Independent Panel Review
of the NMFS Vessel Calibration Analyses for
FSV Henry B Bigelow* and *R/V Albatross IV
August 11-14, 2009
Chair's Consensus Report

1.0 Introduction

The panel was convened to provide a review of the calibration analyses carried out on the data collected during the 2008 pair tow experiment involving the *FSV Henry B Bigelow* and *R/V Albatross IV*. The terms of reference (Appendix 1) were to review and evaluate the suite of statistical methods used to derive calibration factors by species before they are applied in a stock assessment context.

The panel carried out this review and addressed all items in the terms of reference for which analyses were available and where time permitted. The review was based on several working papers prepared by the NMFS staff and the presentations given by them on days 1 and 2.

The panel acknowledges the comprehensive testing, planning, and standardization of trawl gears and vessels, and field data collection conducted in the preceding years leading up to and including the 2008 calibration experiments.

The panel further acknowledges the comprehensive suite of rigorous statistical methods employed to estimate the calibration factors for the 636 paired tows by the R/Vs Albatross IV and Henry B. Bigelow. Below is the panel's review, prepared with full agreement by the entire panel, of the analyses used in the derivation of calibration estimators.

2.0 Synthesis

The panel reviewed the calibration statistics of both Ratio and Beta-binomial estimators along with individual length frequencies (where available) and plots of empirical and beta-binomial based log-p estimators at length for several (18) species, but not all of the 38 species on the FMPs list. The objective was to develop specific protocols for guidance in the selection and use of appropriate estimators based upon the amount of data available and the relative performance of the two candidate estimators. The protocols were then tested by the panel on 2 candidate species to see how well they performed.

2.1 Protocols

2.1.1 In the case where **Ratio estimator = Beta-binomial** estimator or where there was little difference in both estimators based upon the confidence intervals of the ratio estimates, the panel recommends the use of the Beta-binomial estimator. This estimator exhibited superior performance in simulation studies. However, the panel further recommends that length should be added to the Beta-binomial model as a continuous covariate to incorporate length based conversions where appropriate.

Example 1: cases where length is influencing the relationship such as: yellowtail flounder, American plaice, and winter flounder.

Example 2: cases where length and season are influencing the relationship such as: window pane and summer flounders

Example 3: cases where length and season are not influencing the relationship such as: bluefish, butterfish, haddock, black sea bass, white hake and winter skate[1]

2.1.2 In the case where **Ratio estimator >Beta-binomial estimator,** typical of situations where large catches were primarily taken by the Bigelow because of patchiness/density characteristic (spatial heterogeneity) of some schooling pelagic fish the panel recommends using the **Beta-binomial estimator** because it is less influenced by a few large catches by one vessel (i.e., a large catch is treated like a small catch) when compared to the ratio estimator. Examples are silver hake (also influenced by season), herring and spiny dogfish

2.1.3 In the case where **Ratio estimator < Beta-Binomial estimator,** typical of situations where both vessels take large catches at the same time and the Bigelow is also taking a lot of small catches that the Albatross is missing entirely. Such situations are over-influencing the Beta-binomial estimator. Examples are: Little skate (causing a seasonal difference), scup (no small fish less than 5 cm in fall compared to spring, i.e., spring-spatial heterogeneity in catches) and redfish (only a marginal difference). Here the panel recommends using

[1] There may be an identification issue between small winter and little skates which are influencing the numbers in the catches.

the ratio estimator because large catches contain more information about calibration factor.

2.2 General protocol guidelines:

2.2.1 If there are less than 30 ++ total occurrences do not attempt any conversion.

2.2.2 If there are less than 30 ++ in any one seasonal occurrences in the catches seasonal conversion are not appropriate.

2.2.3 If there are 31-50 ++ occurrences in any one season consider a conversion only if one is required but proceed with caution.

2.2.4 If the catches by both vessels were very low then any derived conversion factor would be an unreliable estimate. Examples are: pollock (driven by one large catch), striped bass, Atlantic halibut, Atlantic hagfish, cusk, and Atlantic wolffish.

3.0 Review of the Terms of Reference:
Evaluate the methodology for estimation of conversion factors for catch rates between the FRV Henry Bigelow and NOAA Ship Albatross IV in terms of:

ToR (a.) Statistical appropriateness

- What constitutes a sufficient estimate of calibration effects in terms of precision, bias and other properties?
- (See also g.) below.)

Panel Response
The performance of the estimators was tested in simulation studies and analyses of the calibration data. The panel used these results among others (see Synthesis report above) to develop protocols for integrating the data to respond to ToR "g" below:

ToR (b.) Number of treatment effects to be considered (e.g., time of day, depth)

- Are region-specific estimates feasible and/or necessary?

Panel Response
No information/analyses were presented at the workshop for evaluation.

ToR (c.) Evaluation of calibration implications (if any) of paired tows collected as part of the shadow survey with those based on regional site specific stations.

Panel Response
It was difficult to evaluate these regional site specific stations because they were not always treated separately from the shadow surveys in the analyses presented, e.g. calibration estimates, and the fall shadow and November site length compositions. The panel agreed with the concern expressed by NMFS that site specific stations sampled in June might differ from the spring survey. However, the panel felt that site specific stations should have been separated from survey stations for both spring and fall in the analyses to provide more observations on how the estimators were performing.

> **ToR (d.)** Treatment of matched tows and performance of alternative estimators when one vessel catches a given species but the other does not. (Consider application of zero-inflated, and other mixture distributions for estimation) (See also ToR g. below.)

Panel Response
The panel recommends additional evaluation of model representation of zero frequencies through the use of simulation based model assessment. This involves the use of parametric bootstrap procedures to generate a reference distribution against which to compare the frequencies of observed and expected zero observations under the assumed calibration model.

> **ToR (e.)** Performance of alternative estimators for species with low encounter rates and/or groups of species with potentially similar catchabilities (e.g. flounders, gadids, etc.)

Panel Response
A potential strategy for further methodological development (hierarchical framework for model development and validation) was suggested that could address these issues. Due to shortage of time NMFS directed the panel to provide immediate advice on the use of calibration factors in upcoming fall assessments. However, the panel still recommends that the hierarchical approach be considered in post workshop analyses.

> **ToR (f.)** Estimators of length-specific conversion factors
> - What are appropriate criteria for application?

Panel Response
This issue is addressed in the synthesis report of the panel above.

ToR (g.) For each estimator, develop measures of uncertainty and advise on limits of applicability

i. For which species are there insufficient data for any calibration?
ii. For species to be assessed this fall (butterfish, spiny dogfish), and that have typically relied on spring survey indices, what short-term solutions should be implemented to use the spring 2009 data collected by the Bigelow?
iii. For which species is the current proposed methodology adequate?
iv. For which species does the proposed methodology require adjustment, and what is required?
v. What approaches are appropriate to deal with species or groups with insufficient information: ignore the potential difference, use a mixed category approach, or other approaches?
vi. Recommend approaches for dealing with uncertainty in back-transformations from Bigelow values to Albatross "equivalents." Is a Taylor series expansion appropriate? (This will be most relevant for assessment applications in the next 10 or so years.)

Panel Response
These issues, other than item *vi*, are addressed in detail in the protocols listed in the synthesis report above. Time was not available to address item vi.

ToR (h.) Develop recommendations for ongoing research to improve estimation for specific species groups (e.g. flatfish, pelagics) and potential effects of bottom type.

Panel Response
Time was not available to address this issue.

4.0 Other analysis not in the ToR but presented at the workshop

Analysis of age frequency and size at age data.

Panel Response
The panel suggests that the use of a multinomial model would be appropriate for testing of age frequencies differences between the two vessels using

proportions in the analysis. In addition an equivalence testing model should be used to examine differences in mean length at age between the two vessels. Especially useful when small sample sizes increase the probably of accepting the null hypothesis of no difference. The equivalence model tests the null hypothesis that the mean length at age from the two vessels are different; by rejecting it then conclude the alternate hypothesis that they are the same.

Independent Panel Review
of the NMFS Vessel Calibration Analyses for
FSV Henry B Bigelow* and *R/V Albatross IV
August 11-14, 2009
Chair's Consensus Report

Appendix I: Terms of Reference for Vessel Calibration Analysis Review 11-13 August 2009.

Background: In the spring of 2009, the FRV Henry B. Bigelow replaced the NOAA Ship Albatross IV as the primary vessel platform for conducting the Northeast Fisheries Science Center's (NEFSC) research bottom trawl surveys. In addition to the change in research vessel platforms, the Center also implemented a new fishing system designed in conjunction with fishery stakeholders, and made key changes to survey protocols including tow duration and towing speed. In preparation, Center staff designed a series of experiments to estimate conversion factors for catch rates between the two vessels. The experimental designs were reviewed by a panel of experts 25-27 April 2007, and experiments were conducted in the Spring and Autumn of 2008. Details are provided in NEFSC CRD 07-12 Proposed Vessel Calibration Studies for NOAA Ship Henry B. Bigelow (August 2007). Initial logistical constraints imposed by the limited overlap in time of service by the two vessels were considered in the experimental design; but additional logistical constraints emerged as the experiments proceeded, which affected the type and amount of data originally anticipated. Data collected through the experiments are currently being analyzed and a suite of statistical methods are being developed and implemented. Review of the analytic products is required before final conversion factors are applied in a stock assessment context.

Terms of Reference:

Evaluate the methodology for estimation of conversion factors for catch rates between the FRV Henry Bigelow and NOAA Ship Albatross IV in terms of

a.) Statistical appropriateness

- What constitutes a sufficient estimate of calibration effects in terms of precision, bias and other properties?
- (See also g.) below.)

b.) Number of treatment effects to be considered (e.g., time of day, depth)

- Are region-specific estimates feasible and/or necessary?

c.) Evaluation of calibration implications (if any) of paired tows collected as part of the shadow survey with those based on regional site specific stations.

d.) Treatment of matched tows and performance of alternative estimators when one vessel catches a given species but the other does not. (Consider application of zero-inflated, and other mixture distributions for estimation) (See also g.) below.)

e.) Performance of alternative estimators for species with low encounter rates and/or groups of species with potentially similar catchabilities (e.g. flounders, gadids, etc.)

f.) Estimators of length-specific conversion factors

- What are appropriate criteria for application?

g.) For each estimator, develop measures of uncertainty and advise on limits of applicability

- For which species are there insufficient data for any calibration?
- For species to be assessed this fall (butterfish, spiny dogfish), and that have typically relied on spring survey indices, what short-term solutions should be implemented to use the spring 2009 data collected by the Bigelow?
- For which species is the current proposed methodology adequate?
- For which species does the proposed methodology require adjustment, and what is required?
- What approaches are appropriate to deal with species or groups with insufficient information: ignore the potential difference, use a mixed category approach, or other approaches?
- Recommend approaches for dealing with uncertainty in back-transformations from Bigelow values to Albatross "equivalents." Is a Taylor series expansion appropriate? (This will be most relevant for assessment applications in the next 10 or so years.)

h.) Develop recommendations for ongoing research to improve estimation for specific species groups (eg flatfish, pelagics) and potential effects of bottom type.

If time becomes limiting, priority consideration should be given to species managed under Fishery Mangement Plans (FMPs) where NEFSC trawl survey data are included in stock assessments.