Peer Review of the Southern New England Yellowtail Industry-Based Survey

February 26-27, 2007, Narragansett, RI

Ghislain Chouinard

Fisheries and Oceans Canada
Gulf Fisheries Centre
P.O. Box 5030
Moncton, NB Canada
E1C 9B6

Michael Martin

NOAA Fisheries- RACE Division
Alaska Fisheries Science Center
7600 Sand Point Way NE
Seattle, WA
98011-0070

John Sowles

Department of Marine Resources
PO Box 8
West Boothbay Harbor, ME
04575-0008

Introduction

On February 26 and 27, 2007, a peer review of the Southern New England Industry-Based Survey on Yellowtail flounder was conducted. The main objective of the survey was to assess the temporal and spatial abundance, distribution and size composition of yellowtail flounder (and associated species) within the Nantucket Lightship closed area, other proposed closed areas and adjacent areas. The survey also had a number of secondary objectives. The survey utilized a stratified random design based on strata defined by 30 minute latitude by 30 minute longitude rectangles and an equal number of fixed stations selected by the fishing industry. An external panel was invited to review the technical aspects of this survey following the terms of reference provided in Appendix 1. The organization of this report follows the terms of reference. Participants at the review are listed in Appendix II.

Main Findings and Conclusions of the Panel

The Southern New England Industry-based Survey on Yellowtail flounder is characterized by high sampling intensity and represented a significant amount of effort for the team conducting the survey. Survey team members should be congratulated for their dedication in conducting these surveys.

• The SNE yellowtail survey collected sufficient information to suggest that the Nantucket Lightship closed area does not meet the objective of protection of juvenile yellowtail. The review panel recommends that analysis of the efficacy of the closed area be formally conducted and documented. The survey dataset is considered useful to identify alternate closed areas.

• The review panel is satisfied with the attention to detail taken in the selection of the two vessels used for the survey in an attempt to minimize vessel differences as well as in the selection of the most appropriate trawl gear to be used in the survey.

• The review panel considers age samples collected during the SNE yellowtail survey have been very useful to complement the age-length tables for the assessment of yellowtail flounder, however, the utility of the survey in tracking changes in abundance is low due to the shortness of the time series.

• Many of the questions and concerns of the review panel derive from lack of details to ensure consistency and standardization. Procedures and protocols (e.g. towing speed, guidelines for declaring null sets, swept area, fishing station standardization, analyses) need to be further documented to ensure that data are correctly interpreted and repeatable methods are used if the survey is resumed. The panel recommends that funding be made available to complete the documentation and development of metadata for this dataset to preserve its integrity and usefulness.

• The mixed design of the survey (stratified random and fixed station) poses particular analysis difficulties. Survey estimates using all stations may be biased. Given the high sampling intensity, it should be possible to obtain unbiased indicators of the trends in yellowtail abundance by analyzing stratified random and fixed stations separately.

• If the survey is continued in the future, consideration should be given to using a unique sampling design. Information and knowledge gained during the 2003-2005 surveys would be useful in designing a survey.

• A wealth of information is available for analysis and would be expected to provide new knowledge on the biology of yellowtail flounder in the area, gain insights in survey design and to explore sampling strategies to collect information on multiple species. To the extent possible, the project team members, NEFSC scientists, and others should be encouraged to analyze these data.

• Because of the single-species nature of the survey, integration of this survey, as it now exists, with the NMFS survey is considered to be difficult and not cost effective.

• The Southern New England Industry-based Survey is considered a good example of a cooperative project that provides valuable information on yellowtail flounder in the area. Industry-based surveys are considered more appropriate to address short-term issues than to conduct long-term monitoring.

1. Design and Execution. The review panel should evaluate the statistical and scientific validity of the survey design relative to the program goals and objectives, highlighting strengths, weaknesses, and potential biases.
In particular:
a. evaluate the temporal and spatial design elements relative to survey objectives,

The yellowtail flounder industry –based surveys (IBS) were conducted in spring and fall from 2003 to 2005. The sampling schedule was chosen to closely follow the sampling in this area by Northeast Fisheries Science Center's (NEFSC) spring and fall trawl surveys and seemed well-suited to the main survey objective of evaluating the Nantucket Lightship closed area. The coincident timing could have allowed direct comparison between the two surveys' results and allowed the age and maturity information collected on the IBS surveys to supplement NEFSC survey collections. The primary contribution of the yellowtail IBS to the stock assessment process is the construction of seasonal age-length keys for the southern New England (SNE), supplementing the small age sample sizes for yellowtail flounder from the NEFSC surveys. The

review panel noted some differences in the timing of the surveys over the three-year period due primarily to logistical issues. While the effects of the differences in survey timing are unknown and may have been minimal, the review panel recommends that every effort be made to fix the timing of the surveys if they are continued in the future to ensure valid inter-annual and inter-survey comparisons.

The depth limits of the survey (10-45 fathoms) were determined at the outset of the survey, based on the known depth distribution of yellowtail in SNE. However, in the intervening years evidence from the monkfish fishery showed that yellowtail flounder were captured in substantial numbers at depths greater than 60 fathoms. To ensure that the entire population of yellowtail within the SNE area is sampled, the review panel suggests that if the survey is continued, it be extended to cover depths to at least 62 fathoms. A similar evaluation of the minimum depth limit of the survey would also seem warranted. Survey results suggested that yellowtail distribution likely extends to shallower waters than those covered by the survey. It is understood that the vessels would not be able to operate in very shallow waters.

The geographic extent of the survey seemed generally appropriate for the SNE stock boundaries as described to the peer review team. Historically, yellowtail flounder have been recorded from areas south of the current survey area, although apparently not in recent years. The recent combination of the SNE and the Mid-Atlantic stocks into a single stock necessitates the inclusion of any portion of the yellowtail population in the mid-Atlantic area in the survey to ensure maximum utility of survey results to the fishery management process. Relatively large catches in the fall 2005 survey at the southern limit of the survey suggests that perhaps the survey should be extended farther south to ensure that the entire population is sampled. The stratification scheme with allocation roughly proportional to area, coupled with the high overall density of stations, ensured very good spatial coverage of the survey area.

b. evaluate random versus industry selected sample stations,

About 300 stations were fished in each of the six surveys. About half of the stations were stratified random stations and the other half were fixed stations selected by the fishing industry based on historical distribution of yellowtail in the SNE area. In effect, each design forms a separate survey complicating the estimation of population abundance and its uncertainty.

The review panel was provided with very little information comparing the results of the fixed and stratified random survey designs as all data were combined and analyzed as if sampling was conducted using one sampling strategy. In our opinion, the analysis of the data in this manner is inappropriate given the distinct design elements of the two surveys. The locations of the industry selected tows were chosen based on areas of historically high industry catches. The selection of the locations of the tows in this manner is clearly non-random and very likely to produce biased estimates. A figure presented in the report comparing mean catch rates between the random and industry selected tows showed no significant difference between the two types of tows. However, it seems as if a more powerful test of this difference should be possible taking into account the stratification design. It is unknown whether such results would indicate a difference between the mean catch rates of the different tow types. However, it would be very surprising if

there were not an actual difference in catch rates given the experience and skill of the captains involved in the location and execution of the industry selected tows.

The review panel recognizes that a major impetus for using almost half of the total survey effort in a fixed survey design was to address industry concerns that the NEFSC surveys did not adequately cover areas of traditionally high yellowtail abundance. The review panel also recognizes that given the low sampling effort of the NEFSC surveys within this area, this is a valid concern and the ability of fishery managers to accurately and precisely estimate abundance in the survey area suffers as a result. A fixed station survey design can be quite powerful in monitoring changes in abundance over time as many potential sources of variability have been removed making inter-annual changes in abundance more easily detectable. However, fixed station surveys that are characterized by stations chosen only in areas of expected high catch may mask changes in total abundance over the entire survey area. Since these stations are more likely to be located in the most desirable habitats, they are more likely to be successfully occupied, even at lower overall population densities as seen, for example, in Grand Bank yellowtail flounder (Simpson and Walsh, 2004). Thus, a fixed survey may not reflect recovery or depletion as rapidly as one that does not focus effort on commercially successful areas.

The strata used in the survey were drawn as 30-minute blocks of latitude and longitude and this stratification method was quite effective at ensuring adequate spatial coverage. Although no stratum level analyses were presented, it seems unlikely the strata as currently constituted would be very effective in reducing population variance estimates as known correlates of yellowtail flounder distribution were not used in the development of the strata. Given the relatively high sampling density afforded by the current survey and the inherent advantages of stratified random sampling in total abundance and size composition estimation, the review panel recommends that future surveys be conducted as a single stratified random survey. If the strata boundaries are redrawn according to known correlates of distribution (e.g. depth) and effort is optimally assigned to strata, we believe that the industry would be quite satisfied with the resulting survey design in terms of sufficient coverage of high density areas, while allowing unbiased, precise estimates of relative abundance.

c. evaluate the potential success of estimating total biomass using the swept area method employed,

The swept area method as described used an estimate of distance fished multiplied by an estimate of door spread to estimate total area swept per tow. Estimates of total weight and numbers by species were then divided by this estimate of area swept to calculate density by tow. Distance fished was calculated by multiplying an average vessel speed by the time from brake set to haul back to infer distance. A standard tow was 1.9 nm at 2.8 to 3.1 knots. When haul back occurred before the 1.9 nm standard was reached due to large catches, hangs etc., the tow time was scaled to an average standard tow time before the distance was calculated. The review panel recommends that the distances fished be calculated directly from the recorded latitude and longitude as this involves fewer calculations, allows a single method to be used for all tows, and should provide more defensible results.

Door spread was estimated as a single value from the mean door spreads for all tows in all surveys. The review panel recognizes that using the door spread in the calculation of CPUE provides the most conservative biomass estimate and given the lack of quantitative information on the catchability of the fishing system employed that this may be appropriate. However, this measure almost certainly provides an underestimate of the density and total abundance. If the estimates used are considered relative measures of abundance, this is not of great concern. However, if the estimates are treated as absolute measures of abundance, a much better understanding of the capture process and catchability of yellowtail flounder would be of the highest priority for future research as this assumption would have an enormous influence on absolute abundance estimates. The review panel believes that calculating door spread on a tow by tow basis rather than using a single value for the survey would provide more accurate abundance estimates as differences in door spread with depth, wire out, currents, etc, could have an important influence on total abundance estimation.

Another important aspect in obtaining accurate abundance estimates is the accurate estimation of the weights and numbers of the species to be assessed. The sampling protocols used for yellowtail flounder allowed accurate estimates of weights and numbers on every tow. All yellowtail flounder caught were weighed. Length measurements were obtained from all yellowtail flounder with the exception of fall 2005 when large catches necessitated some subsampling for length. The sampling protocols used to sample other important groundfish species did not consistently allow for accurate estimates of weight and numbers, however. It became clear early in the survey that there was a conflict between the ambitious sampling plan in terms of tows per day, and the sampling protocols for the other groundfish species. The most significant problem that arose from this conflict was the direct use of the hail weight as an estimate of total catch for larger catches. It is clear from the information presented that not only were these estimates quite imprecise, there was also a strong negative bias in the estimation of total catch, leading to underestimates of the CPUE for these species. Consequently, the survey has not consistently provided high quality data to estimate the abundance of these species. It is the recommendation of the review panel that the survey more definitively resolve this conflict in future surveys choosing either fewer tows or less information on other species. If accurate abundance estimates for other species continues to be an objective, the development of a more robust method of estimating total catch size should be a priority. For example, perhaps baskets of the non sub-sampled (unsorted) catch could be counted and their weight estimated from the weighed sub-sample baskets to derive a more unbiased and accurate estimate of total catch.

The stratified random design results were not presented using the strata weightings one would expect from such a design, so it is difficult to ascertain the effectiveness of the stratification used in terms of abundance estimation. Had the total effort been apportioned strictly due to area across all survey strata, this would not have been a major problem. However, the disproportional allocation of effort to the Nantucket Lightship closed area introduced a potential source of bias in the total survey estimates. If a primary goal of future surveys is abundance estimation, it seems likely that much more powerful strata boundaries could be drawn taking advantage of what is now known about yellowtail distribution within the survey area to harness the power of stratified random designs to estimate total abundance.

d. evaluate sampling protocols, sub-sampling procedures and onboard processing of biological materials and total catch, and

In a number of areas, the review panel found that protocols and procedures used in the survey were unclear. Detail was insufficient and, occasionally contradictory (e.g. target tow duration). Much of this could be easily resolved by documenting these details with the assistance of the crews involved in the surveys.

The vessels chosen were of an appropriate size and horsepower to accomplish the survey objectives. The fishing systems (doors, nets, bridles etc.) used in the survey were appropriate for the survey objectives as well and were likely quite efficient at catching yellowtail flounder. The captains and crews of the vessels were by all accounts very familiar with the survey area and very experienced in the yellowtail fishery.

The review panel noted that all stations were fished during daylight hours. Given the large diel variation in catchability for yellowtail flounder observed in other surveys (Casey and Myers 2001); this should contribute to reducing variability in the estimates. Conducting surveys both during the day and night would require the estimation of conversion factors.

The assignment of tows to vessels occurred on an ad-hoc basis in the field in an attempt to use time and fuel efficiently. While the vessels and fishing systems seem well matched and there may well have been little or no vessel effect on CPUE's, such effects are often subtle and difficult to detect. Random allocation of stations to vessels would help mitigate the effects of any potential vessel differences on abundance and size estimation as well as allowing more robust methods for detecting such differences. Therefore, the review panel recommends that stations on future surveys be randomly assigned to vessels. The cost of such a change in terms of time and fuel in such would seem to be minimal, given the high density of stations.

The decision as to whether an assigned station could be sampled was a subjective decision of the captain. Clearly this subjective decision-making is a potential source of bias, but is perhaps unavoidable. Given the experience level of the captains involved and the relatively low number of stations that could not be sampled as assigned, this was probably not a major problem. A great deal of latitude was given to the captain on where to move the sampling effort if the assigned position could not be sampled for any reason. This is another potential source of bias and should be eliminated if possible. For example, captains could be given specific alternate start positions within each stratum if any of the originally assigned tows could not be sampled. In addition, the review panel suggests more clearly defined criteria to specify under what circumstances a tow should be re-sampled. The use of an inferred door angle of attack as a standardized sampling method is unusual, but the review panel has no information to indicate that this is not a valid method. However, an analysis on the repeatability between captains, boats and years of the method is warranted. If it can be shown that the method can, in fact, be standardized, the implemented method should be well documented.

The review panel found that the on deck sampling protocols used to evaluate the abundance, length distribution, maturity, and age of yellowtail flounder were sound and provided high quality information in support of survey goals. The chief scientist was given a great deal of

latitude in establishing sampling procedures from tow to tow. Some latitude is required to allow the field team to complete their mission; however the review panel recommends the establishment of more specific protocols in deck sampling procedures that should always direct the field party. As discussed in the previous section, the sampling protocols for other species were not adhered to consistently (e.g. fall 2003) due to large catches. The issue was addressed through expanded training.

Given the design and apparent efficiency of the survey trawl, catches were generally large which resulted in sampling difficulties. The SNE yellowtail surveys were conducted at a period of low yellowtail abundance which would imply that survey catches would be even larger if abundance were to increase. If future surveys are considered, an evaluation of fishing sets of shorter duration should be performed. Shorter fishing sets could provide adequate information and result in more efficient survey operations. It is understood that an important consideration in determining tow duration in this survey was the acceptability of the results by industry.

e. evaluate data post-processing procedures and archival policy.

The review panel was given only a very broad overview of the post-survey data checking procedures and archival policy, but data appear to have been vetted through essentially the same process used by the NEFSC surveys. Given that this is a mature and much inspected process, we are confident that these procedures would provide adequate data checking and produce high quality data. Storage in the same Oracle database as the NEFSC data seems appropriate and secure, although door spread data (a critical element in the current method of abundance estimation) should ideally be stored here as well. The review panel also applauds the efforts to make the data accessible to the public in an easy to use format via the web.

2. Data Utility. The review panel should evaluate the surveys' utility in assessing:

a. the efficacy of fishery closure areas,

The initial rationale used to determine the boundaries of the Nantucket Lightship closed areas was somewhat unclear but it is believed that the large fishing effort and quantities of juvenile yellowtail flounder being discarded in this particular area prompted the identification of this area for closure. At the time of the closure, there did not appear to be much other information on juvenile yellowtail flounder in the area that would have helped determine if other areas were important.

The high geographic resolution of the survey has provided very good information for the analysis of the efficacy of the closure for juvenile yellowtail. Based on charts of legal and sub-legal yellowtail, it appears that the closed area contains lower biomass of small or juvenile yellowtails than some other areas in the survey (e.g. areas south of Newport and Martha's Vineyard). These other areas would likely be more efficient in protecting juvenile yellowtail. More detailed analyses would need to consider the distribution of fishing effort in relation to the distribution of juvenile yellowtail. Areas of high juvenile abundance, but where there is little fishing effort, may not be as effective in protecting juveniles as areas where juvenile abundance is somewhat

less but high fishing effort results in juvenile bycatch. Other analyses such as the geographic distribution of yellowtail flounder by age groups would provide more detailed information on areas of importance for juveniles. The information collected would be useful to identify alternate closed areas.

It should be noted that these conclusions are based on the three years of information and the observation that some of the areas consistently showed the presence of juvenile flounder. It is unclear whether the patterns observed during the period of the study are indicative of persistent distribution patterns according to length. Based on the 2005 surveys, it appears that when juvenile yellowtail abundance is higher, they appear to be more widely distributed.

b. stock abundance,

Landings and the abundance index from the NMFS trawl survey indicate that this resource is currently at a very low level of abundance. The determination of relative stock abundance requires the examination of longer time-series. Given the shortness of the series, the survey currently has little utility in determining the relative abundance of yellowtail in a longer term context. Despite the low sampling intensity, the NMFS survey appears to capture the general trends in the stock. However, if the survey was to be continued and suggested changes in sampling design incorporated, the panel considers that the survey could provide a good index of abundance.

c. migratory or movement patterns,

Inferring migratory or movement patterns from surveys conducted twice a year can be difficult and often does not provide the same unambiguous patterns that direct methods such as tagging provide. The survey was not designed to answer this particular question and we believe that the utility of the survey in this regard is limited. Detailed examination of distribution by age classes according to depth may provide some insights into distribution by age and potential migration patterns but these would need to be validated using other techniques such as tagging. However, analysis of the fine-scale information collected during the survey could be quite valuable in formulating hypotheses and designing a tagging program. The panel was informed that a complementary yellowtail tagging program was conducted throughout the survey area and this should provide more precise information on movement or migration patterns.

d. reproductive demographics,

Since all fish sampled were sexed, good estimates of the sex ratio of the population should be available. Maturity observations were made during all surveys but there were few details on sampling design and the actual number of observations made. The panel was not presented with any results of the maturity information but given that two surveys with high spatial resolution were conducted per year, data collected during the survey should be very effective in addressing this issue. In New England, yellowtail flounder spawn during the spring so the assignment of maturity stages, particularly differentiating between immature and mature yellowtail, would be expected to be more accurate during the spring survey. Determining maturity stage during the fall survey would be expected to be more difficult and may not be essential. If this has not

already been done, it might be useful to conduct a histological validation of field observed maturities to assess the error rates in this process.

e. and other biological characteristics such as age and growth parameters.

The ageing material (scales) collected during the SNE yellowtail survey has been extremely valuable for the preparation of adequate age-length tables used in the assessment of this stock by NEFSC. Without these samples, age-length tables would likely be deficient resulting in higher uncertainties in estimates of the age structure of the population as well as in other parameters such as weight and length at age. The panel assumes that the techniques used in ageing the scales have been validated and error rates in this process are low. The number of samples collected was quite high and could be used for other studies. For example, if all samples were analyzed the results could determine whether there are fine scale geographic variation in size-at-age in the area. However, in terms of conducting routine abundance surveys, it seems likely that the number of fish sampled for scales could be reduced without a large decline in information for the age-length key, thereby freeing time for the scientists aboard. Given the numerous demands on the scientific crew participating in the survey, the review panel believes that a study to determine an efficient strategy of collecting scales would be appropriate.

3. Consistency. The review panel should evaluate the consistency and comparability among temporal and spatial sampling frames in relation to field procedures, gear selection and maintenance, vessel comparability, data acquisition, and analysis.

The review panel found that throughout project planning and execution, consistency and standardization were generally attempted to ensure reliable comparisons over space and time. As with any pilot project, some modifications in procedures occurred in the early phases of field work when unanticipated problems arose. The Spring 2003 survey was the first opportunity to work out and refine procedures. Overall, adjustments were appropriate and well thought out. One area for strengthening, and already mentioned, is better documentation of procedures to resolve questions around consistency and standardization, among others. Considering the effort and attention that went into completing the survey, these details deserve more thorough documentation while survey personnel are familiar with the details.

a. field procedures

Both spring and fall surveys were timed to coincide with the NEFSC Bottom Trawl surveys. Start and end dates varied slightly over the three years with Spring 2003 survey beginning three weeks later than the two subsequent spring surveys and the Fall 2005 survey beginning two weeks earlier and ending two weeks later than the previous fall surveys. Coincidentally, both the Spring 2003 and Fall 2005 represented the two highest catches of yellowtail flounder. Because this survey was designed to cover the entire Southern New England yellowtail flounder range, the minor change in timing probably had a minimal effect on total abundance estimation if the survey did, in fact, cover the entire SNE population. The effect of differences in survey timing on catch rates are unknown. An analysis of bottom water temperatures and/or climatological data might help to resolve this question.

The survey design apportioned tows proportional to area with the exception of the Nantucket Lightship Area where sampling effort was twice that elsewhere. Assigning different sampling to some areas is acceptable but must be taken into account in subsequent data analyses. Once the design was established, the number and distribution of tows remained substantively the same, season to season.

The use of two vessels and crews operating simultaneously introduces the potential for systematic differences in sampling techniques between vessels. To address and minimize this, commercial crews, project scientists, and land-based project managers frequently consulted by radio and telephone. Tows were not assigned to vessels as part of the planning process, instead captains chose, amongst themselves, which tows to complete. The review panel recommends that tows be randomly assigned to vessels rather than allowing captain discretion to minimize the effects of possible bias in the estimates due to differential vessel catchabilities. The cost of randomizing tows within the 30 minute squares in terms of time and fuel in such a high density survey would seem to be minimal. Each skipper also determined tow direction and where to move sampling effort if an assigned tow was untowable. Although many of these decisions are specific to the immediate circumstances (e.g. bottom obstructions, wind, sea-state and currents) and some flexibility is necessary, the Panel suggests that clearer guidelines be established describing how towable bottom is determined, and how to select alternate tows. This will help ensure consistency between vessels, crews and years and will ensure that the data collected provide maximum value.

The use of a standard scope table ensured that the amount of wire deployed was consistent between tows and vessels. The standardization of fishing tows was unusual in that it required a specific distance (calculated by the product of average speed and tow duration) while maintaining a specific door angle (16^{o}) considered to be optimum for fishing yellowtail flounder. This resulted in some variation in vessel speed to maintain the angle and consequently in tow duration as well. The reported variation in towing speed appeared to be small but it is unclear whether this approach may introduce additional variation. Detailed analysis of the tow duration, vessel speed and door spread data would provide further information.

b. gear selection and maintenance

The decision to equip the vessels with standard, identical nets alleviated many comparability concerns. The only appreciable difference in gear between the vessels were the doors, although they were quite similar in size and weight. Little detailed information was presented, however, on gear maintenance. The panel understands that when holes and tears were encountered, replacement panels and patches were sewn in instead of mending to help ensure nets were maintained to design specifications. Given the experience of the captains and crews in fishing similar gear and the relatively low frequency of net damage, it seems unlikely that deviation from net design was a major problem. Documentation of standard net repair procedures would alleviate doubt as well as ensure standard practices for future surveys. The review panel would also recommend the use of a net repair log that would document the details of any net damage event and how the net was repaired.

c. vessel comparability

Two very similar vessels were selected to simultaneously survey each of the 30 minute blocks. The F/V Mary Elena was 6 ft. longer but otherwise each was equipped with identical CAT3508 engines of 800 horsepower, the same reduction gear ratio of 6:1, same wheel size, and same Bollard pull at optimal towing speed rpm. Based on their virtually identical attributes, many of the concerns the committee might otherwise have had concerning vessel comparability were minimized and such comparisons are not a high priority. Preliminary statistical summaries of all data from all tows conducted over the three years did not reveal differences in vessel catchability suggesting consistency between vessels. However, by pooling all data for the three years, the ability to detect differences is greatly reduced. More properly, comparison of paired catches from adjacent areas during the same survey or at least comparing vessel catches for each of the six surveys would be a more rigorous and meaningful comparison. If differences were found after those analyses, then vessel comparison studies should be considered to test whether catchability differences between the vessels is an issue.

Throughout the duration of the six surveys, each of the two vessels was operated by the same captain. Although science crews and, to a lesser extent, fishing crews changed from survey to survey, survey scientists briefed captains and crews prior to each survey to ensure consistency and that standard protocols were followed. Daily communication between Rhode Island Department of Fisheries and Wildlife scientists, vessel crews, and project managers helped to minimize inconsistency. The review panel reiterates the need for clear and detailed documentation to ensure that standard procedures are followed consistently when vessels and crews change. If the survey continues, it would be beneficial to retain or contract with observers and other employees who have demonstrated ability to collect high quality data as this is perhaps the greatest asset to any survey.

d. data acquisition

All yellowtail flounder were weighed, sexed, enumerated and measured for length for all six surveys. Biological sampling protocols (e.g. scale samples, maturity, and stomach contents) appear consistent throughout the project; however few details are available on actual methods (e.g. fork vs. total lengths). Sample intensity varied according to workload and statistical needs. For example, initially, scales were collected from 1 yellowtail in 5 cm increments but by 2005, scales were collected from all yellowtail longer than 40cm. Varying sampling intensity is appropriate and not an issue of consistency.

On the other hand, methods used on species other than yellowtail flounder appear to have varied from cruise to cruise and year to year and at the crew's discretion depending on workloads and other circumstances. For example, a criterion of a total catch weight of 1000 pounds, based on hail weight, was set as the threshold for subsampling non-yellowtail species. Since hail weight can be considerably inaccurate, expanding sub-sample weights produces unreliable estimates of species-specific catch weights. The panel understands that collecting data on species beyond yellowtail was not the primary goal of this project. However, where additional information is collected, it is important that protocols be standardized and documented. If workloads prevent this, the Review Panel believes it is preferable to focus on processing a subset of species well rather than attempting to collect data on all species. Prior to each cruise, a prioritization of non-yellowtail species would be helpful.

NMFS Observer logs were employed in the initial survey. Beginning with the second survey, logs were reformatted to those used on the NEFSC standard trawl survey. This change does not appear to have introduced inter-annual comparability issues and likely reduced the probability of errors by eliminating the conversion between the two formats.

e. analysis

Data analysis for the project to date has focused on descriptive statistics and presentation of general patterns of distribution and abundance. Although analyses were not in-depth, they were consistently performed and presented. All figures were standardized making it easy to compare distributions across surveys. The review panel feels that many more analyses of the data are necessary and appropriate as discussed below.

4. Quantitative Analysis. The review panel should evaluate quantitative analysis techniques, measures of statistical precision, and recommend design or analytical processes that will improve the utility of existing survey data.

The panel noted that details of the computational methods used in the analyses were generally lacking in the report. Documentation of the methods is essential both for determining whether proper techniques have been used and for consistency in approaches between successive surveys.

A mixed survey design was used, but all stations were included in most of the analyses (e.g. length frequencies, biomass estimates, age composition) as if a single random design (all data pooled) had been used. The use of fixed stations selected on the basis that they were areas of yellowtail abundance would be expected to introduce some bias in the results. The review panel considered that, given the data, there are three possible ways of moving forward. The best approach would be to obtain unbiased estimates of biomass and length frequency distributions by using only the stratified random stations. These estimates could then be compared with those produced by the fixed stations. Another approach would be to investigate the use of geostatistical methods (e.g. kriging, see Petitgas (2001) for an overview) to analyze these data. These methods take into account the spatial variability and can be applied to various sampling designs to produce unbiased estimates. The panel notes that the appropriate use of these techniques requires significant expertise. Finally, if the bias introduced by using the fixed stations could be estimated by stratum and year then the bias could be taken into account (fixed stations would need to be 'corrected') and all stations could be used in the analyses. Estimation of bias usually requires large sample sizes and it is unclear whether the number of stations per stratum would be adequate to robustly estimate the bias introduced by fixed stations. In addition, the bias would need to be examined on the basis of length. The review panel considered that this is likely not a viable alternative.

The procedures and analyses related to catch standardization between sets should be reviewed and revised and documented where necessary. Catches were standardized to a standard tow length. The length of each tow was estimated by multiplying the time of the fishing set by the average speed. In some cases, it appeared that the standardization was done on time instead of tow length. The review panel was informed that data on distance towed based on GPS output

were available. At a minimum, the review panel recommends that standardization should be done on tow length based on distance towed calculated from start and end positions or better yet, course traveled. As noted earlier, trawl mensuration data are available for all stations so that catch standardization could be more accurately done on area swept by combining the trawl data with distance towed.

In general, the summary statistics provided a good overview but the aggregation of the data likely resulted in much information being missed. Aggregation of data from different surveys is unlikely to be very informative. Given the very high sampling intensity and the large amount of data collected, detailed analyses are encouraged. It is considered that these analyses could yield significant insights in the distribution patterns and biological characteristics of the species in the area. For example, in examining distribution, analyses on finer size-classes or by age classes may be useful in uncovering potential ontogenetic changes in distribution, geographic variation in size and weight-at-age or sex segregation. For the purposes of examining changes in distribution, all stations could be utilized and the use of contouring software (GIS), modeling (GLMs and GAMs) and geostatistical techniques may help uncover patterns.

Given the large number of stations, assumptions about the comparability of the two vessels could be examined by creating a comparative fishing data set using stations fished by the two vessels in close proximity and at about the same time. Comparisons of the results of paired fishing stations from the two vessels should be conducted on the aggregate yellowtail catch as well as on length frequency distributions. If significant differences between the two vessels are uncovered, these could be used to generate conversion factors to adjust catches to a standard vessel.

If further surveys are to be conducted, analyses of this data set would be extremely valuable in various aspects of sampling design such as determining the optimum sampling intensity required to minimize variance in abundance estimates and other parameters of interest. Some of these analyses have already been conducted by the IBS Design Subcommittee but could be enhanced with the additional data. In addition, analyses of the data should be conducted to identify the optimum sampling routines for biological characteristics (e.g. age, maturity, etc). It may be possible to reduce sampling of some characteristics without significant loss of precision thereby allowing more time to ensure that sub-sampling is conducted optimally or to collect other useful information on yellowtail or other species of interest.

The review panel concluded that the dataset produced by this project presents a significant opportunity for in-depth analyses that should not be lost. To the extent possible, the project team members, NEFSC scientists, and others should be encouraged to work with these data.

5. Integration. The review panel should evaluate the potential for integrating the IBS surveys with NMFS or other inshore trawl surveys or fishery independent monitoring programs. This includes interoperability and comparability of NEFSC current (R/V Albatross IV) and future (R/V Bigelow) bottom trawl surveys and states' near shore trawl survey programs.

The most useful aspect of these surveys as a complement to the NEFSC surveys to date has been the contribution of the length at age information. Due to the high spatial and temporal resolution

of the sampling effort and the higher catchability of the gear designed for retaining yellowtail, the number of scale samples collected is roughly an order of magnitude greater than that collected during a typical NEFSC survey providing much better information for stock assessment scientists. The high resolution sampling should also provide an excellent time series of yellowtail abundance within the survey area if the survey is continued. However, the large differences between the NEFSC research vessel and the IBS vessels, gear and survey designs preclude direct integration with the NEFSC survey results. A paired trawling experiment to estimate fishing power differences between vessels would be required for full integration, although two independent unbiased estimates of SNE yellowtail abundance would not be an entirely undesirable result. The SNE yellowtail survey also appears to have potential in providing an index of yellowtail flounder recruitment as strong year classes were detected on two occasions (Spring 2003 and Fall 2005).

Another path to better integration with other surveys would be better data collection for species besides yellowtail flounder. While collection of high quality information on other species was not a primary goal of this survey, better abundance and length information for these species would greatly improve its utility to other stock assessments. This strategy could potentially fill one of the last remaining gaps in regional inshore groundfish surveys. But this strategy is far removed from the original intentions of the survey. There was some effort to collect data on other species, but it was inconsistent and affected by use of hail weights which proved to be inaccurate as mentioned earlier. To improve collection of data on other species would require reducing the total number of tows and perhaps reducing the volume of the catches. Eliminating fixed stations and use of a completely stratified random design would allow some reduction in the number of tows while still allowing adequate spatial coverage and estimation of population parameters. To reduce the volume of the tows, some consideration should be given to reducing the tow duration. Given the early stages of the time series, if there was support to fill this region's inshore survey gap, the project team might consider a net design more typical of the other inshore surveys (e.g. MA, NEAMAP, ME-NH,) that could capture roundfish without significantly diminishing catchability for yellowtail.

6. Future of IBS and Other Initiatives. The review panel should be prepared to make recommendations concerning the continuation of IBS program and development of future fishery independent programs under NCRPP.

The SNE yellowtail survey team successfully reinforced the idea that industry based surveys can be effective, efficient and valuable. In addition to achieving the primary objectives of collecting information on yellowtail flounder, the review panel believes other unstated benefits that accrue from industry based surveys should be considered. The first is a common understanding of purpose, intent, and methods that comes from state and federal biologists working side by side with fishers to accomplish a common goal. The day to day interaction in the planning and execution of a survey engenders understanding and mutual respect between the parties that carries benefits well beyond the completion of the survey for both parties. Second, the manner in which this survey was executed has the potential to strengthen ties between state and federal agencies whose missions are coincident. This has obvious benefits for future interactions and collaboration, especially once the data are applied to management decisions. Third, fishers know

their equipment, vessels, and waters far better than scientists. And fourth, most industry members want to constructively engage in working toward solving problems with their industry.

Despite these benefits, there are also potential pitfalls of IBS surveys. Perhaps most importantly is the need to very carefully choose the projects. Although the committee believes that surveys of this type can be extremely efficient in terms of data collected per dollar spent compared to larger, government run surveys, the expense can still be quite large. Industry based projects must be chosen to provide the maximum amount of scientifically sound information at the lowest cost. Including the large number of fixed tows in the SNE yellowtail survey project design was one example of how industry collaboration may have increased costs with little gain in information. In this case, the added expense was probably worthwhile to build industry confidence. Such accommodations are to be expected, especially early in projects until more trust and understanding of sample theory is developed. However, once the data are analyzed, presented, and understood, efficiency can be improved by revising the survey design while remaining acceptable and credible to industry.

The Northeast Cooperative Research Partners Project was designed to engage commercial fishermen and address research questions of mutual concern between fishermen, scientists, and managers. As noted earlier, fishermen have expressed concern, if not frustration, over the delay in answers and the incorporation of the collected data into the management process. For example, during this technical review, one fisherman expressed an initial expectation that results of this survey would be used for the recent yellowtail flounder quota. He wondered whether the benefit of the IBS survey had been delayed to a point when it is no longer relevant. In this regard, long term monitoring, where patterns emerge over a long period of time, may not be able to realistically meet industry expectations.

Another important consideration is the need to be clear at the outset what the final product of a survey will be and what effect this survey is likely to have (or not have) on the management process. It would seem to be quite easy for the industry to expect immediate changes in the stock assessment and increased access to their fishery if the estimated biomass is found to be increasing, although this may not be the case.

Long-term monitoring, and fishery independent surveys in particular, requires a very different commitment than does short term research. For all the benefits of industry based surveys, two issues, longevity and bias, stand out those needing to be addressed in order for any industry based survey to succeed. A time series is most valuable when it consistently adheres to standard methods. To be economically viable, however, commercial vessels are continually upgrading to new and improved technologies (e.g. engines, winches, even hull modifications) that result in improved catch efficiency. Commercial vessels are routinely sold and therefore have the potential to be lost to the survey. Although the economics of research today are competitive with commercial fishing, it is uncertain how costs of commercial vessel contracts will be affected as stocks rebuild or even whether the interest to participate will remain. A dedicated survey vessel reduces many of these variables although at a monetary expense.

Fishery independent surveys are predicated on their ability to control bias, in particular a bias toward catching fish. Bias can be managed as captains understand the difference between

sampling and fishing, however, scientists must pay particular attention and be diligent throughout all phases of each survey to ensure the integrity of fishery independent data.

As a pilot or proof of concept project, the SNE yellowtail survey successfully demonstrated the value of industry collaboration while answering specific questions about yellowtail distribution and abundance as well as validity of the NLS Closed area. In this regard, the SNE yellowtail survey was an appropriate and successful discrete project of NCRPP. However, it is probably inappropriate and certainly undesirable and counter to the goals of building a continuous time series to rely on funding the SNE yellowtail survey as a long term monitoring project through a competitive granting process on a year by year basis.

References

Casey, J.M. and Myers, R.A. 1998. Diel variation in trawl catchability: is it clear as day and night? Can. J. Fish. Aquat. Sci. 55:2329-2340.

Petitgas, P. 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries, 2:231-249

Simpson, M.R. and Walsh, S.J. 2004. Changes in the spatial structure of Grand Bank yellowtail flounder: testing MacCall's basin hypothesis. Journal of Sea Research, 51:199-210.

Appendix I: Terms of Reference for Technical Peer Review of the NCRPP Industry Based Survey (IBS) Program

1. Design and Execution. The review panel should evaluate the statistical and scientific validity of the survey design relative to the program goals and objectives, highlighting strengths, weaknesses, and potential biases.
In particular:

a. evaluate the temporal and spatial design elements relative to survey objectives,
b. evaluate random versus industry selected sample stations,
c. evaluate the potential success of estimating total biomass using the swept area method employed,
d. evaluate sampling protocols, sub-sampling procedures and onboard processing of biological materials and total catch, and
e. evaluate data post-processing procedures and archival policy.

2. Data Utility. The review panel should evaluate the surveys' utility in assessing:

a. the efficacy of fishery closure areas,
b. stock abundance,
c. migratory or movement patterns,
d. reproductive demographics,
e. and other biological characteristics such as age and growth parameters.

3. Consistency. The review panel should evaluate the consistency and comparability among temporal and spatial sampling frames in relation to field procedures, gear selection and maintenance, vessel comparability, data acquisition, and analysis.

4. Quantitative Analysis. The review panel should evaluate quantitative analysis techniques, measures of statistical precision, and recommend design or analytical processes that will improve the utility of existing survey data.

5. Integration. The review panel should evaluate the potential for integrating the IBS surveys with NMFS or other inshore trawl surveys or fishery independent monitoring programs. This includes interoperability and comparability of NEFSC current (R/V Albatross IV) and future (R/V Bigelow) bottom trawl surveys and states' near shore trawl survey programs.

6. Future of IBS and Other Initiatives. The review panel should be prepared to make recommendations concerning the continuation of IBS program and development of future fishery independent programs under NCRPP.

Appendix II: List of participants at the Southern New England Yellowtail Industry Based Survey Peer Review Meeting, Narragansett, February 26 and 27, 2007

Review Panel

Ghislain Chouinard, DFO, Gulf Fisheries Centre, Moncton, N.B. Canada
Michael Martin, NOAA Fisheries, AFSC, Seattle, WA
John Sowles, Maine Department of Marine Resources, Boothbay Harbour, Maine

Monday, February 26, 2007

Name	Agency	email
Ghislain Chouinard	DFO/GFC,Canada	chouinardg@dfo-mpo.gc.ca
Wendy Gabriel	NOAA/NEFSC	wendy.gabriel@noaa.gov
Mark Gibson	Rhode Island DEM	mark.gibson@dem.ri.gov
John Hoey	NOAA/NEFSC	john.hoey@noaa.gov
Robert Johnston	NOAA/NEFSC	robert.johnston@noaa.gov
Laura Lee	ASMFC	llee@asmfc.org
Michael Martin	NOAA/AFSC	michael.martin@noaa.gov
Earl Meredith	NOAA/NEFSC	earl.meredith@noaa.gov
Sarah Pierce	NOAA	sarah.pierce@noaa.gov
John Sowles	Maine DMR	john.sowles@maine.gov
Fred Serchuk	NMFS/NEFSC	fred.serchuk@noaa.gov
April Valliere	Rhode Island DEM	april.valliere@dem.ri.gov

Tuesday, February 27, 2007

Name	Agency	email
Ghislain Chouinard	DFO/GFC,Canada	chouinardg@dfo-mpo.gc.ca
John Hoey	NMFS/NEFSC	john.hoey@noaa.gov
Bill Hoffman	Massachussetts DMF	bill.hoffman@state.ma.us
Robert Johnston	NMFS/NEFSC	robert.johnston@noaa.gov
Michael Martin	NOAA/AFSC	michael.martin@noaa.gov
Fred Mattera	NESTCo.	fm@nestco.necoxmail.com
Earl Meredith	NOAA/NEFSC	earl.meredith@noaa.gov
John Sowles	Maine DMR	john.sowles@maine.gov
April Valliere	Rhode Island DEM	april.valliere@dem.ri.gov