	**36**
	45	18	18	36
746		**26**	**8**	**34**
	46	26	8	34
749		**1**	**1**	**2**
	49	1	1	2
750		**3**	**1**	**4**
793		**10**	**30**	**40**
	33	1	0	1
	93	9	30	39
794		**15**	**0**	**15**
	94	15	0	15
795		**2**	**33**	**35**
	95	2	33	35
796		**3**	**9**	**12**
	96	3	9	12
71112		**16**	**6**	**22**
	11	16	6	22
71516		**24**	**7**	**31**
	15	24	7	31
71920		**23**	**6**	**29**
	19	23	6	29
72324		**43**	**0**	**43**
	23	43	0	43
72728		**26**	**5**	**31**

Table 2.3 (cont.)

717		**27**	**56**	**83**
	14	0	1	1
	17	27	54	81
	87	0	1	1
718		**23**	**6**	**29**
	17	1	0	1
	18	22	6	28
721		**123**	**226**	**349**
	21	123	225	348
	25	0	1	1
722		**17**	**16**	**33**
	22	17	16	33
725		**5**	**91**	**96**
	21	0	1	1
	25	4	89	93
	26	1	0	1
	50	0	1	1
726		**12**	**7**	**19**
	26	12	7	19
729		**78**	**37**	**115**
	25	10	5	15
	29	67	31	98
	90	1	0	1
	92	0	1	1
	49	1	0	1
	50	2	1	3
781		**6**	**3**	**9**
	81	6	3	9
782		**6**	**0**	**6**
	82	6	0	6
783		**7**	**6**	**13**
	83	7	6	13
784		**15**	**13**	**28**
	84	15	13	28
785		**26**	**21**	**47**
	85	26	21	47
786		**18**	**7**	**25**
	86	18	7	25
787		**17**	**53**	**70**
	21	0	1	1
	87	17	52	69
788		**18**	**173**	**191**
	88	18	172	190
	89	0	1	1
789		**9**	**142**	**151**
	89	9	142	151
790		**5**	**15**	**20**
	90	5	15	20
	27	26	5	31
73132		**70**	**5**	**75**
	26	1	0	1
	27	1	0	1
	30	1	0	1
	31	65	5	70
	32	2	0	2
73536		**53**	**0**	**53**
	34	3	0	3
	35	48	0	48
	36	1	0	1
	39	1	0	1
73940		**45**	**1**	**46**
	39	43	1	44
	40	2	0	2
74344		**1**	**0**	**1**
	39	1	0	1
75152		**3**	**0**	**3**
	51	3	0	3
747148		**56**	**3**	**59**
	47	48	3	51
	48	7	0	7
	50	1	0	1

Table 2.3 (cont.)

730		**64**	**10**	**74**	**791**		**21**	**36**	**57**
	26	3	3	6		90	5	5	10
	29	0	1	1		91	16	31	47
	30	61	6	67	**792**		**9**	**56**	**65**
						29	0	10	10

Table 2.4. Numbers of positive, zero and total stations for ocean quahogs in the southern area during 1997-2017 that were used in this analysis, by building block and original stratum number (final data set reflecting all decisions). Building blocks omitted based on the 1% rule are shaded. See footnote for Table 2.1.

Area building block, orig. stratum	N positive tows	N zero tows	Total	Area building block, orig. stratum	N positive tows	N zero tows	Total	Area building block, orig. stratum	N positive tows	N zero tows	Total
SOUTH	**1457**	**1524**	**2981**	**838**	**7**	**42**	**49**	**895**	**12**	**23**	**35**
805	**9**	**1**	**10**	38	7	41	48	95	12	23	35
5	9	1	10	39	0	1	1	**896**	**12**	**0**	**12**
809	**161**	**34**	**195**	**841**	**3**	**50**	**53**	96	12	0	12
5	0	1	1	41	3	50	53	**81112**	**1**	**21**	**22**
9	160	33	193	**845**	**16**	**20**	**36**	11	1	21	22
10	1	0	1	45	16	20	36	**81516**	**1**	**30**	**31**
810	**8**	**23**	**31**	**846**	**8**	**26**	**34**	15	1	30	31
6	2	2	4	46	8	26	34	**81920**	**0**	**29**	**29**
9	0	1	1	**849**	**1**	**1**	**2**	19	0	29	29
10	6	20	26	49	1	1	2	**82324**	**1**	**42**	**43**
813	**38**	**65**	**103**	**850**	**4**	**0**	**4**	23	1	42	43
13	38	65	103	49	1	0	1	**82728**	**2**	**29**	**31**

Table 2.4 (cont.)

814		**1**	**48**	**49**		50	3	0	3		27	2	29	31
	14	1	48	49	**881**		**9**	**0**	**9**	**83132**		**11**	**64**	**75**
817		**5**	**54**	**59**		81	9	0	9		26	0	1	1
	14	0	1	1	**882**		**6**	**0**	**6**		27	1	0	1
	17	5	53	58		82	6	0	6		30	0	1	1
818		**1**	**28**	**29**	**883**		**13**	**0**	**13**		31	9	61	70
	17	0	1	1		83	13	0	13		32	1	1	2
	18	1	27	28	**884**		**23**	**5**	**28**	**83536**		**7**	**46**	**53**
821		**173**	**176**	**349**		84	23	5	28		34	0	3	3
	21	173	175	348	**885**		**45**	**2**	**47**		35	7	41	48
	25	0	1	1		85	45	2	47		36	0	1	1
822		**1**	**32**	**33**	**886**		**25**	**0**	**25**		39	0	1	1
	22	1	32	33		86	25	0	25	**83940**		**9**	**37**	**46**
825		**12**	**84**	**96**	**887**		**68**	**2**	**70**		39	7	37	44
	21	0	1	1		21	1	0	1		40	2	0	2
	25	11	82	93		87	67	2	69	**84344**		**1**	**0**	**1**

	26	1	0	1	**888**		**183**	**8**	**191**		39	1	0	1
	50	0	1	1		88	182	8	190	**85152**		**3**	**0**	**3**
826		**4**	**15**	**19**		89	1	0	1		51	3	0	3
	26	4	15	19	**889**		**146**	**5**	**151**	**89905**		**39**	**0**	**39**
829		**6**	**109**	**115**		89	146	5	151		5	38	0	38
	25	0	15	15	**890**		**14**	**6**	**20**		9	1	0	1
	29	6	92	98		90	14	6	20	**89909**		**165**	**4**	**169**
	90	0	1	1	**891**		**26**	**31**	**57**		9	163	4	167
	92	0	1	1		90	0	10	10		83	1	0	1
830		**4**	**70**	**74**		91	26	21	47		84	1	0	1
	26	1	5	6	**892**		**23**	**42**	**65**	**89913**		**42**	**1**	**43**
	29	0	1	1		29	3	7	10		13	42	1	43
	30	3	64	67		92	20	35	55	**89917**		**22**	**2**	**24**
833		**6**	**70**	**76**	**893**		**1**	**39**	**40**		17	21	2	23
	33	6	70	76		33	0	1	1		87	1	0	1
834		**52**	**39**	**91**		93	1	38	39	**847148**		**15**	**44**	**59**
	34	52	39	91	**894**		**7**	**8**	**15**		47	15	36	51
837		**5**	**17**	**22**		94	7	8	15		48	0	7	7
	37	5	17	22							50	0	1	1

Table 2.5. Area and descriptive information for surfclam and ocean quahog building blocks based on current strata. Areas were taken originally from NEFSC shellfish strata shape files and prorated where necessary based on a fine scale grid.

Area or region	Building block ID number Surfclam	Building block ID number Quahog	Area (km2)	Min depth (m)	Max depth (m)	Mean depth (m)	Mean latitude (decimal degrees)	Mean longitude (decimal degrees)
GBK	763	863	1,052	41.0	80.0	60.7	41.659	-66.900
GBK	767	867	672	18.1	59.9	48.9	41.377	-68.394
GBK	768	868	1,303	11.0	55.0	34.0	41.533	-68.122
GBK	769	869	3,093	21.0	54.0	40.5	41.206	-67.998
GBK	770	870	1,865	37.1	61.2	51.3	40.940	-68.174
GBK	771	871	576	34.0	59.8	49.2	41.995	-67.388
GBK	772	872	1,618	11.1	59.0	37.0	41.808	-67.683
GBK	773	873	1,660	17.0	65.6	40.3	41.525	-67.474
GBK	774	874	1,662	34.5	67.4	52.7	41.663	-67.180
GBK	75354	85354	1,230	42.9	80.0	65.4	41.237	-68.624
GBK	75556	85556	1,476	48.1	80.0	65.7	40.710	-68.741
GBK	75758	85758	781	49.8	80.0	66.6	40.735	-68.224
GBK	75960	85960	2,309	52.8	80.0	68.2	40.908	-67.581
GBK	76162	86162	2,805	41.0	80.0	64.1	41.247	-66.894
GBK	76566	86566	466	37.6	80.0	63.3	41.928	-67.787
GBK	747256	847256	1,052	62.0	80.0	71.6	40.647	-68.981
SNE	737	837	2,263	19.2	55.9	36.6	41.162	-71.242
SNE	738	838	919	43.8	56.5	51.3	40.949	-71.165
SNE	741	841	1,989	22.4	47.8	37.7	41.007	-70.401
SNE	742	842	1,221	44.4	55.6	50.3	40.830	-70.480
SNE	745	845	1,395	20.2	53.7	38.9	40.812	-69.621
SNE	746	846	703	30.3	66.1	51.6	40.705	-69.501
SNE	749	849	761	22.7	52.4	36.9	41.457	-69.571
SNE	750	850	535	37.8	66.6	52.4	41.333	-69.426
SNE	794	894	736	9.0	46.3	24.5	41.275	-71.318
SNE	795	895	953	13.0	31.3	23.3	41.104	-70.263
SNE	796	896	1,680	11.1	44.6	27.3	41.209	-69.615
SNE	73940	83940	3,770	51.0	80.0	64.5	40.678	-71.162
SNE	74344	84344	1,835	53.0	80.0	65.1	40.555	-70.383
SNE	75152	85152	574	49.1	80.0	68.4	41.450	-69.448
SNE	747148	847148	3,178	52.1	80.0	68.2	40.536	-69.499

Table 2.5 (cont.)

LI	729	829	4,067	25.0	58.4	37.9	40.376	-73.048
LI	730	830	2,398	38.6	78.0	51.4	40.163	-72.797
LI	733	833	1,238	25.0	48.8	38.5	40.802	-72.120
LI	734	834	709	43.8	56.6	50.6	40.666	-72.043
LI	791	891	1,236	9.1	37.5	22.4	40.460	-73.559
LI	792	892	562	9.0	30.9	21.9	40.663	-72.896
LI	793	893	331	9.0	29.1	20.7	40.930	-72.116
LI	73132	83132	3,749	52.0	80.0	65.5	40.048	-72.448
LI	73536	83536	2,775	52.9	80.0	68.1	40.466	-71.815
NJ	717	817	2,411	11.0	52.0	35.6	38.741	-74.353
NJ	718	818	823	42.0	61.0	51.9	38.547	-74.044
NJ	721	821	5,806	18.0	50.4	36.3	39.289	-73.765
NJ	722	822	1,046	42.7	58.8	51.0	39.056	-73.518
NJ	725	825	1,849	25.1	52.5	37.4	39.862	-73.468
NJ	726	826	541	38.0	66.1	50.4	39.846	-73.246
NJ	787	887	1,208	9.0	31.1	16.7	38.985	-74.613
NJ	788	888	1,660	11.6	33.0	21.2	39.383	-74.158
NJ	789	889	1,176	12.1	29.9	22.1	39.884	-73.933
NJ	790	890	337	12.2	37.0	21.5	40.297	-73.891
NJ	71920	81920	1,039	51.3	80.0	64.0	38.460	-73.823
NJ	72324	82324	3,110	47.1	80.0	67.2	38.988	-73.290
NJ	72728	82728	1,728	46.3	80.0	67.6	39.594	-72.920
DMVSVA	705	---	2,366	15.1	50.0	27.2	36.663	-75.288
DMVSVA	---	805	507	26.0	50.0	34.9	36.602	-74.964
DMVSVA	---	89905	1,860	15.1	38.8	25.1	36.680	-75.376
DMVSVA	709	---	6,496	16.2	58.8	32.6	37.275	-75.126
DMVSVA	---	809	2,614	26.0	58.8	39.2	37.229	-74.943
DMVSVA	---	89909	3,882	16.2	38.2	28.1	37.306	-75.249
DMVSVA	710	810	727	41.5	71.9	53.4	37.146	-74.791
DMVSVA	713	---	3,940	12.1	51.0	33.6	38.220	-74.639
DMVSVA	---	813	2,282	23.8	51.0	38.8	38.189	-74.535
DMVSVA	---	89913	1,659	12.1	39.8	26.5	38.263	-74.782
DMVSVA	714	814	703	38.2	63.6	51.4	38.078	-74.390
DMVSVA	781	881	1,241	9.0	25.9	17.3	36.752	-75.770
DMVSVA	782	882	420	9.0	22.0	15.6	37.056	-75.704
DMVSVA	783	883	740	9.0	26.1	17.2	37.356	-75.546
DMVSVA	784	884	1,149	9.0	26.9	16.6	37.771	-75.303

Table 2.5 (cont.)

DMVSVA	785	885	1,188	9.0	25.7	16.7	38.244	-74.994
DMVSVA	786	886	675	9.0	34.7	18.3	38.657	-74.942
DMVSVA	71112	81112	1,001	52.0	80.0	64.4	37.319	-74.670
DMVSVA	71516	81516	1,533	45.6	80.0	65.6	37.979	-74.278

Table 2.6. Parameters used in simulation analyses to evaluate effects of survey changes on stock assessments and management advice for ocean quahogs in the GBK and southern areas. CVs for survey errors are derived from relative errors= (estimate-truth)/truth in bootstrap analyses assuming either the original survey design and average station allocation or the recommended design with 5 strata and optimal allocation.

Parameter	**GBK**		**Southern area**	
Survey design	**Current**	**Recommended**	**Current**	**Recommended**
N strata[1]	14	6	40	7
Initial abundance (relative)	0.765	1.25	1.34	1.89
CV for survey errors	0.178	0.103	0.153	0.059
Increment parameter $\boldsymbol{\delta}$	0.01	0.01	0.01	0.01
Survey interval (years)	3	6	3	6

[1] In recommended option (not simulation).

Table 3.1. Total area of current survey strata at depths of 9-80 m, area retained based on the 1% rule and percent change.

Area (km^2)	GBK		South	
	Surfclam	Ocean quahog	Surfclam	Ocean quahog
Total	23,630	23,630	83,290	83,290
Included	17,514	13,652	46,499	54,051
% reduction	26%	42%	44%	35%

Table 3.2. Root mean squared error (RMSE) statistics measuring stability of potential stratification options with 2-6 new strata based on univariate cluster analysis and multivariate cluster analysis with preliminary data and a range of weights on catch density. Results for different species, areas, numbers of strata and type of building blocks (current strata vs. FMSQ) are not comparable. The most stable option in each set (lowest RMSE) is grey and the least stable (highest RMSE) are bold and italicized.

Surfclams										
Catch Weight	GBK current strata					South current strata				
N new strata	2	3	4	5	6	2	3	4	5	6
0.25	***74***	***66***	***60***	***55***	39	79	80	81	70	65
0.50	73	65	45	46	40	78	79	80	69	64
1.00	70	62	50	42	***42***	82	84	82	72	65
Univariate	63	54	51	44	36	***93***	***89***	***84***	***78***	***67***
Range	11	12	15	13	6	15	10	4	9	4
Catch Weight	GBK FMSQ					South FMSQ				
0.25	0	8	10	30	7	21	4	19	31	24
0.50	0	20	24	30	12	23	19	27	24	34
1.00	19	18	28	28	23	**48**	45	42	36	44
Univariate	***24***	***42***	***31***	***38***	***32***	30	***66***	***62***	***60***	***73***
Range	24	35	21	10	25	26	62	43	36	49

Ocean quahogs										
Catch Weight	GBK current strata					South current strata				
N new strata	2	3	4	5	6	2	3	4	5	6
0.25	63	***59***	37	39	37	116	***105***	94	***87***	***87***
0.50	61	56	46	39	38	116	***105***	97	***87***	78
1.00	61	55	***52***	39	38	116	107	95	84	77
Univariate	***64***	46	49	***42***	***39***	***120***	101	***99***	84	76
Range	3	13	16	3	1	4	6	5	3	11
Catch Weight	GBK FMSQ					South FMSQ				
0.25	19	10	1	8	20	18	0	22	23	21
0.50	20	16	1	30	32	17	14	24	17	46
1.00	23	26	9	17	25	35	42	47	68	31
Univariate	***36***	***33***	***33***	***32***	***26***	***61***	***63***	***88***	***85***	***77***
Range	17	23	32	24	12	44	63	67	68	57

Table 4.1. Six options for scheduling NEFSC clam surveys for surfclams and ocean quahogs over an eighteen year planning horizon. In the second column, "separate" means that surfclams and ocean quahogs are targeted separately (i.e. during separate years), "test in 3rd year" refers to time for gear testing (currently every third year), and "2x surfclams" means that the frequency of surfclam surveys is double the current frequency. "Sn", "Ss", "Qs" and "Qs" refer to surveys for surfclams and ocean quahogs in the northern (GBK) and southern assessment areas. Option 5 (target the two species separately, double the survey frequency for surfclams and halve the frequency for ocean quahogs) is recommended.

Option	**Description**	**Yr 1**	**Yr 2**	**Yr 3**	**Yr 4**	**Yr 5**	**Yr 6**	**Yr 7**	**Yr 8**	**Yr 9**	**Yr 10**	**Yr 11**	**Yr 12**	**Yr 13**	**Yr 14**	**Yr 15**	**Yr 16**	**Yr 17**	**Yr 18**	N surveys per year	
																				Surfclam	Quahogs
1	Current	Ss	Sn	T	Ss	Sn	T	Ss	Sn	T	Ss	Sn	T	Ss	Sn	T	■	■	■	0.33	0.33
		Qs	Qn		Qs	Qn		Qs	Qn		Qs	Qn		Qs	Qn		■	■	■		
2	Separate, test in 3rd year	Ss		T	Qs		T	Ss		T	Qs		T	Ss		T	■	■	■	0.20	0.20
			Sn			Qn			Sn			Qn			Sn		■	■	■		
3	Separate, test in 3rd year, 2x surfclams	Ss		T	Ss		T	Qs		T	Ss		T	Ss		T	Qs		T	0.22	0.11
			Sn			Sn			Qn			Sn			Sn			Qn			
4	Separate, no testing	Ss		Qs		Ss		Qs		Ss		Qs		Ss		Qs		■	■	0.25	0.25
			Sn		Qn		Sn		Qn		Sn		Qn		Sn		Qn	■	■		
5	**Separate, no testing, 2x surfclams**	**Ss**		**Ss**		**Qs**		**Ss**		**Ss**		**Qs**		**Ss**		**Ss**		**Qs**		**0.33**	**0.17**
			Sn		**Sn**		**Qn**		**Sn**		**Sn**		**Qn**		**Sn**		**Sn**		**Qn**		
6	Separate, no testing, 3x surfclams	Ss		Ss		Ss		Qs		Ss		Ss		Ss		Qs		■	■	0.38	0.13
			Sn		Sn		Sn		Qn		Sn		Sn		Sn		Qn	■	■		

Table 4.2. Median design efficiency (DEF) statistics for recommended stratification options for the NEFSC clam survey from bootstrap analyses (3000 iterations). DEF measures the percent reduction in variance of stratified random means relative to the variance from a random design. For example, the recommended option for GBK surfclams would be expected to reduce variance by 25% relative to a random design. DEF total = DEF allocation + DEF stratification, where the latter terms are the benefits of optimal allocation and the stratification scheme.

Species	**Area**	**DEF allocation**	**DEF stratification**	**DEF total**
Surfclams	GBK	18	7	25
	South	6	4	10
Ocean quahogs	GBK	10	14	25
	South	27	9	35

Table 4.3. Bootstrap results for recommended design options in the NEFSC clam survey (stratum level, 3000 iterations). “Mean variance” is the average within-stratum variances across bootstrap samples, “mean allocation” is the average allocation to each stratum and “perfect allocation” gives the optimal Neyman sample sizes for comparison. There were a total of 150 random station on GBK and 200 in the south. “Allocation/100 sq km” is the sampling intensity (number of random stations per area). Strata numbers correspond to Figures 4.1-4.4.

Species	Area	N strata	Statistic	New stratum						
				1	2	3	4	5	6	7
Surfclams	GBK	6	Area (sq km)	467	5632	4583	2239	2310	1231	--
			Mean variance	0.29	4.47	23.74	18.55	4.33	5.64	--
			Mean allocation	2	20	78	31	13	6	--
			Perfect allocation	1	34	65	28	14	8	--
			Allocation/100 sq km	0.44	0.35	1.71	1.38	0.54	0.52	--
Surfclams	South	6	Area (sq km)	4733	18076	10104	5305	4221	3300	--
			Mean variance	1.63	5.67	7.17	5.77	4.13	2.88	--
			Mean allocation	9	94	57	27	6	7	--
			Perfect allocation	12	84	53	25	17	11	--
			Allocation/100 sq km	0.19	0.52	0.57	0.50	0.14	0.21	--
Ocean quahogs	GBK	6	Area (sq km)	1663	3760	2806	1866	467	3091	--
			Mean variance	0.89	1.54	1.99	3.69	10.78	7.54	--
			Mean allocation	6	31	27	24	7	56	--
			Perfect allocation	10	29	25	23	10	54	--
			Allocation/100 sq km	0.34	0.81	0.97	1.28	1.54	1.80	--
Ocean quahogs	South	7	Area (sq km)	13037	5828	6837	11364	4689	8414	3881
			Mean variance	1.08	1.86	4.18	11.97	9.26	8.41	87.96
			Mean allocation	17	10	18	55	19	33	48
			Perfect allocation	18	11	19	52	19	33	49
			Allocation/100 sq km	0.13	0.17	0.27	0.48	0.40	0.40	1.24