SUMMARY REPORT OF THE
NEFSC GROUNDFISH OPERATIONAL ASSESSMENTS
GEAR EFFICIENCY RESEARCH EXPERT REVIEW PANEL

Northeast Fisheries Science Center

Woods Hole, Massachusetts

Prepared by the Gear Efficiency Research Expert Review Panel

July 21, 2017

Panel Members

Steven X. Cadrin (Chair)

Micah Dean

Gregory DeCelles

James Gartland

Table of Contents

Background

The Northeast Fisheries Science Center is carrying out Operational Stock Assessments of species in the New England Fishery Management Council's Northeast Multispecies Fishery Management Plan. Peer review of the stock assessments will take place the week of September 11, 2017. These assessments will provide the scientific basis for fishery management decisions for the 2018-2020 Fishing Years. Several flatfish species are among the stocks being assessed.

The Operational Assessment Process was designed to be relatively streamlined and efficient. Operational Assessments incorporate additional years of data, but new sources of information are not to be introduced. To provide flexibility in the process, some exceptions are made by the Assessment Oversight Panel on a case by case basis depending on their importance and relevance.

Cooperative field studies were carried out in 2015 and 2016 to estimate relative catch efficiency of the NEFSC trawl survey gear, particularly regarding flatfish. Results from these efficiency studies could be used to provide context and estimates of catchability for the Operational Assessments. Before using results from the efficiency studies, this independent peer review is being conducted to evaluate whether the field studies were carried out with acceptable scientific methods and standards, and whether enough data were collected to reliably estimate relative efficiency by species. Peer review of the efficiency studies needs to be completed in advance of the groundfish Operational Assessment working papers in order to meet the schedule leading to the Assessment Peer Review during the week of 11 September.

A twin trawl was used with two identical nets with the exception of the sweep. One net had a rockhopper sweep, and one net had a chain sweep. The 2015 study was conducted south of Cape Cod and on Georges Bank, targeting yellowtail flounder. The 2016 study was conducted on the northern edge of Georges Bank and in western Gulf of Maine, targeting witch flounder. If the chain sweep efficiency is assumed to be 100%, the catch in the rockhopper sweep trawl can then be expressed relative to the chain sweep trawl and expressed as maximum efficiency with a rockhopper sweep. The estimate of maximum efficiency can be compared with model based catchability estimates to evaluate the plausibility of model estimates. The efficiency estimates can also be used in empirical assessment approaches to calculate a minimum swept-area biomass.

In addition to the efficiency estimates, information on bridle efficiency will be presented. The study presented does not allow for an estimate of bridle efficiency, but supports the conclusion that bridle efficiency is negligible. In addition, observations made during the twin trawl studies described above supports the inference of low bridle efficiency.

The panel will be asked to evaluate the scientific quality of the efficiency estimates from the twin trawl studies and the bridle efficiency study and provide a recommendation on the appropriateness of using this work in the flatfish assessments that are part of the upcoming Operational Assessments.

Process

This report summarizes independent reviews by Panel members. Panelists reviewed working papers in advance of a webinar. During the webinar, background information was provided, analysts presented information from the working paper, and Panel members asked for clarification and additional information. Panel members submitted individual reviews to the chair, and the chair drafted this summary. This report should be considered a compilation of independent reviews rather than a consensus summary, but panelists were in agreement on many conclusions.

Schedule

- Papers for Review & Terms of Reference Distributed to Reviewers- July 14, 2017
- Panel and NEFSC Webinar- Tuesday, July 18, 2017 9:30 am to 11:30 am
- Panel provides input to Chair to address TOR and Chair delivers Review Report- July 21, 2017

Background Reference Documents

- Brooks EN & KJ Curran. 2016. Eastern Georges Bank Cod and Haddock, and Georges Bank Yellowtail Flounder TRAC. TRAC Proceedings 2016/01.
- Miller TJ. 2017. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. Can. J. Fish. Aquat. Sci. 70: 1306–1316
- Miller TJ, C Das, PJ Politis, AS Miller, SM Lucey, CM Legault, RW Brown & PJ Rago. 2010. Estimation of Albatross IV to Henry B. Bigelow Calibration Factors. NEFSC Ref. Doc. 10-05.
- NEFSC. 2017. SAW/SARC 62 Black Sea Bass and Witch Flounder. NEFSC Ref. Doc. 17-10.

Working Papers for Review

- Hare J, J Hoey, J Manderson, M Martin, P Politis, D Richardson & C Roebuck. 2016. Empirical estimates of maximum efficiency of Witch Flounder *Glyptocephalus cynoglossus* L. on the Northeast Fisheries Science Center Fall bottom trawl survey. SAW62 Working Paper.
- Miller TJ, M Martin, P Politis, CM Legault & J Blaylock. 2017. Some statistical approaches to combine paired observations of chain sweep and rockhopper gear and catches from NEFSC and DFO trawl surveys in estimating Georges Bank yellowtail flounder biomass. TRAC Working Paper 2017/xx.
- Politis PJ & TJ Miller. 2017. Bridle herding efficiency of a survey bottom trawl with different bridle configurations. TRAC Working Paper 2017/xx.
- Richardson D, J Hoey, J Manderson, M Martin & C Roebuck. 2017. Empirical estimates of maximum efficiency and minimum biomass of Georges Bank yellowtail flounder on the NEFSC bottom trawl survey. TRAC Working Paper 2017/xx.

Webinar Presentations

- Gear Efficiency Research Expert Review Panel Webinar – Russell Brown
- Development of Maximum Catchability Estimates for Flatfish Species on the Northeast Fisheries Science Center Bottom Trawl Survey – David Richardson
- A statistical approach to evaluating effects on and estimation of relative catch efficiency, and scaled abundance or biomass indices – Tim Miller

In addition to the working papers, the Panel requested results from the 2015 and 2016 Chain Sweep Study for all target species (Table 1).

Sampling Design and Data

Term of Reference 1) Is the sampling design and the resulting data collected by the twin trawl studies appropriate to estimate maximum efficiency by the NEFSC Trawl Survey for use in the following operational assessments?

- ***Cape Cod/Gulf of Maine yellowtail flounder***
- ***Southern New England/Mid-Atlantic yellowtail flounder***
- ***Georges Bank yellowtail flounder***
- ***Georges Bank winter flounder***
- ***Gulf of Maine winter flounder***
- ***Southern New England/Mid-Atlantic winter flounder***
- ***Gulf of Maine/Georges Bank American plaice***
- ***Witch flounder***
- ***Northern windowpane flounder***
- ***Southern windowpane flounder***

The Review Panel concludes that the sampling design and the resulting number of stations with positive catches in the twin trawl studies are appropriate to estimate maximum efficiency by the NEFSC Trawl Survey for most of the flatfishes listed. The twin trawl experimental design was well conceived and well executed, including a relatively large number of tows over a variety of conditions that are relevant to the NEFSC Trawl Survey, and involved collaboration with fishermen who have expertise in targeting flatfish. The maximum efficiency estimates derived from the twin trawl studies provide an informative diagnostic to verify model estimates of catchability and stock size.

The use of a twin-trawl approach to quantify relative catchability is considered to be the most effective way to control for a multitude of confounding variables. Comparing the NEFSC Trawl Survey gear to identical gear with a chain sweep was appropriate for the estimation of maximum efficiency, because chain sweeps maintain close proximity of the footrope with the bottom, particularly when compared with the NEFSC Trawl Survey's rockhopper sweep. Removing flotation from the chain sweep net was also appropriate for maintaining consistent bottom contact. The restrictor lines maintained consistent wingspread and area swept for both nets over a range of depths, but are not likely to bias results because of the behavior of many flatfishes in the mouth of a trawl.

The seasonal timing of experiments matched the fall NEFSC Trawl Survey well, and sampling during day and night improves the applicability to NEFSC Trawl Surveys. Station locations were selected in collaboration with fishermen to target flatfish in Southern New England, Georges Bank, and in the Gulf of Maine in 2015 and 2016, representing a diversity of flatfish habitats.

The number of tows with positive observations is the unit of observation that should be considered in determining adequate sample size. The data and resulting estimates are appropriate for use in operational assessments of witch flounder (53,297 fish at 117 stations), yellowtail flounder (13,040 fish at 121 stations), American plaice (30,306 fish at 118 stations). The data available for windowpane flounder (4,926 fish at 110 stations) are marginally acceptable for estimating relative efficiency. There may be too few stations with winter flounder (2,212 fish at 63 stations) to produce reliable estimates of relative efficiency. Sample sizes may also be insufficient for determining other factors of efficiency such as diel period or fish length.

The Northeast Trawl Advisory Panel recently considered information from several other experiments (www.mafmc.org/council-events/2017/ntap-meeting) that may provide additional information on maximum efficiency of the NEFSC rockhopper survey gear that can be considered in the 2017 operational assessment update. Since 2009, NEFSC has collaborated with fishermen to evaluate efficiency of the NEFSC survey trawl. For example, a paired trawl study conducted in 2009 and 2010 may provide additional information on maximum efficiency of the NEFSC rockhopper survey gear that can be considered in the 2017 operational assessment update. For example, 33,796 winter flounder were caught at 353 stations, and 14,880 windowpane flounder were caught at 344 stations in the paired trawl study.

Analytical Approaches

Term of Reference 2) Are the analytical approaches presented in the working papers for witch flounder and yellowtail flounder from the twin trawl studies in 2015 and 2016 appropriate for estimating 1) relative efficiency of the chain sweep and Bigelow rockhopper gear, 2) size effects on efficiency, and 3) diel effects on efficiency for the operational assessments of the stocks listed above? (Note: this TOR is not being broadened to include review of other approaches in the working papers that address diel effects in the annual trawl survey data collected with NOAA ship Henry B. Bigelow rockhopper gear. That topic would require considerable time to review and will not be addressed in detail here.)

The working papers provide several alternative approaches to estimating relative efficiency, including relatively simple empirical estimates (Richardson et al. 2017) and more complicated model-based estimates (Hare et al. 2016, Miller et al. 2017). The methods used to estimate relative efficiency have been applied to paired-tow data in the past and are considered to be acceptable methods. The application of both methods produced similar estimates of relative efficiency for witch flounder.

Determining the optimal estimator is a challenge for both methods, with both approaches offering alternatives for considering diel effects and the effect of fish size. Unfortunately, sample sizes are small for some categories of fish size and time of day. Therefore, parsimony and representativeness of NEFSC Trawl Survey data should be considered in the determination of an optimal model.

An important assumption in the application of maximum efficiency estimates to flatfish stock assessments is that the stock is 100% available to the NEFSC Trawl Survey. This assumption should be tested for each stock when calculating area-swept biomass or survey catchability for comparison with assessment models.

Although the 2015 and 2016 net efficiency studies were conducted throughout Georges Bank, the Gulf of Maine, and southern New England, efficiency of the NEFSC Survey trawl may vary among habitats and among stock areas. Sample sizes may be insufficient to test for habitat or stock area effects or to estimate efficiency by stock area, but the available data should be explored to investigate differences among stock areas. For example, the evaluation of habitat effects on relative efficiency of yellowtail flounder should be applied to other flatfishes, and the approach could be applied to investigate potential biases associated with excluding the deepest NEFSC Trawl Survey habitats from the experiment. Habitat may also influence the effect of fish size on relative efficiency. The effect of fish density on relative efficiency should be evaluated to test if the targeting of flatfish may have biased estimates of efficiency.

Another aspect of applying maximum efficiency results to operational approaches is the time series of NEFSC Trawl Surveys to be considered. The empirical approach for Georges Bank yellowtail flounder is limited to the Bigelow survey time series (TRAC 2017). Alternatively, the empirical approach for witch flounder includes the entire time series of NEFSC Trawl Survey data (NEFSC 2017), and the estimate of maximum efficiency for the Bigelow survey was multiplied by the Bigelow/Albatross calibration coefficient (Miller et al. 2010, Miller 2013) thereby compounding uncertainty in the estimates of relative efficiency and inter-survey calibration.

Richardson et al. 2017. TRAC Working Paper 2017/xx

The methods described by Richardson et al. (2017) are reasonable, clearly interpretable and demonstrate the sensitivity of the resulting biomass estimates to alternative statistical models. The analytical approach is appropriate for estimating the relative efficiency if there are sufficient number of stations with positive catch.

Selecting the optimal model should be based on the significance of diel effects on relative efficiency and evidence of <100% efficiency of the chain sweep during the day. An evaluation of the influence of each variable on rockhopper efficiency should be considered in the application of maximum efficiency in the 2017 operational stock assessments. Based on the results of the 2015 and 2016 twin trawl studies and previous comparative fishing experiments, there are significant diel differences in net efficiency for the five flatfish species being considered for the 2017 operational assessments. Such diel differences are consistent with behavioral studies that show how that flatfish react to trawl gear differently, as a function of ambient light levels, and there are considerable differences in daytime and nighttime catch rates in commercial fisheries for most flatfish. Flatfish behavior also varies between day and night, with some flatfish (e.g., yellowtail flounder) exhibiting more frequent off bottom movements at night. Applying diel efficiency estimates to individual tows seems preferable to using the proportion of day and night tows to calculate a weighted average efficiency that is then applied to an aggregate index.

Catch ratio results suggest that relative efficiency varies by fish size, so that length-based estimates of relative efficiency are most appropriate for application to the operational assessments. The general approach applied to witch flounder and yellowtail flounder is appropriate for estimating size effects of net efficiency. However, estimating relative efficiency of yellowtail flounder in 1cm length bins produces a complex result that is difficult to interpret.

Hare et al. 2016. SAW62 Working Paper; and Miller et al. 2017. TRAC Working Paper 2017/xx

The modeling framework of Miller et al. (2017) applied to yellowtail flounder and Hare et al (2016) applied to witch flounder is more complex, including multiple random effects, beta-binomial dispersion, and cubic spline smoothed functions for length. This statistical approach was developed to derive calibration coefficients between the Albatross and Bigelow survey systems (Miller 2013). The Panel concludes that simpler models may be more appropriate for estimating the efficiency of the Bigelow NEFSC Trawl Survey from the twin-trawl data, because of more limited sample sizes and the simpler sampling design (twin trawls with only sweep differences).

Relative efficiency is expected to be a relatively simple function of fish size, and the flexibility of cubic spline smoothers do not improve the estimation of relative efficiency. Considering the limited number of stations for estimating relative efficiency and the significance of other factors (e.g., diel, size, habitat, stock area, density), cubic splines require more parameters than necessary. Multimodal length functions of footrope escapement are difficult to interpret and may be artifacts of the sampled length distribution. The Panel is concerned that the cubic splines for length effects of witch flounder (Hare et al. 2016) and yellowtail flounder (Miller et al. 2017) are overfitting sample data and may not be robust for application to NEFSC Trawl Survey data. Candidate models should include simple parametric length models (e.g., linear, logistic, log-logistic), and cross-validation (i.e., out-of-sample prediction) should be applied to test the robustness of estimates to extrinsic data, so that estimated length functions can be reliably applied to a wide range of size distributions. The investigation of size effects should also examine catch-ratio-at-length observations, model predictions and residuals. Model selection according to parsimony and the Akaike information criterion (AIC) should only consider valid models with no diagnostic problems with model residuals.

The available information from NEFSC Trawl Surveys may also be insufficient to estimate length-weight relationships by year and season. Several of these flatfish stocks are depleted, and few specimens are sampled in each survey. Multi-year length-weight estimates may be more reliable.

Area Swept

Term of Reference 3)Should the stock assessments use wing swept area, as done for yellowtail flounder and witch flounder assessments, in the operational assessments for above listed stocks?

Assessments of flatfish stocks should use net wingspread instead of doorspread when calculating swept area. Politis & Miller (2017) and Richardson et al. (2017) review previous studies and consider the design of the Bigelow survey gear (e.g., narrow door spread, minimal contact of ground cables) to conclude that wing spread is the most appropriate measure to determine area swept for flatfishes.

The first Northeast Trawl Survey Advisory Panel, including fishermen and scientists, designed the NEFSC survey trawl with the intention of minimizing herding between the net and doors, and net manufacturers collaborated to determine the minimum bridle length necessary to achieve the optimal net spread in an effort to avoid herding. Groundgear of flatfish trawls is typically covered with rubber disks (‘cookies’), and interspaced leads and lengths of chain are attached at both terminal ends of the bridle wire to maintain consistent contact of bridles with the bottom. Bare wire is not expected to produce much herding, particularly at the length used by the survey trawl.

Field-based observations during the bridle herding and twin trawl experiments indicated that herding by the bottom bridles on the survey trawl is negligible. Video images showed that the bottom bridle is not in contact with the seafloor for at least 2m forward of the bunt bobbin. The grease wear patterns on the bridles, and the paint applied to the bridles during the twin trawl study provide evidence that the bridles are not consistently on the bottom, especially near the wing ends.

Behavioral studies suggest that flatfish will not herd in response to the bridles unless they are struck by the bridle itself. Because there is not consistent bottom contact by the lower bridles, it is unlikely that the bridles are herding flatfish effectively. Any flatfishes that are herded toward the trawl by the forward sections of the bottom bridle escape beneath the bridle in the zone ahead of the bobbin. Previous studies conclude that herding by the bridles should be the most effective during the day time. Experienced trawl fishermen also report that flatfish herd in response to commercial trawl gear during the day, but not at night.

Unfortunately, the results of the bridle herding experiment are inconclusive for most species (Politis & Miller 2017). The experimental design is generally appropriate and has been successfully applied in a previous study. However, sample size may be insufficient for the variability in an alternate tow design, particularly if herding is negligible in the Bigelow survey system. There is no evidence of a significant herding effect during the day for any of the flatfish species that were examined, and no precise estimates of day time herding efficiency were calculated for the six flatfish species. Herding efficiency was poorly estimated for the six flatfish species when all of the data were combined. The relatively small sample sizes for each species relative to the previous study on which it was designed may be problematic for estimating bridle efficiency.

The model with both day and night observations may not have converged because of the variation introduced by diel period. The models based on day or night observations alone minimize this variation, but also substantially reduce the sample size. An alternative approach would be to include a fixed diel effect. A saturated model with diel period as a predictor variable may better integrate all of the information contained in the full dataset.

Sampling several blocks at the same precise locations may involve depletion in successive samples. If the footrope efficiency estimates are accurate, up to ~30% fewer yellowtail flounder would remain after each tow. Potential depletion should be tested by removing such successive tows at the same location. All of the tows on eastern Georges Bank were in the same direction as the prevailing current, which is not the NEFSC Trawl Survey protocol. Similarly, these stations should be removed from the analysis to investigate the influence of the different protocols.

Acknowledgments

The Panel thanks the field staff, fishermen and analysts involved in the relative efficiency studies. Andrew Lipsky coordinated the review, and background information was provided by Russell Brown and James Weinberg.

Table 1. Catch Ratios of the Chain Sweep to the Rockhopper sweep by taxa by daytime and nightime tows, with 95% bootstrap confidence intervals in parentheses. For the three skate taxa the total catch was weighed but not measured or counted.

Species	Day Number	Day Weight	Night Number	Night Weight
Barndoor skate	-	1.89 (1.61 -2.24)	-	1.12 (0.88 -1.40)
Smooth skate	-	4.83 (3.54 -7.07)	-	3.48 (2.59 -5.19)
Thorny skate	3.32 (2.44 -4.41)	1.82 (1.27 -2.70)	1.49 (1.07 -2.16)	1.06 (0.61 -1.81)
Leucoraja skate	-	4.02 (2.78 -8.23)	-	2.55 (2.35 -2.76)
American Plaice	1.45 (1.33 -1.59)	1.39 (1.30 -1.51)	1.24 (1.05 -1.58)	1.28 (1.06 -1.66)
Summer Flounder	1.66 (1.51 -1.94)	1.61 (1.37 -1.98)	1.49 (1.29 -1.73)	1.36 (1.17 -1.60)
Fourspot Flounder	10.45 (9.08 -12.59)	8.91 (7.65 -10.67)	4.29 (3.86 -4.82)	4.11 (3.68 -4.61)
Yellowtail Flounder	5.17 (4.47 -5.90)	4.92 (4.24 -5.71)	3.09 (2.81 -3.37)	3.05 (2.79 -3.33)
Winter Flounder	1.88 (1.51 -2.18)	1.64 (1.33 -1.92)	1.03 (0.93 -1.15)	1.04 (0.90 -1.24)
Witch Flounder	4.89 (4.32 -5.57)	4.68 (4.18 -5.25)	3.25 (2.78 -3.85)	2.92 (2.58 -3.29)
Windowpane Flounder	8.15 (6.79 -10.09)	7.86 (6.65 -9.37)	2.31 (2.09 -2.58)	2.32 (2.10 -2.59)
Gulf Stream Flounder	10.46 (8.33 -13.34)	9.58 (8.05 -11.62)	3.12 (2.54 -4.16)	3.14 (2.62 -3.92)
Goosefish	3.15 (2.86 -3.47)	2.32 (2.13 -2.55)	3.18 (2.80 -3.61)	2.22 (1.93 -2.60)

Appendix 1. Terms of Reference

1) Is the sampling design and the resulting data collected by the twin trawl studies appropriate to estimate maximum efficiency by the NEFSC Trawl Survey for use in the following operational assessments?

- Cape Cod/Gulf of Maine yellowtail flounder
- Southern New England/Mid-Atlantic yellowtail flounder
- Georges Bank yellowtail flounder
- Georges Bank winter flounder
- Gulf of Maine winter flounder
- Southern New England/Mid-Atlantic winter flounder
- Gulf of Maine/Georges Bank American plaice
- Witch flounder
- Northern windowpane flounder
- Southern windowpane flounder

2) Are the analytical approaches presented in the working papers for witch flounder and yellowtail flounder from the twin trawl studies in 2015 and 2016 appropriate for estimating 1) relative efficiency of the chain sweep and Bigelow rockhopper gear, 2) size effects on efficiency, and 3) diel effects on efficiency for the operational assessments of the stocks listed above? (Note: this TOR is not being broadened to include review of other approaches in the working papers that address diel effects in the annual trawl survey data collected with NOAA ship Henry B. Bigelow rockhopper gear. That topic would require considerable time to review and will not be addressed in detail here.)

3) Should the stock assessments use wing swept area, as done for yellowtail flounder and witch flounder assessments, in the operational assessments for above listed stocks?

Appendix 2. Webinar Agenda

9:30 AM - Introduction & Overview of Process: Russ Brown/Chris Legault, NEFSC Population Dynamics Branch

9:40 AM - Overview of Gear Survey Methodologies: Phil Politis, NEFSC Ecosystem Surveys Branch

9:55 AM - Overview presentations of the Analyses & Approaches: Dave Richardson, NEFSC Oceans & Climate; Tim Miller, Population Dynamics Branch

10:30 AM - Panel Discussion with Principal Investigators

11:30 AM - Adjourn

Appendix 3. Webinar Participants

- Panel Members: Steve Cadrin (Chair), Micah Dean, Greg DeCelles, Jim Gartland
- NEFSC: Larry Alade, Jessica Blaylock, Russ Brown, John Hoey, Chris Legault, Andy Lipsky, Sean Lucey, Tim Miller, Paul Nitschke, Phil Politis, Dave Richardson, Jim Weinberg, Tony Wood
- NEFMC: Calvin Alexander, Jamie Cournane, Libby Etrie, Robin Frede, Chris Kellogg, Tom Nies, Mike Sissenwine,
- GARFO: Emily Keiley, Ryan Silva
- Others: Katie Almeida, Bonnie Brady, Adam Poquette,
- Elizabeth-Last Name not provided
- Susan-Last Name not provided
- Chris-Last Name not provided

Appendix 4. NEFSC Gear Efficiency Research Review

Micah Dean – July 18, 2017

My comments below are focused more on the general appropriateness of the methods, and less on the relevance to the specific stocks listed in the terms of reference, given that we were provided analyses for only two species. In general, my feeling is that the experiments were well designed and executed, and that useful estimates of catchability and area-swept biomass can be constructed from these data. The analyses contained in the working papers provide examples of several different approaches to producing these estimates. I have tried to offer suggestions where alternative approaches could be pursued that may be more appropriate for the available data and focus of the investigation.

Terms of Reference

1. *Is the sampling design and the resulting data collected by the twin trawl studies appropriate to estimate maximum efficiency by the NEFSC Trawl Survey for use in the following operational assessments?*

   The twin trawl experimental design appears well conceived and thorough. A large number of tows were conducted over a variety of habitats/depths/diel periods, creating general relevance to the broader survey datasets. A limiting factor of the available data for a given species may be the number of tows with positive observations, particularly when the data are broken into subsets by diel period. Similarly, the relevance of length bins at the tails of the size distribution should be judged more on the number of positive tows in each bin than on the number of fish observed. Another data consideration will be the spatial heterogeneity of the catches-at-length. For example, if all the large individuals for a given species were caught within a small area, other factors (e.g., substrate, temperature, depth, current) may be influencing the relative chain:rockhopper efficiency of that size class.

2. *Are the analytical approaches presented in the working papers for witch flounder and yellowtail flounder from the twin trawl studies in 2015 and 2016 appropriate for estimating 1) relative efficiency of the chain sweep and Bigelow rockhopper gear, 2) size effects on efficiency, and 3) diel effects on efficiency for the operational assessments of the stocks listed above?*

   The various calculation options of the Richardson working paper are reasonable, straightforward, and serve to demonstrate the sensitivity of the resulting biomass estimates to different analytical choices. Selecting a preferred option for a given species should be based on whether diel period and/or size are considered significant predictors of chain:rockhopper efficiency, and whether there is evidence of <100% efficiency of the chain sweep during the day (i.e., survey $rockhopper_{night}:rockhopper_{day}$ >> $chain_{day}:rockhopper_{day}$; or, $chain_{night}:chain_{day}$ >> 1). Such an evaluation of the influence of each variable on rockhopper efficiency was not addressed by the Richardson paper, but should be included if these methods are to be applied to specific stocks.

   Applying day & night efficiency estimates to individual tows seems preferable to using the proportion of day and night tows to calculate a weighted average efficiency that is then applied to an aggregate index. If a day/night difference in efficiency truly exists, you are essentially collecting information in different units, and it would be inappropriate to average those values together and then apply an overall weighted average efficiency. If areas with higher abundance were sampled more intensively during either night or day, you would want to apply the appropriate efficiency estimate for that period to achieve a more accurate estimate of abundance.

For yellowtail flounder, estimating relative efficiency in 1 cm length bins yielded a somewhat 'noisy' result, with adjacent length bins changing by a factor of 2 or more. This noise gets propagated on to the final biomass calculations, and becomes exaggerated when the survey size distribution differs from that of the twin-trawl experiment. For some species, it may be preferable to aggregate the data into wider length bins to increase the number of paired samples per length bin and reduce this noise. Alternatively, efficiency could be estimated as a function of length, as opposed to assuming independence among length bins.

The efficiency modeling framework of Miller (including the Miller-yellowtail and Hare-witch working papers) is more complex in that it includes multiple random effects, beta-binomial dispersion, and cubic spline smoothed functions for length. This approach and the candidate model set were originally developed to address the multitude of differences between the Albatross and Bigelow survey gears/methodologies (Miller 2013). Simpler models may be more appropriate for estimating the efficiency of the Bigelow net from the twin-trawl data, given the smaller sample size and the sweep being the only substantive difference between gears. Essentially, the goal is to determine the probability that a fish in the trawl path will go above the footrope and enter the net. I would expect this probability to be a monotonically increasing function of length (less likely decreasing), or perhaps a unimodal function. The additional flexibility offered by cubic spline smoothers may be unnecessary and not worth the added parameters and complexity. Complex multimodal length functions seem generally implausible for the process under investigation (trawl footrope escapement) and may be an artifact of the population length distribution when the experiment was conducted. Ideally, length-based efficiency estimates would remain relevant as the population length distribution changes over time. Going forward, I would recommend expanding the candidate model set to include some simple parametric length models (e.g., logistic, log-logistic) to compare against the current 'best' models.

Evaluating the appropriateness of the best efficiency model should also include an examination of the model fit to the raw catch-ratio-at-length observations. Model fit is difficult to interpret from plots of pair-wise and aggregate model fits alone; residuals plots or model fit over observed data would be more useful. A delta-AIC model selection approach is only as good as your candidate model set, and care should be taken to include a range of probable and practical models.

The methods provided in the working papers assume 100% availability of yellowtail and witch flounder to the NEFSC spring and fall surveys. This assumption should be re-visited for each species/stock when calculating area-swept biomass or survey catchability for comparison with assessment models. Although the 95% confidence intervals for Georges Bank yellowtail biomass overlap between the spring and fall (Richardson et al. - Tables 4 and 6), the median values for the fall survey are consistently higher than the spring survey. This suggests either a difference in availability between the two survey cruises, or an additional unaccounted-for variable that may be influencing efficiency (e.g., temperature).

3. *Should the stock assessments use wing swept area, as done for yellowtail flounder and witch flounder assessments, in the operational assessments for above listed stocks?*

   The Politis and Richardson working papers provide a decent review of published studies; their conclusion that wing spread is the appropriate default assumption for flounder species is logical, given the design of the Bigelow survey gear (narrow door spread and minimal contact of ground cables). Unfortunately, the results of the bridle herding experiment are inconclusive for most species. The model conditioned on total catches appeared to perform best and managed to produce a bridle efficiency (Eb) estimate for yellowtail flounder. However, the model failed to converge when applied to night-only catches of yellowtail, despite an ANOVA model finding bridle length to be a significant predictor using the same night-only dataset. It is difficult to explain this contradictory result with the information provided. The experimental design seems well thought out, and a substantial number of tows were conducted. However, the sample size may be insufficient to overcome the variability inherent in an “alternate tows” design, particularly given the expected weak herding effect of the Bigelow bridles/doors. Still, further analyses may yield improved model performance.

   The fact that several blocks and tows within a block occurred at the same precise locations is concerning. If the footrope efficiency estimates are reasonably accurate, then ~30% less yellowtail flounder would be left behind after each tow was made. Depending on the mobility of the fish and time between tows, re-sampling the same locations could have influenced the results of the bridle efficiency experiment. If these tows can be identified and omitted, I suggest re-fitting the model to see if performance improves. Another possibility is to attempt a model with diel period as a fixed effect. The model with both day and night observations may not have converged because of the variation introduced by diel period. The models based on day or night observations alone minimize this variation, but also substantially reduce the sample size. A full model with diel period as a predictor variable may better integrate all of the information contained in the full dataset.

Appendix 5. DeCelles Review Report

Gear Efficiency Research Expert Peer Review Panel

Greg DeCelles
Massachusetts Division of Marine Fisheries
Gregory.DeCelles@state.ma.us

Term of Reference #1 – Is the sampling design and the resulting data collected by the twin trawl studies appropriate to estimate the maximum efficiency by the NEFSC survey trawl for use in the following operational assessments?

- Cape Cod/Gulf of Maine yellowtail flounder, Southern New England/Mid-Atlantic yellowtail flounder, Georges Bank yellowtail flounder
- Georges Bank winter flounder, Gulf of Maine winter flounder, Southern New England/Mid-Atlantic winter flounder
- Gulf of Maine/Georges Bank American plaice
- Witch flounder
- Northern windowpane flounder, Southern windowpane flounder

Yes, overall the sampling design is appropriate for estimating the maximum efficiency of the NEFSC survey trawl for use in the operational assessments. Based on the presentations and technical reports I have reviewed, it is evident that the field work and data analysis were capably performed by NEFSC staff. A major strength of this approach is that the experimental work was carried out in collaboration with an industry vessel.

For witch flounder the experimental approach yielded a robust sample size (n = 53,297), and witch flounder were present in 117 chain sweep tows. These data are appropriate to estimate the efficiency of the Bigelow trawl for witch flounder.

For yellowtail flounder the experimental approach yielded a large sample size (n = 13,040), and yellowtail flounder were present in 121 chain sweep tows. These data are appropriate for estimating maximum net efficiency for all three yellowtail flounder stocks.

For American plaice, 30,306 individuals were sampled in the 2015 and 2016 twin trawl experiment, and plaice were captured in 118 tows made with the chain sweep. This sample size is appropriate to estimate net efficiency for American plaice.

4,926 windowpane flounder were sampled in the twin trawl experiments, and windowpane flounder were present in 110 chain sweep tows. These sample sizes are adequate to estimate net efficiency for both the northern and southern stocks of windowpane flounder. I note that a previous comparative fishing experiment (2009/2010 paired trawl study) also produced robust sample sizes for windowpane flounder. As a sensitivity analysis, I would recommend that the net efficiency estimates from the 2015 and 2016 twin trawl studies be compared to the net efficiency estimates from the 2009/2010 paired trawl study for consideration at the 2017 operational assessment update.

2,212 winter flounder were captured in the twin trawl studies, and winter flounder were only captured in 63 of the chain sweep tows. Based on the catch ratios presented in the summary table, the sample size does not appear to be adequate to estimate length-based net efficiency estimates for winter flounder. However, the data could be pooled across all length classes to generate a single estimate of net efficiency for the three winter flounder stocks. If the data is available prior to the 2017 assessment, I recommend that the data from the planned 2017 twin trawl study in southern New England be considered to increase the sample size for winter flounder. It should also be noted that the 2009/2010 paired trawl experiment yielded very large sample sizes (n > 16,000) for winter flounder. I recommend that the data from the 2009/2010 study be investigated further and that the net efficiency values from both studies be compared and considered during the 2017 operational assessments.

The net efficiency values derived from the twin trawl studies provide an important diagnostic to assess the accuracy of age-structured assessment models. During the 2017 operational assessment updates, I recommend that catchability values that are estimated internally by the age-structured models be compared to the net efficiency estimates from the 2015 and 2016 twin trawl studies. Divergence between the net efficiency estimate and the model estimated catchability may indicate scaling issues with the age structured assessment.

Based on the presentation and summary tables provided by Phil Politis at the June 2017 Northeast Trawl Advisory Panel meeting, the Northeast Fisheries Science Center has a large repository of data available to investigate the efficiency of the Bigelow survey trawl. Since 2009, the Science Center has dedicated considerable time and resources to estimate survey net efficiency, and has partnered with industry vessels to conduct extensive fieldwork throughout the region. Although it is outside of the Terms of Reference for this review, I would encourage the Science Center to further analyze and publish the data from these field experiments, and to disseminate the results from these experiments to regional stakeholders. The summary tables suggest there is a wealth of information related to net efficiency that can be used to inform stock assessments for flatfish, along with other species such as monkfish, red hake, and skates.

The 2015 and 2016 net efficiency studies were conducted throughout Georges Bank, the Gulf of Maine, and southern New England. For the Georges Bank yellowtail flounder empirical assessment, the estimated value of net efficiency was based on observations that were collected across multiple stock areas. This approach assumed that there were no geographic differences in net efficiency, fish behavior, or size structure. For the sake of consistency, I recommend that the same value of net efficiency should be used in the empirical assessment of each stock for a given species. In other words, the empirical approaches for Georges Bank yellowtail flounder and Cape Cod/Gulf of Maine yellowtail flounder should use the same value of net efficiency.

If empirical approaches are developed for more flatfish stocks at the 2017 operational assessments, the analysts will need to decide whether the survey time series should be truncated to only include survey data collected on the Bigelow, or whether the full time series of NEFSC survey data should be included. For witch flounder, the SAW 62 Working Group decided to utilize the entire time series of NEFSC survey data, and the experimentally derived estimate of net efficiency for the Bigelow trawl was multiplied by the Bigelow/Albatross calibration coefficient to estimate the net efficiency of the Albatross. However, when an empirical approach was developed for Georges Bank yellowtail flounder, the TRAC Working Group decided to restrict the NEFSC time series to only include survey data collected on the Bigelow, as a means to reduce the uncertainty associated with the change in survey vessel. Thus, there is no precedent for what range of years should be included in an empirical approach. For the 2017 operational assessments I recommend that empirical approaches should be restricted to include survey data collected from 2009-2017, in order to avoid the uncertainty associated with the change in survey vessel (and the calibration coefficient). I believe that further analysis is needed (on a stock by stock basis) to determine whether it is appropriate to combine the Bigelow net efficiency estimates and the Bigelow/Albatross calibration coefficients to estimate the efficiency of the Albatross survey trawl.

Term of Reference #2 – Are the analytical approaches presented in the working papers for witch flounder and yellowtail flounder from the twin trawl studies in 2015 and 2016 appropriate for estimating: 1) relative efficiency of the chain sweep and Bigelow rockhopper gear, 2) size effects on efficiency, and 3) diel effects on efficiency for the operational assessments of the stocks listed above?

The analytical approaches presented in the Miller et al. Working Paper (binomial/beta-binomial analysis) are not appropriate for estimating the relative efficiency of the chain sweep and Bigelow rockhopper gear. The analytical approach is unnecessarily complicated to address the question at hand. In particular, I am concerned that the models have too many parameters (e.g., too many knots in the smoother) relative to the sample size (# of paired tows). For example, the best model for witch flounder during the day-time tows (Figure 8 in the Hare et. al. 2016 Working Paper) produces a smoothed fit with five inflexion points. From a biological perspective, it is difficult to explain how the behavior of this animal and the characteristics of the trawl gear could produce this type of complex response. I am concerned that the length based relationship would not persist over time if the size structure of the resource were to change. In deference to the principle of parsimony, these complex model fits should be compared to simpler parametric models (e.g., linear or logistic) to determine whether a simpler model provides a comparable or improved fit to the data. In addition, I am concerned about the decision to estimate annual and survey specific length- weight relationships for these species. For species that are at low abundance, it is expected that low numbers of animals would be captured in a single survey, and estimating the length-weight relationship from such a small sample size may unnecessarily introduce considerable uncertainty into the model. Although this uncertainty is considered through bootstrapping, I would suggest either using established length-weight relationships (e.g., Wigley et al, 1994), or using length weight relationships based on multiple years of survey data.

Provided that the sample sizes are sufficient, the analytical approach presented by Richardson et. al. is appropriate for estimating the relative efficiency of the chain sweep and rockhopper gear.

Based on the catch ratio tables it is apparent that there are meaningful size effects on net efficiency for most, if not all, of the flatfish species that are included in the Terms of Reference. Therefore, it is important that length based estimates of net efficiency are calculated. The analytical approach presented by Richardson et al for witch flounder and yellowtail flounder is appropriate for estimating size effects of net efficiency. However, I would suggest that future analyses consider the appropriateness of using larger length bins (e.g., 3 or 5cm) rather than trying to estimate net efficiency for each centimeter length bin. In Table 2 of the Richardson et al Working Paper, it is noted that there are often quite large differences in the catch ratio for yellowtail flounder from one length bin to the next (e.g., median ratio is 4.09 at 40cm and 2.4 at 41cm). Binning the data across larger length bins may reduce some of this variability, produce tighter confidence intervals, and allow the functional form of the length based net efficiency relationship to be more apparent.

Based on the results of the 2015 and 2016 twin trawl studies and earlier comparative fishing experiments (e.g., 2009 and 2010 cookie/rockhopper, 2014 Hera/Bigelow), it is apparent that there are important diel differences in net efficiency for the five flatfish species. These diel differences in net efficiency are not surprising. Previous research has shown that flatfish react differently to trawl gear depending upon ambient light levels, and there are considerable differences in daytime and nighttime catch rates in commercial fisheries for most flatfish. Further, flatfish behavior is known to vary between day and night, with some flatfish (e.g., yellowtail flounder) exhibiting off bottom movements at night. Therefore, I strongly recommend that diel effects be incorporated in estimates of net efficiency for the five flatfish included in the Terms of Reference.

During his webinar presentation (slide 14), Dave Richardson recommended that the approach used to estimate net efficiency for witch flounder (use a length specific approach, separate the day and night twin trawl results and weight based on the proportion of nighttime and daytime tows for that stock) be considered for other stocks. I concur that this approach is appropriate for estimating both size based and diel effects on net efficiency, and I agree with Dave's recommendation that this approach be applied to the other flatfish stocks at the 2017 operational assessments.

Term of Reference #3 – Should the stock assessments use the wing spread to calculate area swept in the operational assessments?

Yes, the area swept by the survey net should be based on the distance between the wings, and not the doorspread. Further work is needed to estimate the length of the bridle that is in contact with the bottom (i.e., W_{off} from Somerton and Munro, 2001) before herding efficiency and effective area swept can be quantified for flatfish.

The published literature suggests that herding by the bridles should be the most effective during the day time. Experienced trawl fishermen also report that flatfish herd in response to commercial trawl gear during the day, but not at night. There is no evidence of a significant herding effect during the day for any of the flatfish species that were examined, and no precise estimates of day time herding efficiency were calculated for the six flatfish species. In addition, herding efficiency was poorly estimated (or could not be estimated at all) for the six flatfish species when all of the data were combined. The relatively small sample sizes for each species (relative to the sample sizes used by Somerton and Munro, 2001) may be problematic for estimating bridle efficiency.

I am also concerned about the decision to make all of the tows on eastern Georges Bank in the same direction as the prevailing current, because this protocol very different than the protocols used on the Bigelow during the trawl survey.

The two model estimates (NLL vs. NLS; Table 5 of the Politis and Miller Working Paper) provide much different efficiency estimates for some species, and in some cases the estimate from one model is outside of the confidence interval of the other model (Table 5 in Politis and Miller). The divergent efficiency estimates between the two models suggest that the model fit is poor.

The survey net was specifically designed to minimize variability due to herding. The grease wear patterns on the bridles, and the orange spray paint applied to the bridles during the 2016 twin trawl study provide evidence that the bridles are not consistently on the bottom, especially near the wing ends. The published literature on fish behavior in response to trawl gear indicates that flatfish will not herd in response to the bridles unless they are struck by the bridle itself. Because there is not consistent bottom contact by the lower bridles, it is unlikely that the bridles are herding flatfish effectively.

Appendix 6. Gartland Review Report

NEFSC Gear Efficiency Research Review

J. Gartland – July 20, 2017

TOR 1 – Is the sampling design and the resulting data collected by the twin trawl studies appropriate to estimate maximum efficiency by the NEFSC Trawl Survey for use in the following operational assessments?

Response – The Northeast Fisheries Science Center (NEFSC) undertook a series of paired-tow experiments in the fall of 2015 and late summer and fall of 2016 in an effort to estimate maximum bounds on catchability, specifically detectability, for a variety of flatfish species. Field surveys were conducted using the *F/V Karen Elizabeth*, a commercial vessel designed to operate a twin-trawl configuration where two bottom trawls of similar or differing design can be towed simultaneously. While yellowtail flounder (*Limanda ferruginea*) and witch flounder (*Glyptocephalus cynoglossus*) were the primary target species, data on a number of other flatfish species were collected as well.

The field experiments meant to quantify the maximum catchability of these flatfishes were well-designed and appropriate to address the research objectives. The use of a twin-trawl approach to quantify relative catchability of a gear of interest (e.g., NEFSC survey trawl) has precedent in the literature and represents the most effective way to control for a multitude of confounding variables. By towing both nets from the same vessel simultaneously, concerns regarding the influence of fishing power, captain effects, and spatiotemporal patchiness of the target animals, among other considerations, were eliminated.

The seasonal timing of these experiments matched that of the fall bottom trawl surveys (BTS) fairly well, and the incorporation of sampling during both the day and night strengthened the subsequent analyses and applicability of the catchability estimates to the NEFSC BTS data. Sampling was conducted over a wide spatial extent, occurring in Southern New England and Georges Bank in 2015 and again on Georges Bank and in the Gulf of Maine in 2016, effectively covering much of the distributional ranges of a number of flatfish stocks. Sites sampled within each of these regions were selected based on perceived high abundances of target animals to maximize data yield, which is reasonable given the limited resources available to these experiments. Such an approach may have introduced some bias, however, if the catchabilities of the survey trawl and benchmark gear were differentially impacted by changes in flatfish density.

The investigators sampled as broad of a range of bottom types as possible, given the limitations imposed by the sweep design of the benchmark (i.e., chain sweep) net, and subsequent analyses indicated that maximum estimates of potential bias due to the avoidance of hard-bottom areas were only on the order of ~7% for yellowtail flounder. Estimates of habitat bias were not provided for other species, but would be a valuable addition should these results be considered for an Operational Assessment. The depth ranges sampled were reasonable, and likely covered an appreciable portion of the depth ranges of the target species. While sampling between 153m and 366m (maximum depth of NEFSC survey sampling) did not occur, perhaps it would be possible to conduct an analyses similar to the bottom-type analysis described above to estimate the magnitude of potential bias due to the truncation of the range of sampling depths.

The design of the benchmark gear was appropriate for the development of estimates of the upper bounds on flatfish catchability by the NEFSC survey trawl. Chain sweep designs maintain the bolschline/footrope in close proximity with the bottom, particularly when compared with the survey 40.64cm/35.56cm rockhopper sweep. The decision to remove flotation from the benchmark net to maintain consistent bottom contact for this gear was appropriate. Given the well-documented behavior of many flatfishes in the mouth of a trawl, the use of restrictor lines likely was not a source of bias in the experiment, and were valuable tools in the maintenance of consistent and constant wingspread for both nets over a range of depths.

Overall, these experiments were very well designed and executed, and are acceptable for the estimation of an upper bound on catchability for those flatfishes listed as such below. The deficiencies in the field sampling design and execution noted above were relatively minor, and likely had very little (if any) impact on the results or their applicability. The potential sources of bias identified above likely would only serve to inflate the reported estimates of catchability for the survey gear, and therefore do not represent fatal flaws as the objective was to generate estimates of maximum catchability. Further, the decision to conduct this research in a collaborative manner with an industry platform is laudable, as it leverages industry expertise in the location of the target animals and in the fishing operations, which in turn yields greater confidence in and acceptance of the results across a broad range of user groups and stakeholders.

Cape Cod/Gulf of Maine Yellowtail Flounder

The catchability experiments conducted by the NEFSC in 2015 were specifically designed to target yellowtail flounder, and additional data were collected in 2016. While the majority of the effort was focused on Georges Bank over those two years, a number of tows were conducted in the Gulf of Maine as well. Because catchability is thought to vary by size and between the day and night for flatfish species, maximum estimates of catchability were quantified by each of these variables for yellowtail. Sample sizes for individuals greater than 30 cm were relatively large for both day and night samples. Yellowtail were collected on a majority of paired tows. Although fewer fish were captured at smaller sizes, and confidence intervals for catchability estimates were wide for daylight tows, these juvenile fish have a smaller impact on total stock biomass. As such, the estimates of maximum catchability for yellowtail flounder are acceptable for use in the operational assessment, including for the Cape Cod/Gulf of Maine stock.

Southern New England/Mid-Atlantic Yellowtail Flounder

As noted above, yellowtail flounder was a primary focus of the 2015 and 2016 catchability studies. Although comparatively fewer tows were made in Southern New England, habitat types are similar to those sampled on Georges Bank, and as such the yellowtail catchability values are believed to be acceptable for use for this stock as well.

Georges Bank Yellowtail Flounder

Estimating an upper bound on catchability for the Georges Bank stock of yellowtail flounder was the primary focus of the field experiments conducted in 2015. A large number of tows were conducted on

Georges Bank in 2015 and 2016 as a result, and samples sizes were relatively large for most size categories for both the day and night tows. Catchability estimates are acceptable for use in the Operational Assessment for this stock.

Winter Flounder (Gulf of Maine, Georges Bank, Southern New England/Mid-Atlantic Stocks)
Winter flounder (*Pseudopleuronectes americanus*) were not a target species for the 2015 and 2016 experimental field studies. Accordingly, this species was caught on relatively few tows, and overall sample sizes were relatively low, as were sample sizes within each size category for both the day and night. Size-specific estimates of catchability varied widely for each time period, and confidence intervals were rather large. As such, the recommendation is that these estimates of maximum catchability are not acceptable for use in operational assessments.

Gulf of Maine/Georges Bank American Plaice
While American place (*Hippoglossoides platessoides*) were not targeted by these field experiments, a large number of animals representing a broad size range were collected over the two-year period. This species was also collected on a majority of the paired tows. Estimates of maximum catchability were stable across size categories, and the ranges of the associated confidence intervals were relatively small. These estimates of maximum catchability are acceptable for use in the Operational Assessments.

Witch Flounder
Witch flounder was the primary target species for the 2016 field experiments. Sample sizes are very large across a broad size range of animals, estimates of maximum catchability are stable across size classes, and associated confidence intervals are reasonable. These catchability estimates are acceptable for use in the Operational Assessments.

Windowpane Flounder (Northern and Southern)
Windowpane flounder (*Scophthalmus aquosus*) were not targeted by the field sampling efforts in 2015 or 2016. Although sample sizes were somewhat reasonable, estimates of maximum catchability spanned a large range, particularly during the day, and confidence intervals were wide. As a result, the daytime catchability estimates for this species are not acceptable for use in the Operational Assessments. The nighttime estimates are more stable with reasonable confidence intervals, however, and are acceptable. Perhaps reliable daytime estimates could be attained assuming 100% night efficiency for the benchmark gear, coupled with the day:night ratio for windowpane derived from the survey.

TOR 2 – Are the analytical approaches presented in the working papers for witch flounder and yellowtail flounder from the twin trawl studies in 2015 and 2016 appropriate for estimating 1) relative efficiency of the chain sweep and Bigelow rockhopper gear, 2) size effects on efficiency, and 3) diel effects on efficiency for the operational assessments of the stocks listed above?

The analytical approaches used to develop maximum estimates of catchability for flatfish species included both a model-based approach (i.e., binomial and beta-binomial models) that accounted for within and between pair variation as well as a more direct method of estimation that involved bootstrapping. The application of both methods appears to have been well thought out and yielded similar estimates of maximum catchability for witch flounder, which serves to validate the appropriateness of each and increases confidence in the reliability of the results.

The models used to estimate maximum catchability of these flatfish species have been applied to paired-tow data in the past and represent an accepted method by which to quantify the relative catchability of two gears. The approach used to identify the most appropriate model from among the candidate forms appears to be reasonable. While ability of these models to characterize maximum catchability for smaller-sized individuals was rather limited, likely due to the small sample sizes available for the juvenile stages, catchability estimates somewhat stabilized and confidence intervals tightened at larger sizes. Based on the results of the witch flounder work, the bootstrapping approach seemed to follow similar trends.

Given that the results of the model-based and bootstrapping approaches were comparable, but that the latter represented a computationally simpler, more intuitive approach, it is recommended that the bootstrapping approach be used to generate estimates of maximum catchability for use in the Operational Assessments of these flatfishes. It follows, then, that the bootstrapping approaches outlined in the yellowtail and witch flounder working papers are appropriate for estimating the relative efficiency of the chain sweep and the Bigelow rockhopper gear. Further, the general analytical recommendations outlined in the presentation of the bootstrapping methodology should be favored over those given in the working paper that outlined the modelling approach.

Previous investigations have found length-specific and diel differences in flatfish catchability, and the results of the field experiments conducted in 2015 and 2016 are consistent with these findings. As such, the decision to account for size and day/night variability in catchability was appropriate. The lower and upper size length bins designated for the four species with acceptable catchability estimates (i.e., yellowtail flounder, witch flounder, American plaice, windowpane flounder [night] – see above) are reasonable given the smaller sample sizes of the juvenile and the largest individuals.

Differences in catchability of these flatfishes between daylight and night tows was apparent and properly quantified. For yellowtail and witch flounders, the assumption was made that the catchability of the benchmark (chain) net in both the day and night was 100%, which is reasonable given that the day:night ratio of benchmark-to-survey trawl conversion factors was similar to the day:night ratio of survey tows. For those species where the aforementioned relationship does not hold, it is recommended that the maximum night catchability of the survey trawl be quantified using paired tows conducted at nighttime and assuming that the benchmark net is 100% efficient, while maximum day catchability for the survey trawl be estimated using the nighttime catchability and the survey day:night ratio. Again, the decision as to whether to assume that the benchmark net is 100% efficient in both day and night or only during the night should be species-specific. This recommendation varies from that given in the presentation on the bootstrapping approach.

For each species, a single estimate of catchability was generated by weighting day and night estimates of catchability by the proportion of survey tows conducted during each time period from 2009-2016, rather than developing annual estimates of catchability by calculating the proportion of day and night tows each year. This approach is reasonable given that the proportion of tows occurring during each time period was approximately constant across years. It would be advisable, however, to generate an annual estimate of catchability should a situation arise where the timing of the survey changes appreciably for some reason such that the proportion of tows conducted during the day and at night differs appreciably from the average condition. This recommendation is a slight modification of that put forward in the bootstrapping presentation.

TOR 3 – Should the stock assessments use wing swept area, as done for yellowtail flounder and witch flounder assessments, in the operational assessments for above listed stocks?

Response – Assessments of these flatfish stocks should use net wingspread instead of doorspread when calculating swept area. The background information provided in the working paper on the field experiments designed to quantify bridle herding correctly pointed out that the industry representatives, along with the academic and federal scientists comprising the first Mid-Atlantic/New England Trawl Survey Advisory Panel, developed the current survey trawl with the intention of minimizing herding between the net and doors. It also would be valuable to note that representatives of three net manufacturing businesses in New England worked together to calculate the minimum bridle length necessary to achieve the optimal net spread and opening, in an effort to avoid herding, and that this length (i.e., 20 fathom) was chosen as the standard bridle length for this survey.

In addition to relying on the expert advice and calculations discussed above, field-based observations indicated that herding by the bottom bridles on the survey trawl is at most very minimal, and likely nonexistent. Field experiments designed to quantify bridle herding through the use increasing bridle length were conducted in 2014, and results showed no clear relationship between catch and bridle length for each of the flatfishes considered. Further, still images provided by underwater video equipment showed that the bottom bridle is not in contact with the seafloor for at least 2m forward of the bunt bobbin. The behavior of flatfishes in the presence of a trawl is well-documented in the literature, and based on these studies it is likely that any flatfishes that are herded toward the trawl by the forward sections of the bottom bridle escape beneath the bridle in the zone ahead of the bobbin. While images were only available from the aft portion of the bottom bridle, observations of the wear patterns of this wire, made during the 2014 bridle efficiency experiments and 2015/2016 twin-trawl investigations, indicated that bottom contact in the forward sections of this bridle is minimal and inconsistent, even at lengths that are greater than twice that of the standard design. Indeed, to ensure that bridles maintain consistent and continuous bottom contact when targeting flatfishes in industry applications, groundgear is typically covered with rubber disks (i.e., cookies) and interspaced leads and lengths of chain are attached at both terminal ends of the bridle wire. It is unlikely then that bare wire yields any appreciable herding effect, particularly at the length used by the survey trawl.

Overall, it is apparent that bridle-herding for the NEFSC survey trawl is very unlikely, and that stock assessment efforts for these flatfishes should use wing swept area as opposed to door swept area.