Independent Peer Review Report on the

60th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Scup and Bluefish

Prepared for:
The Center for Independent Experts

Dr Kevin Stokes
Stokes.net.nz Ltd
59 Jubilee Rd
Khandallah
Wellington 6035
New Zealand
Ph: +64 (04) 973 7305
E-mail: kevin@stokes.net.nz

EXECUTIVE SUMMARY

The 60th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC) meeting took place at the NEFSC, Woods Hole, MA, from 2nd to 5th June 2015. The review was hosted by NEFSC. Two stocks were considered at SARC 60: Scup (*Stenotomus chrysops*) and Bluefish (*Pomatomus saltatrix*).

Time was limited for in-depth consideration of all aspects pertaining to two distinct stocks but the process was well-ordered and the Panel was able to complete its tasks with support from NEFSC staff and all SAW representatives. The process allowed sufficient opportunity for public input.

Scup

The SAW has made a thorough job of compiling and processing data as inputs to the assessment model. These have been applied methodically and carefully when fitting the existing model to new data and in development of a new benchmark assessment. Model development has been systematic and thorough. The resulting assessment could be explored further, particularly with respect to time-series length and selectivity assumptions, but it is unlikely that any modified model would provide fundamentally different perspectives on stock status or prognosis. Overall, the model, reference point definition, stock status determination, and advice implicit in projections are reliable and robust. In my view, the SAW has provided scientific assessments which are adequate to serve as a basis for developing fishery management advice.

Bluefish

The SAW has made a thorough job of compiling and processing data as inputs to the assessment model. These have been applied methodically and carefully when fitting the existing model to new data and in development of a new benchmark assessment. Model development has been systematic and thorough. The new model is a major improvement over the existing benchmark, bringing in new data sources, disaggregating data, attending to developments in best practice, and exploring more thoroughly selectivity patterns. Overall, the model, reference point definition, stock status determination, and advice implicit in projections are reliable and robust. (Note that at the time of this report, the final projections have yet to be made available.) In my view, the SAW has provided scientific assessments which are adequate to serve as a basis for developing fishery management advice.

BACKGROUND

Scup

Scup (*Stenotomus chrysops*) is a schooling, shelf species that grows to a maximum size of over 40cm and reaches a maximum age of about 15 years. Full maturity is reached by age 3 with more than 50% mature at age 2. Scup are a prey item for a large number of species, including a range of commercially important species.

While tagging and morphometric studies have suggested the possibility of more than one stock, for the purposes of assessment, scup have traditionally been considered as a single stock in US waters, ranging from Cape Cod in the north southward to Cape Hatteras.

Scup are caught commercially, primarily in trawls, and recreationally, both privately and from for-hire vessels. Historically, total catches have ranged from 30,000mt or greater in the 1960s to a low of near 3,000mt in the late 1990s, at which time a recovery plan was implemented. Current annual catches are of the order of 10,000mt. Scup are managed jointly by MAFMC and ASMFC under a FMP. The FMP established a recovery plan for scup with intended (and achieved) major reductions in fishing mortality during the period 1997-1999, attained through a variety of management mechanisms including commercial quota constraints, recreational limits, and various input controls. In 2000, Gear Restricted Areas (GRA) were additionally introduced to reduce scup discards in squid and silver hake fisheries. The Recovery Plan was intended to ensure the scup stock was rebuilt to the biomass target by 2015, but the stock assessment in 2008 indicated the stock was neither overfished nor experiencing overfishing in 2007.

Scup was last assessed in 2008 using a statistical catch at age (SCAA) model implemented in the ASAP program. The assessment was conducted as part of the 2008 Northeast Data Poor Stocks Working Group (DPSWG) and Peer review process. The 2008 benchmark, accepted by the DPSWG, was updated in 2011 using the same model configuration. The assessment indicated an F of 0.034 in 2011, well below the accepted (in 2008) FMSY proxy pf F40%SPR = 0.177, and SSB of 190,424 mt, well above the accepted SSBMSY proxy of SSB40%SPR = 92,044 mt.

The 2015 stock assessment is a new benchmark with opportunity to consider new data, data updates, model configuration, and potentially biological reference points.

Bluefish

Bluefish (*Pomatomus saltatrix*) is a globally distributed pelagic species found in inshore and offshore waters off the east coast of the United States from Florida to Maine. Bluefish live to slightly over 15 years of age and grow and mature very early, with over 30% of maximum length and 50% maturity for both sexes at just over 1 year old. Younger fish are found predominantly inshore while larger fish are primarily offshore.

Bluefish are widely distributed in temperate and tropical waters and extend south from Florida in to the Gulf Mexico. There is no evidence of stock structure within the area managed by the ASFMC and MAFMC, and for stock management and assessment purposes the stock is defined as that portion occurring along the Atlantic coast from Maine to Florida.

Bluefish are caught both commercially in a wide variety of gears and recreationally, both privately and from for-hire vessels, with the large majority of the catch being taken recreationally during May to October. Bluefish is a highly regarded and sought after recreational species. Historically, commercial catches have ranged from more than 7,000 mt in 1983 to a recent low of less than 2,000 mt in 2013. Recreational catches have averaged over 14,000 mt since 1981 though catches in the last decade have been smaller with recent catches *circa* 5,000 mt per year.

Bluefish is managed jointly by MAFMC and ASMFC under a FMP. The FMP lays out a clear allocation between recreational and commercial sectors and for state-specific quotas. Various amendments to the FMP have progressively tightened and clarified management measures as well as setting monitoring requirements.

The current statistical catch at age (SCAA) model, implemented in the ASAP program, was accepted for bluefish at SARC 41 in 2005 following extensive consideration of alternatives. MSY-based reference points were also adopted at that time, using Thompson-Bell YPR as a basis for estimating FMSY and the Shepherd-Sissenwine approach to estimate BMSY. SARC 41 concluded that bluefish were neither overfished nor experiencing overfishing.

The 2015 stock assessment is a new benchmark with opportunity to consider new data, data updates, model configuration, and potentially biological reference points.

REVIEW PROCESS

The 60th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC), considering benchmark stock assessments for Scup and Bluefish, took place at the NEFSC, Woods Hole, MA, from 2nd to 5th June 2015. The review was hosted by NEFSC.

The SARC Review Panel comprised an appointed Chair (Jones; MAFMC SSC) and three CIE reviewers (Hall, Kupschus, and Stokes). Rapporteurs for all sessions were drawn from the NEFSC Population Dynamics Team. The Panel was tasked with providing separate SARC Reports for Scup and Bluefish. The Panel was also tasked to work with the Stock Assessment Working Groups (*via* the Chairs, Lead Analysts, and other members present) to develop an agreed Summary Report for each stock. CIE participants are further tasked with providing independent reports (of which this is one).

The SARC 60 review meeting included numerous staff from the NEFSC, a few Council and State management agency staff and limited academic participants (see Appendix 3). A written perspective on the Scup fishery and resource was received from commercial fishermen (Lapp, Almeida and Cadrin, 2015). I am not aware of any problems with notification of the meetings. All participants were able to participate throughout the meeting according to recently revised rules of procedure circulated in advance and opportunity was explicitly given for public comment at each session, in the room and *via* a constantly running conference link. Many non-Panel participants contributed usefully to discussion, and I believe that all were provided appropriate opportunity for involvement both during the Panel meeting and during extra-mural discussions.

Notification of the meeting and dissemination of papers followed closely the schedule laid out in the CIE Statement of Work (see Appendix 2). For both Scup and Bluefish, main assessment reports and draft summaries, together with past reviews and some background papers, were provided in advance *via* a dedicated web link (see Appendix 1). No presentations were made available in advance. During the meeting an internal file server was used and IT support was provided. Overall, administration of the review was sound. Other regional review processes use different methods for making information available. The use of ftp is common and it is not unusual for materials to be made available at least to reviewers also using, e.g., Dropbox, Google Drive, or similar cloud products. All of these methods allow much simpler synchronization of materials before, during, and after meetings. The web link method used by SARC was frustratingly cumbersome and the file server did not allow use of synchronization tools (e.g. WinSCP). I would suggest that SARC considers the use of ftp or cloud-based alternatives.

The scope of the SARC is stated as: *The Northeast Regional SARC meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC peer review is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development and report preparation (which is done by SAW Working Groups or ASMFC technical committees),*

assessment peer review (by the SARC), public presentations, and document publication. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fisheries within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO).

In my view, the critical aspect of this is that the *[the] review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice.* Given this background, I consider it important that the review process should not try to serve as an in-depth science process *per se*, or to attempt to make every assessment perfect. Adequacy as input to advisory processes is all important, with the key questions pertaining to robustness and reliability. Would advice change or management be different if the assessments were changed?

The SARC review allows four days to consider all issues relevant to two stocks, agree stock summaries and draft SARC Panel reports. The agenda at SARC 60 allowed 4 hours for initial scup presentations and 3:45 for bluefish. No presentations were available in advance. The scup presentations on data and the model summed to 214 slides, while for bluefish the combined data and model presentation was 293 slides. There was an additional 43-slide presentation relevant to bluefish ToR3, considering the effects of temperature on bluefish distributions and availability to surveys. Overall, there was approximately 1 minute per scup slide and just 40 seconds per bluefish slide. In general, doing a review with so much material in so little time is difficult for presenters and reviewers alike. It can be frustrating. It is therefore essential to keep in mind the critical issue of scope noted above. Only so much delving is necessary. The question of adequacy is paramount.

The Terms of Reference (ToR) for the review are given in Appendix 2, Annex 2. Often, reviews including CIE experts focus on a particular phase of the stock assessment process – either the data inputs or the assessment *per se*, and often deal only with a single stock. The ToR set for the SAW/SARC 60 review is very wide, spanning for each of two stocks, data quality (including collection and analysis), the stock assessment, status advice, projections, and research recommendations. I have noted in previous CIE reports that it was not always possible to devote as much time as would be desirable to every issue area when two stocks and wide ranging issues are to be considered. I do not make any recommendations, but reiterate that focus on one stock or just the modelling might lead to a better consideration of reliability, robustness, and adequacy.

REVIEWER'S ROLE IN THE REVIEW ACTIVITIES

The role of the reviewer is set out in the CIE Statement of Work, Attachment A, attached here in Appendix 2, Attachment A. CIE reviewers are tasked with producing an independent report to the CIE. The reviewers are additionally tasked with writing the SARC Panel Reports for each of Scup and Bluefish. The reviewers are further tasked, together with the SAW representatives present, to develop the SAW Summary reports for each stock.

In addition to *becoming familiar with the draft stock assessments(s) and background materials*, I participated in all discussions and contributed to the SARC reports (leading on ToR 5, 6, and 7 for each stock). Both SARC reports were drafted reasonably quickly, but the bluefish report was left incomplete due to the need for revised projections and stock status summaries. The three CIE reviewers worked collaboratively by e-mail to ensure the SARC reports were finalized; this was complicated by the reviewers and chair working across three time zones spanning 17 hours. My edits on the draft SARC reports were sent to the Panel by 13th June (EST) and all CIE reviewer comments were completed by 14th June (EST), with the Chair responsible for final editing. This was effectively completed on the timescale of this report though some revised runs for bluefish are still outstanding. During the meeting, along with other Panel members, I participated in development of the SAW Summary reports. Both were completed with relative ease but, as for the SARC report, the final bluefish Summary still required revised projections at the end of the Review meeting. I have not seen any new materials via the web link, so am unsure of their status.

FINDINGS BY STOCK AND ToR

Scup

1. Estimate catch from all sources including landings and discards. Include recreational discards, as appropriate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

In my view, the SAW thoroughly and carefully considered commercial and recreational landings and discard data, describing all sources and potential issues related to building the assessment model inputs. The SAW comprehensively described the spatial and temporal distribution of landings, discards, and effort and commented on uncertainties. There is little to be added to what has been reported by the SARC Panel, itself reflecting the SAW Report.

Commercial Fisheries Data were obtained from the standard sources and the SAW commented on potential sources of historical bias in e.g. landings records and CPUE/LPUE, the latter due to reporting only positive scup trips (and hence ignoring effective scup effort when none were caught). The SAW dealt with the latter by an arbitrary, but standard and reasonable inclusion only of trips catching more than a given percentage (75%) of scup.

Recreational landings were similarly obtained from the standard for-hire and angler sources and the SAW developed model inputs using standard approaches given known issues in the MRFFS and MRIP data collection schema. As noted by the SARC Panel, scup-specific issues related to temporal variability in distribution in relation to the MRFFS/MRIP sampling frames and lack of discard length frequencies all add to data uncertainty. These issues were appropriately recognized and addressed by the SAW.

The SAW considered discard estimation for both commercial and recreational sectors. The commercial discard tonnage is of the same order as recreational landings but recreational discards are low and relatively unimportant. Nevertheless, the SAW was thorough in its treatment of all catch components and made adjustments to all compared to previous assessments.

A major development by the SAW was the adoption of an alternative approach to estimating commercial discards. Until 2008/09, discard estimation for scup and other Northeast region stocks used a Geometric Mean Discards-to-Landings Ratio (GMDL) method. This method was used in the existing benchmark assessment of scup done in 2008 for trawl gear only (which accounts for approximately 83% of the commercial catch). This approach is limited to using only trips with both non-zero catches and discards. A new method, Standardized Bycatch Report Method (SBRM), was adopted in 2008 and has been applied to most NEFSC-assessed stocks since that time. The SAW implemented the SBRM with three stratification alternatives and compared estimates with those using the older GMDL. The method is still applied only to trawl data and assumes a discard mortality of 100%. The performance of the different SBRM estimators and comparisons to a GMDL estimate and Dealer total landings were used to determine the best estimator. The chosen estimator (MESH240) results in lower mean discard estimates than the GMDL with different annual patterns, but the estimates are considered to be more robust. The SAW has done a thorough job implementing the new method and comparing its performance to the existing GMDL method. It is appropriate to use the new discard estimates.

2. Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty and any bias in these sources of data.

In my view, as at ToR 1, the SAW thoroughly and carefully considered the multiple sources of survey data, describing all sources and potential issues related to building the assessment model inputs. There is little to be added to what has been reported by the SARC Panel, itself reflecting the SAW Report.

The SAW provided brief outlines on a range of surveys, resulting indices, and composition data. These include the NMFS winter, spring, and fall trawl surveys, and a large number of State and academic institution surveys, carried out in waters ranging from Massachusetts to Cape Hatteras, North Carolina. Given the spatial and temporal heterogeneity of scup, it is unsurprising that the various indices display high inter-annual variability, with such variability being greater for surveys conducted in spring (catching 0 and 1-year olds) rather than fall (in which a wider range of older fish is caught). The indices also display lack of coherence, making interpretation of abundance signals difficult.

The SAW therefore applied GLM and hierarchical analyses in an attempt to determine any underlying trends and to explore the extent to which the individual indices varied from that trend.

The SAW did not use indices derived from either of the aggregating methods in the final assessment model (below) but considered each, independently in sensitivity testing. I am not especially attracted to aggregating methods, though recognize they can serve a useful exploratory purpose. In my view, each survey needs to be considered separately from first principles of design and implementation to weigh its potential value as an index of overall stock abundance. If a survey is in the spring and only catches 0 and 1 year olds, but is in a constrained location, it may be that the spatial and temporal heterogeneity of scup at those ages is such that the survey cannot be expected to index anything of value. Aggregating all available indices may be helpful if the overall coverage is sufficient to capture the extent of the population, but this will depend on the individual survey sampling protocols and selectivity and on the spatial and temporal complexity of the population and its relationship to multiple covariates. Without a clear *a priori* design accounting for the biology of the stock, any aggregating method will introduce unknown biases. I am not convinced that simply extending the number of surveys would necessarily reduce bias. In this latter point my view is not wholly consistent with the SARC Panel.

Rather than focusing on aggregating methods, in my view it would be more useful for SAWs to concentrate on the utility of individual surveys and, as in the SEDAR process, rate/rank those indices as representative of abundance. Highly ranked indices should be given priority in model tuning. Where indices are inconsistent, either they should be discarded on first principles or used for sensitivity testing.

3. Describe the thermal habitat and its influence on the distribution and abundance of scup, and attempt to integrate the results into the stock assessment.

This ToR was specific to the scup SAW. Given the difficulties associated with interpretation of multiple (often incoherent) indices, the ToR is highly relevant, at least potentially, if scup spatial and temporal heterogeneity is primarily linked to the thermal environment rather than other factors and if indices considered for modification are based on survey designs generally considered appropriate for sampling of scup.

The approach reported by the SAW involved the development and evaluation of time series of varying estimates of the proportion of thermal habitat suitability for scup surveyed by the NEFSC and NEAMAP bottom trawl surveys from 1975-2012 and integrating those indices in the model building (ToR 4). Neither survey is designed specifically to sample scup or to account for its spatial and temporal variability which may be driven by factors other than thermal habitat. In the model testing, it is unclear how other (inconsistent) indices were integrated and overall tuning was conducted for the (then) base and sensitivity tests, but the initial testing as reported suggested little impact of including the suitability indices estimated using the thermal habitat modelling.

I am not convinced the approach will provide great insights for scup and would rather a systematic approach to *a priori* consideration of index suitability as noted at ToR 2. However,

the general approach to thermal habitat monitoring could be applied widely to stocks and to multiple surveys; it is likely that at least for some stocks and surveys it will be possible to improve indices. Of course, even if statistical relationships are found, their stability is not guaranteed; the need, as ever, is to understand the processes that give rise to the complex dynamics.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.

In my view, the SAW conducted a thorough and systematic approach to both the continuity model update and to the new model-building exercise, starting with standard tuning approaches to the multiple datasets and progressing to changes in configuration and multiple sensitivity tests. The SAW report covers the multiple steps taken in developing the new model and provides useful, though necessarily summarized, final diagnostics. It is not possible to show all diagnostics for all developmental stages and to explain simply how decisions were made, though the report text does a very good job in this respect, in a concise but helpful manner. My own preference would be for more comparative diagnostics to be shown to clarify the decision-making steps, but I recognize this is a matter of degree, preference, and 'house style'. I do think that for review purposes, more emphasis could be given to comparative diagnostics rather than comparative results and fuller final model diagnostics (already shown in the main report). I do not say this with any intent of criticism, because it is difficult for the SAW and its representatives to deliver complex messages in constrained review times to diverse and variably focused reviewers. Overall, I repeat that the SAW has done a thorough and systematic job.

I agree with the SARC Panel view that the final model (S60_BASE_18) provides a robust picture of the trends in the scup stock spawning biomass and fishing mortality rates, but that the absolute values estimated are less clear. During the review there was time for only limited exploration of time-series length and selectivity assumptions. The Panel concluded that the final model is very stable in terms of stock status despite major changes in configuration. The reasons for this stability could not be explored in detail during the review process, but, because the trends in recent SSB are increasing under stable exploitation given increasing catches, the assessment was accepted as suitable for the provision of management advice. I remain sure the model is adequate for providing advice, and after revisiting the SAW report and materials am more so than expressed during the SARC Review meeting. The key issues raised in review relate to i) the length of time-series fitted, and ii) the use of domed selectivity and cryptic biomass. In particular, the Panel was concerned at the very steep rise in SSB and the very low values of estimated F, especially compared to historic levels, and given the lack of obvious reasons for such a large reduction in the exploitation history and management measures.

During review, a run was made with the time series starting in 1989 (S60_BASE_18_1989). This resulted in a reduction in the contrast in SSB and F over the time series. I note the SAW report

includes figures A88-A90, showing the same thing. This does not change the direction of the trend, but alters the rate of change in, and results in lower values of, SSB. It also leads to overall higher F estimates since the late 1990s (about 50% greater since 2001). I agree with Panel comments that a lack of age information when fitting the model prior to 1989, and back to 1963, is potentially problematic. My preference would be for a run starting in 1989 as fitting to earlier data has no obvious benefits in terms of information and potentially can be misleading. Also, with reference points based on %SPR, it makes sense to use YPR information from the most reliable part of the dataset and assessment.

The accepted assessment model indicates a very steep increase in SSB between 2000 and 2010, by a factor of more than 30. An increase in SSB is supported by almost all the data sources, but the rate of increase is much less than that estimated for each data source, in most cases being a factor of three to four. The accepted assessment has an estimated domed selectivity pattern in all fleets, peaking at age 4, remaining near the maximum at ages 5 and 6, but declining sharply at ages 7 and 8 plus. The SARC Panel requested a model run with a flat topped selectivity (S60_BASE_18_FLATL) for the catches in all periods (in a run fitting data back to 1963). Diagnostics were examined and it was considered the model fit was similar and potentially better than the accepted assessment and opined that the estimated SSB and F series were more realistic. The estimated SSB has been reduced and the trend since 2000 is less steep, while estimated F shows a slightly reducing trend but still very low values which are hard to reconcile with the catches and exploitation history. I am comfortable with the amended model fits based on very limited diagnostics but do not see a great improvement in model realism over S60_BASE-18 and am concerned that the changes in selectivity effectively only on ages 7 and 8 plus would have such an effect. I am also aware of written comments provided by commercial fishermen, which give a clear basis for accepting a domed selectivity pattern (Lapp, Almeida and Cadrin, 2015).

I am of the view that the SAW has done a good job and has provided a robust and adequate basis for stating stock status and running short-term projections in 2015. The relative stock status (ToR 6) appears to be robust to the model configurations, notwithstanding concern that the absolute levels of the indicators (SSB and F) may be less so. I am less convinced that the new assessment is suitable as a benchmark, only to be updated for a prolonged period. There is more that might yet be explored, but I am not sure how the SARC process might or might not allow for more than simple data updates between benchmarks.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, BTHRESHOLD, FMSY and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

Scup reference points were agreed following work of the 2008 Data-Poor Stocks Working Group (DPSWG) Peer Review Panel (NEFSC 2009). The adopted threshold fishing mortality reference point was $F_{MSY} = F_{40\%SPR} = 0.177$ and the corresponding spawning stock biomass, $SSB_{40\%} =$ 92,044 mt, was adopted as the target stock biomass reference point. $B_{THRESHOLD}$ was accepted as ½ $SSB_{40\%}$. The SAW correctly reported on the background to and values of these reference points.

The SAW used the new ASAP model (see ToR 4) as a basis for recommending new point estimates for reference points, accepting the same 40%SPR basis as agreed at the DPSWG in 2008. This is standard practice and was not debated by the SAW or the SARC Panel; in my view it is appropriate for scup. The new, SAW-recommended reference points are $F_{MSYproxy} = F_{40\%} =$ 0.220 (where fishing mortality is measured as 'apical' F at true age 3); $B_{MSYproxy} = SSB_{40\%} =$ 87,302 mt; and $B_{THRESHOLD} = ½\ SSB_{40\%} =$ 43,651 mt. The consequential proxy estimate for MSY = $MSY_{40\%} =$ 11,752 mt (comprised of landings of 9,445 mt and discards of 2,307 mt).

The ToR calls for estimates of uncertainty around the reference points. This is not common practice for SPR reference points and is not necessary. It is more normal to consider status with respect to reference points (ToR 6) probabilistically. The SAW provided estimates of uncertainty on the indicators (F and SSB) which are compared to reference points to guide management. In my view, while the ToR was not strictly, completely met, the SAW did all that was necessary to meet the intentions of the ToR.

6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review.

a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

The SAW updated the existing assessment model and developed a new benchmark assessment (ToR 4). For both the existing and new models, the SAW evaluated stock status with respect to the appropriate %SPR reference points (see ToR 5). In either case, the scup stock was evaluated to be neither overfished nor experiencing overfishing.

As noted at ToR 4, there are some concerns that the benchmark assessment does not fully capture recent trends in SSB and F, possibly overestimating the former and underestimating the latter. A key issue is the use of strongly domed selectivity in the benchmark assessment, creating an apparent, large cryptic biomass. Limited explorations conducted during the SARC Review considered *inter alia* alternative selectivity patterns (see ToR 4). Based on these, I am confident the current status evaluation is robust, but agree with the Panel conclusions, reflected in the revised Stock Summary, that care will be needed in future to consider this issue when monitoring

status. The domed selectivity could mask underlying status changes due to reduced recruitment or increasing fishing mortality.

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to SAW TORs for definitions).

- *a) Provide numerical annual projections (3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).*
- *b) Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions*
- *c) Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.*

a) The SAW provided projections under two 2015 catch assumptions, the difference being that either 100% or 75% of the 2015 ACL would be taken. Based on catch uptake trends and absolute values, the 75% uptake is the more likely and was favoured by the SAW. This seems reasonable.

The ToR calls for a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment). The SAW recognized the uncertainty in terminal year abundance and variability in recruitment by undertaking 100 stochastic projections using each of the 1000 MCMC estimates of terminal stock size and drawing random samples of future recruitment from the cumulative density function of the estimates of recruitment from 1984 to 2014. The SAW stated that all biological inputs to the scup assessment are well-founded, and has attended to process error and retrospective errors in terminal year estimates of abundance by inflating (by 50-100%) the OFL CV used in the projection. The SAW deemed this sufficient to account for the uncertainty seen in final year abundance/SSB estimates across a set of 25 sensitivity tests to investigate selectivity specification, ageing errors and discard determination. This approach tries to ensure enough uncertainty is reflected in the projections to cover all but major uncertainties associated with alternative states of nature. Generally, I would have preferred to see more explicit sensitivity testing of the implications of alternative selectivity specification, S-R assumptions, etc. However, I am reasonably confident that the projections provided give a sufficient view of the short term uncertainty associated with future management. This is especially so given that during the SARC review a request was made to check if the range of model estimates covered by the inflated CV included the outputs of model

S60_BASE_18_FLATL (with forced flat selectivity and reduced time-series length) explored during SARC 60 (see ToR 4); the check provided confidence this was so.

The ToR calls for the provision of annual probabilities of exceeding threshold reference points (below for SSB and above for F). These were not shown by the SAW, but it is clear from the projected estimates of SSB and from the text that the probabilities in each of the years 2016-2018 of exceeding the biomass threshold is zero.

b) I note that this ToR component (b) seems to repeat component (a). For completeness, however, the SAW considered two 2015 catch options and dealt with uncertainty *via* inflation of the CV on the OFL to reflect the range of uncertainty exposed by model sensitivity testing. However, no explicit sensitivity testing of projections to alternative states of nature was carried out.

c) The SAW ToR defines vulnerability in terms of both productivity and susceptibility, where the latter is defined as the potential for the stock to be impacted directly and indirectly by the fishery (e.g. loss of habitat). The SAW dealt with the productivity component by considering landings compared to MSY proxies and historical values. The SAW view is that the stock has low probability of becoming overfished in the short term if fishing is at the OFL. I agree with this view which is supported by the projections. The SAW did not made any comment on the matter of susceptibility but there is no indication that the fishery creates any susceptibility concerns. I am comfortable at the lack of comment by the SAW.

8. Review, evaluate and report on the status of the SARC, SSC, and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

I have commented many times in CIE reports that SAWs in various regions provide research recommendations which read as shopping lists, often lacking clarity as to purpose and which assessment or management issues are being addressed and why. Often, research recommendations appear to have been rushed due to lack of time and simply tacked on at the end. It is also of note that review panels typically deal with research recommendation ToRs in a similar fashion. Indeed, the SARC 60 Panel dealt with this ToR only on the final day, whilst also attempting to draft the SARC Report.

In my view, research recommendations and progress updates need to be approached systematically, and usually in a separate (to SAW) process. Progress needs to be looked at critically with respect to objectives, milestones, etc. New proposals need to clarify issues being considered and be specific as to how they would be addressed. Ideally, prioritization would be commented on with respect to a number of factors.

The SAW provided some progress updates on previous recommendations, though it is not clear for each item how those previous recommendations were translated into research projects or specified tasks. Some recommendations seem to have been to the SAW itself while some may have led to funded projects to feed in to the SAW. Notwithstanding, the SAW has reported progress in a few areas and no or some progress in others. It is unclear if lack of progress is due to lack of funding, changed priorities, lack of interest or capability, etc. Some previously identified research recommendations, e.g. doing experimental work to better characterize discard mortality rate in various commercial gears, seem unnecessary given their likely lack of impact on status determination or short-term management needs. Others, e.g. on evaluation of indicators of signs of potential reductions in productivity, have been reported on but unclearly; while the method of Conn (2010) might be relevant to the recommendation, it is not clear exactly how or if that work was scheduled as part of an effort to deal with the recommendation.

The SAW made six new research recommendations:

i) *A standardized fishery dependent CPUE of scup targeted tows, from either NEFOP observer samples or the commercial study fleet, might be considered as an additional index of abundance to complement survey indices in future benchmark assessments.* Such tuning information might be useful given the lack of coherent information in fishery-independent indices, but commercial selectivity is already an issue and is subject to change through time. Continued use of a new index could be problematic.

ii) *Explore additional sources of length/age data from fisheries and surveys in the early parts of the time series to provide additional context for model results.* This would be interesting given problems with lack of information prior to 1989 and model sensitivity, but may not be important in defining stock status.

iii) *Explore experiments to estimate the catchability of scup in NEFSC and other research trawl surveys (side-by-side, camera, gear mensuration, acoustics, etc.).* This seems to address the issue of catchability as a component of selectivity but does not address the apparently more critical issue of availability and spatial and temporal heterogeneity of scup for each survey. For NEFSC alone it might be useful, but it is not clear that the utility would extend to other surveys.

iv) *Refine and update the Manderson et al. availability analysis when/if a new ocean model is available (need additional support). Explore alternative niche model parameterizations including laboratory experiments on thermal preference and tolerance.* Cf iii, this would have greater potential utility though as noted at ToR 3 thermal habitat alone may be only one of a multiple of factors. Whether this is useful for scup is debatable but the recommendation is applicable more widely.

v) *Explore the Study fleet data in general for information that could provide additional context and/or input for the assessment.* This is unclear and seems to be what should be standard practice. See comments at ToR 2 regarding the need for a consideration in principle as to the utility and potential use of each survey.

vi) *A scientifically designed survey to sample larger and older scup would likely prove useful in improving knowledge of the relative abundance of these large fish.*

Information on older fish is lacking in the assessment but there is also the issue of selectivity to contend with. More generally, surveys are expensive and planning needs to take account of other species and survey needs.

Scup Conclusions and Recommendations

Conclusions

The SAW has made a thorough job of compiling and processing data as inputs to the assessment model. These have been applied methodically and carefully when fitting the existing model to new data and in development of a new benchmark assessment.

Model development has been systematic and thorough. The resulting assessment could be explored further, particularly with respect to time-series length and selectivity assumptions, but it is unlikely that any modified model would provide fundamentally different perspectives on stock status or prognosis. Overall, the model, reference point definition, stock status determination, and advice implicit in projections is reliable and robust.

In my view, the SAW has provided scientific assessments which are adequate to serve as a basis for developing fishery management advice.

Bluefish

1. Estimate catch from all sources including landings and discards. Evaluate and if necessary update the discard mortality estimate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

In my view the SAW thoroughly and carefully considered commercial and recreational landings and discard data, describing all sources and potential issues related to building the assessment model inputs. The SAW comprehensively described the spatial and temporal distribution of landings, discards, and effort and commented on uncertainties. There is little to be added to what has been reported by the SARC Panel, itself reflecting the SAW Report. Overall, the SAW took a careful and methodical approach to constructing all datasets.

Commercial Fisheries Data were obtained from standard sources, but for SARC 60, for the first time, landings data were sourced not from the NEFSC Commercial Fisheries Database but from the Atlantic Coastal Cooperative Statistics Program (ACCSP). As noted by the SAW and Panel, a discrepancy of 1.5% was seen between the ACCSP Virginia commercial reported data and Virginia's Fishery Mandatory Reporting Program Trip (FSMRPT) landings database, and the Potomac River Fishery Commission (PRFC) data that resulted from problems with quality control during data uploading and failure to synchronize data across programs when updating. To deal with this discrepancy, the Working Group (WG) chose to use either the ACCSP or the Virginia historic landings in a given year for whichever one was greater. The discrepancy is small and it is unlikely to impact the assessment or the adequacy of any conclusions drawn.

Recreational landings were obtained from the standard for-hire and angler sources and the SAW developed model inputs using standard approaches given known issues in the MRFFS and MRIP data collection schema.

Because of the declining commercial landings and minimal discards rates for the sector, the WG did not include commercial discard data into the assessment. This introduces a negligible source of bias and would not affect the adequacy of any conclusions drawn.

Recreational discards represent the second highest catch component but are uncertain as they are estimated from self-reporting by anglers during the MRFSS/MRIP survey. There has been a major trend in the proportion of fish released through time, rising from just a few percent in 1981 to approximately 60% in 2014. The increasing trend and self-reporting add uncertainty and potential bias to input data and model estimates. So too, potentially, do imprecise estimates of discard mortality. The SAW considered a working paper that presented four analyses used to estimate recreational discard mortality. The SAW also carried out a meta-analysis to determine an overall estimate of 15%, the same as previously used. The value of 15% was applied to all release estimates to derive a discard amount.

Given the recreational discard component of the catch is relatively high, the thoroughness of the SAW is to be commended.

2. Present and evaluate data and trends on life history information including, age, growth, natural mortality, food habits, and maturity.

In my view, the SAW thoroughly and carefully considered all aspects of the ToR. The SAW spent a lot of time considering age data as well as other aspects of life history. New data and/or analyses were available relevant to growth, natural mortality, maturity and food habitats. The SAW reviewed all of these, but concluded existing estimates for all remained valid. The SAW also considered stock definition and habitat used by bluefish though neither impacted on the assessment. In all respects the SAW made appropriate decisions.

The key issue under this ToR is the quantity and quality of age information used in the assessment. The SAW clearly expended considerable energy tracking scale and otolith samples from a variety of sources in order to construct and in many cases reconstruct age length keys. A major step forward by the SAW is the greater use of otoliths as opposed to scales. Where possible, otoliths have been used. Additionally, as the main discrepancies between scale and otolith ages are generally at ages greater than six, the catch-at-age matrices were constructed with a plus group at age six. Use of otoliths where possible and the use of a six plus group should have resulted in more accurate estimates of age and eliminated one source of previously identified error. The SAW should be commended on its efforts in this area.

It was noted during the review meeting that clear cohort patterns could be seen in the recreational catch-at-age data and in the MRIP recreational CPUE catch compositions, as well, to a lesser extent, in the commercial data. Caveats noted about interpretation included that the first two datasets included a considerable amount of overlap in the composition data and could not be used as confirmation of one another (both include recreational landings but different discard components), and that commercial catches are dominated by just the youngest ages, making cohort tracking debatable. Nevertheless, while the ability to track cohorts is not proof of correct age assignation, it is a strong signal thereof. In particular, the SARC Panel noted that the use of scale-read ages does not appear to be a major concern or likely source of imprecision and hence uncertainty in the assessment.

3. Present the survey data available for use in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.), evaluate the utility of the age-length key for use in stock assessment, and explore standardization of fishery-independent indices. Investigate the utility of recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data, including exploring environmentally driven changes in availability and related changes in size

structure. Explore the spatial distribution of the stock over time, and whether there are consistent distributional shifts.

In my view, as at ToR 1, the SAW thoroughly and carefully considered the multiple sources of survey data, describing all sources and potential issues related to building the assessment model inputs. There is little to be added to what has been reported by the SARC Panel, itself reflecting the SAW Report.

The SAW provided brief outlines on a range of surveys, resulting indices, and composition data. These include the NEFSC spring and fall inshore trawl surveys, NEAMAP, SEAMAP, and a large number of State surveys carried out in waters ranging from Florida to New Hampshire. Length data were available for the more extensive surveys and age compositions were constructed for these using age-length keys from commercial catches.

The SAW explored standardization of fishery-independent indices through application of a hierarchical analysis to produce a single index representing the underlying trend in bluefish YOY. The six beach seine surveys considered in the analysis are conducted by States ranging from New Hampshire to Virginia, each covering only a small geographic area relative to the overall bluefish and YOY distribution. The analysis treats all surveys as representative of the true abundance of YOY but subject to different observation and sampling error. The individual survey CVs are used to represent sampling error and hence weighting in the analysis. Effectively, the composite index is a weighted smoother of the six surveys. Its validity as an index of abundance is dependent on each being consistently representative of the annual variation in abundance and being consistently sampled in a way that maintains the relative selectivity of the surveys. Descriptions of the surveys are brief, but in most respects the technical conditions apparently hold, though none are designed specifically for bluefish. The validity of the index, which is used in the final assessment, is ultimately dependent on their joint and consistent representativeness which is unknown.

The SAW investigated the utility of a recreational LPUE index of relative abundance by querying the MRIP intercept data, defining bluefish trips as those where targeting was reported by the angler. Over 200,000 trips were identified with near half being positive. A GLM was used, with the final model selection including year, wave, mode (shore, for-hire, private/rental boat), State, and avidity. Note Area was also available and explained 5% of variance but was not selected on relative quality (AIC) grounds. During SARC 60, discussion on avidity (effectively angler experience) suggested the importance of its inclusion. The GLM-derived MRIP-CPUE index was used in the final assessment model with the associated age composition having major influence and with a sensitivity to the choice of selectivity formulation (ToR 4). Choice of selectivity (ToR 4) needs to be understood in terms of data used to develop the MRIP-CPUE index. Essentially, recreational landings are comprised of landings plus 'dead discarded' fish with many live discards due to unpalatability of older fish. The CPUE index includes landings plus all discards (including live). While recreational landings are expected to have a domed

selectivity –at-age, the selectivity for the CPUE index is expected to be flat. This caused some confusion in the discussion at ToR 4, but I am swayed by the rationale provided.

4. Estimate relative fishing mortality, annual fishing mortality, recruitment, total abundance, and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Explore inclusion of multiple fleets in the model. Include both internal and historical retrospective analyses to allow a comparison with previous assessment results and previous projections. Explore alternative modeling approaches if feasible.

In my view, the SAW conducted a thorough and systematic approach to both the continuity model update and to the new Statistical Catch-at-Age (SCAA in ASAP) model-building exercise, starting with standard tuning approaches to the multiple datasets and progressing to changes in configuration and multiple sensitivity tests. The SAW additionally considered two (now standard in the Pacific Northwest) alternative assessment models (DCAC and DBSRA). DCAC and DBSRA are simple models that do not fit to catch-at-age data as are available for bluefish. It is nevertheless useful to use alternative approaches as rough checks on advice emanating from the core assessment. (Both DCAC and DBSRA suggest that recent annual harvests were at sustainable levels).

The new model is very different to that from SARC 41. It includes new and influential data sets (e.g. MRIP-CPUE), incorporates surveys as catch-at-age rather than at-age indices, separates commercial and recreational fleets, fits selectivities by fleet (in blocks), and makes a number of technical tuning changes. Many of the changes are necessary (new data) or ensuring keeping up with best practice. Others are due to exploration. The SAW report covers the multiple steps taken in developing the new model and provides useful, though necessarily summarized, diagnostics. Some intermediate diagnostics are presented, primarily on likelihood components, to help explain model development. It is not possible to show all diagnostics for all developmental stages and to explain simply how decisions were made, though the report text does an excellent job in this respect, with concise but sufficient wording. My own preference would be for more comparative diagnostics to be shown to clarify the decision-making steps, but I recognize this is a matter of degree, preference, and 'house style'. In any case, many of the model building steps are necessary to incorporate new data or allow for technical development and do not as such require diagnostics to inform decisions. Overall, I repeat that the SAW has done a thorough, professional and systematic job.

The model is heavily influenced now by index (esp. MRIP) and catch-at-age compositions. This is a major change from SARC 41, which was weighted towards the single catch fleet.

The SARC Panel quickly accepted the final model (B043) as the basis for advice. On closer examination, however, it was noticed by one reviewer (Kupschus) that there had been a small misspecification of the model (fixing the alpha 50% for the logistic selectivity for the MRIP index at true age 0 rather than as intended at true age 1; a consequence perhaps of a coding issue

identified as a general problem). Following some further model runs and discussion with the WG, the Panel thought the most appropriate way to fix the issue was to determine the alpha 50% within the model. This resulted in the estimation of one additional parameter, but also very slightly reduced the systematic age effects in the age composition residuals of the index, though not reducing the LL overall. Because the original model specification was not the one intended by the WG, and because the freeing up of the parameter was considered more objective, the Panel, in agreement with the WG accepted the new model specification as the final model (B044). It is not a new model developed during the Review; it is just a small correction to the unintended misspecification. Because of the minor differences in model results the existing sensitivity analyses on the basis of model B043 were accepted as highly likely to be representative of the sensitivities of B044. Similarly, the projection results (ToR 7), although different in absolute terms, were also expected to be only marginally affected in relative terms. The methodology for assessing stock status and determining OFL was therefore accepted on the basis of the presentation based on Model B043, but values to be included in reports will be made using Model B044. I agree fully with this approach and commend the SAW on its quick turnaround of the new run and provision of diagnostics. I note the final projection runs (ToR 7) were not available for the Stock Summary or Panel reports (as at the time of this report).

During the SARC Review, when discussing Model B044 diagnostics and results, it was noted that the magnitude of the retrospective pattern increased relative to Model B043 (e.g. Mohn's Rho increased from 0.076 to 0.19 for SSB while displaying a near identical pattern). Two issues need to be noted. First, the size of the retrospective bias in F and SSB for model B044 fell within their respective confidence limits as determined by MCMC analyses of model B044, giving confidence that projection starting conditions and CVs will sufficiently reflect uncertainty due to the changed selectivity specification. Second, internal retrospective analyses are not model diagnostics as such, and the magnitude of Rho or retrospective pattern is not a reason to reject one model over another. In this case, the revised selectivity specification amplified the existing retrospective pattern, but was determined not to be the cause of the pattern. Because of this, and because the retrospective pattern in absolute terms was small when compared to the assessments of other species, it was deemed appropriate to use B044 as the basis for advice despite the amplified retrospective bias (cf B043). I fully agree with this view expressed by the Panel.

The accepted assessment model (B044) for bluefish represents the stock and exploitation history and is suitable for the provision of advice on stock status and short-term exploitation. However, as noted by the Panel, the model is strongly driven by the MRIP-CPUE index, which provides the majority of information at ages two and older. This index is derived from MRFSS/MRIP data, and is thus at least partially correlated with the recreational catch-at-age data derived from the same sources. The two data inputs are the predominant influences of model behavior (and hence advice) and the partial correlation is a major concern. The Panel was confident that that the SAW has done everything possible to minimize the impact of the lack of complete independence on the assessment and noted a number of factors: i) The MRFSS/MRIP CPUE were corrected for potential avidity bias in the GLM model. Avidity is a clearly important variable in deriving the MRIP-CPUE index and its inclusion is appropriate, but it does not *per se*

deal with the (lack of) independence issue; ii) The recreational catch and CPUE series differed in that the CPUE series included all the released fish, while the catch series only included 15% of the released catches. This is noted also at ToR 3. The released alive component of the total recreational catch is relatively small but it is the major component on older ages – this does not reduce the overlap in younger ages, but it is a critical difference in the information provided by the two data sources; iii) Of far greater concern given its importance to model fitting was the fact that the majority of age information for the recreational fleet was re-used in the index calculation. It was not possible to determine the degree of overlap exactly as it varies by age and time. This remains the key issue; iv) Examination of the estimated catch-at-age for the recreational fleet indicated good cohort coherence particularly in the earlier part where both aging and MRFSS data were thought to be less certain. This is true, but perhaps suggests the lack of independence is a major issue as the lack of coherence potentially introduces a very strong signal twice to the model. Given these considerations, the Panel was of the view that any future work in attaining a more independent index of the abundance of older ages should be given priority over other types of data collection to improve the objectivity of the assessment. I strongly agree with this view.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

The current statistical catch at age (SCAA) model, implemented in the ASAP program, was accepted for bluefish at SARC 41 in 2005 following extensive consideration of alternatives. MSY-based reference points were also adopted at that time, using Thompson-Bell YPR as a basis for estimating F_{MSY} and the Shepherd-Sissenwine approach to estimate B_{MSY}. The SAW correctly reported on the background to and values of these reference points (F_{MSY} = 0.19 and B_{MSY} = 147,052 mt, where B refers to total biomass; conducted an update of the existing assessment; and concluded on that basis that bluefish was neither overfished nor subject to overfishing in 2014.

The SAW argued that in order to reliably estimate MSY-based reference points as used since 2005, a stock-recruitment relationship is required, and derivation of one is not feasible for bluefish given the lack of information on recruitment at low stock sizes. The SAW therefore proposed new MSY proxy reference points based on per recruit and projection methods. The SARC Panel accepted this approach. The SAW proposed new reference points related to total biomass but, after discussion at SARC 60, it was agreed to use SSB-based proxies. It was recognized, however, that, given the current estimates of selectivity of bluefish at ages 0 and 1, prior to maturity, there would be little difference in effect.

The SAW used the new ASAP model (see ToR 4) as a basis for recommending new point estimates for reference points, based on 40%SPR. This is standard practice and was not debated by the SAW or the SARC Panel; in my view it is appropriate for scup. I agree with this approach which is standard for many east coast US stocks and internationally. The new, SAW-recommended reference points are $F_{MSYproxy} = F_{40\%} = 0.170$; $B_{MSYproxy} = SSB_{40\%} = 111{,}228$ mt; and $B_{THRESHOLD} = ½\ SSB_{40\%} = 55{,}614$ mt. The consequential proxy estimate for MSY = $MSY_{40\%}$ = 13,967 mt.

In passing, I note that at SARC 56, considering white hake (*Urophycis tenuis*), a similar approach was taken for the same underlying reasons. At that time, the SAW chose to reconsider BRPs from the %SPR family, noting that the use of $F_{40\%}$ as a common proxy was due primarily to the work of Clark. The SAW chose to use the approach of Clark but with white hake specific parameterization. I accepted this generally as a good way to proceed at that time and still consider it appropriate.

SARC 60 discussed the choice of alternative percentage values for the SPR-based reference points. It was recognized that bluefish productivity is high and that 40%SPR is a default level typically associated with less productive demersal species. While considering a lower percentage basis for MSY proxy reference points; however, it was noted that a higher value might be necessary to account for the portion of immature catches taken. I am not inclined to this view as the SPR calculation in any case takes account of the selectivity and maturity schedules. It was also noted as a personal communication by the lead assessor that mako shark have been estimated to take approximately the same amount of bluefish as the current commercial activities. This raised issues of multispecies aspects of M estimation not considered by the working group and was beyond the scope of the review. The SARC Panel agreed that the 40%SPR basis should be maintained at this time, but that appropriateness of the %SPR levels should be given further consideration in future. I agree with this and note that the methods applied at SARC 56 might be the simplest way forward, notwithstanding multispecies issues.

The ToR calls for estimates of uncertainty around the reference points. This is not common practice for SPR reference points and is not necessary. It is more normal to consider status (ToR 6) probabilistically. The SAW provided estimates of uncertainty on the indicators (F and SSB), which are compared to reference points to guide management. In my view, while the ToR was not strictly, completely met, the SAW did all that was necessary to meet the intentions of the ToR.

6. Evaluate stock status with respect to the existing model (from previous peer review accepted assessment) and with respect to a new model developed for this peer review.

A When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

B Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

The SAW updated the existing assessment model and developed a new benchmark assessment (ToR 4). For both the existing and new models, the SAW evaluated stock status with respect to the appropriate %SPR reference points (see ToR 5). In either case, the scup stock was evaluated to be neither overfished nor experiencing overfishing.

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level; see Appendix to the SAW TORs).

a) Provide annual projections (3 years). For given catches, each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
b) Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
c) Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

Projections from the accepted model were made using standard NEFSC AGEPRO software. Projections for a range of constant F scenarios for 2016-2018 and 2015 removals set to the 2015 quota were run, all starting with initial abundance estimates drawn from MCMC runs of the accepted model B044 (run during SARC 60 and intended to be finalized immediately following but not available at the time of submitting this CIE report). The projections were done using a single fleet; this required development of a combined selectivity curve, based on the last three years of the model estimates. A small CV was added to the selectivity-at-age estimates, as was also done for biological parameters drawn from lognormal distributions. The SAW ran a number of sensitivity tests using model B043 to address potential uncertainties using the model as presented to the SARC; based on those tests, conclusions drawn using the accepted model B044 are expected to be robust. The ToR calls for the provision of annual probabilities of exceeding threshold reference points (below for SSB and above for F); these are not shown by the SAW but it is clear from the projected estimates of SSB and from the text provided by the SAW that the probabilities in each of the years 2016-2018 of the SSB falling below the biomass threshold is zero.

It is not clear why the wide range of constant F projections is run. I can see why *status quo*, $F_{MSYProxy}$ and F_{TARGET} would be considered but runs using F_{low} and $F_{0.1}$ seem to be vestigial.

Use of a single fleet is odd, given the assessment itself used disaggregated commercial and recreational datasets. Model B043 estimated selectivities for the two sectors were not dissimilar. The changed selectivity at age formulation in Model B044 applies only to the MRIP-CPUE index and the estimated commercial and recreational selectivities are little affected, still being similarly domed. I would expect the combining of fleets with a common selectivity to remain similar for the new (as yet unseen) Model B044-based projections.

Sensitivity testing was constrained to recruitment time-series, constant M and increased CV on biological parameters. The first of these is typically important to consider recent versus longer recruitment levels. In this case the two are similar. Results on changed M and increased input CVs are as expected. Overall, the tests appear sufficient to test short-term robustness of any conclusions that might be drawn from the projections.

Overall, the projections based on Model B043 (and Model B044 to come) appear to be the most realistic for use in advisory and decision-making processes. There do not appear to be any exceptional issues related to vulnerability (as productivity or susceptibility) that would influence decision-making.

8. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports, as well as MAFMC SSC model recommendations from 2005 and the research recommendations contained in its 23 September 2013 report to the MAFMC. Identify new research recommendations.

I have commented many times in CIE reports that SAWs in various regions provide research recommendations which read as shopping lists, often lacking clarity as to purpose and which assessment or management issues are being addressed and why. Often, research recommendations appear to have been rushed due to lack of time and simply tacked on at the end. It is also of note that review panels typically deal with research recommendation ToRs in a similar fashion. Indeed, the SARC 60 Panel dealt with this ToR only on the final day, whilst also attempting to draft the SARC Report.

In my view, research recommendations and progress updates need to be approached systematically, and usually in a separate (to SAW) process. Progress needs to be looked at critically with respect to objectives, milestones, etc. New proposals need to clarify issues being considered and be specific as to how they would be addressed. Ideally, prioritization would be commented on with respect to a number of factors.

The SARC Panel noted these issues and that many of the previously listed research recommendations are unclear and possibly out of date. It was therefore difficult to comment on those items. In response, the SAW Chair worked with the SAW during SARC 60 to revise the SAW Report sections on progress on previous research recommendations and on new recommendations. The SAW's responsiveness during SARC 60 is much appreciated; it would be useful if the SAW (all SAWs) could build on this to provide more informative research recommendation sections, with clear opinion on the utility of specified work in furthering assessment and especially management needs, possibly with consideration of costs *vs* benefits.

The SAW made eight new research recommendations:

i) High Priority: Determine whether NC scale data from 1985-1995 are available for age determination; if available, re-age based on protocols outlined in ASMFC (2011); if re-aging results in changes to age assignments, quantify the effects of scale data on the assessment. The SAW noted this *"Would allow for validation of the adjustments to the early NC spring age data made by WG at model meeting"* (WP B6). The Panel noted that while this would improve data input accuracy, it would have a lower impact on model results and suggested it should not be only a moderate priority undertaking. I agree with this view.

ii) High Priority: Develop additional adult bluefish indices of abundance (e.g., broad spatial scale longline survey or gillnet survey). The SAW noted that *Given the limited information on older (e.g., age 2+) bluefish collected by existing fishery independent surveys this item addresses the need to adequately characterize dynamics of older fish that are currently not well sampled by fishery independent trawl surveys.* The Panel noted the high importance of this recommendation. I agree with this view but see (iii), below.

iii) High Priority: Expand age structure of SEAMAP index. The SAW noted that *Given patterns of bluefish migration and recruitment (Shepherd et al. 2006, Wuenschel et al. 2012), it is important to monitor bluefish abundance in SAB; currently, the SEAMAP index used in the assessment indexes age 0 abundance only, but recent age data from SEAMAP suggests collection of age 1 and 2 fish that would help inform the SAB age structure.* The Panel noted this would address concerns over incomplete mixing of the population at younger ages (see ToR 3) and, if designed correctly, this could be addressed by the survey discussed in the previous research recommendation (ii). I agree with this but note the emphasis is on deriving the same expansion of information on older fish through extended sampling of the existing SEAMAP survey rather than development of a new survey. Cost-benefit considerations need to be taken in to account in considering how to prioritize recommendations (ii) and (iii).

iv) Moderate Priority: Investigate species associations with recreational angler trips targeting bluefish (on a regional and seasonal basis) to potentially modify the MRIP index used in the assessment model. The SAW noted that *"Given the importance of the MRIP index in the assessment model, this addresses a need to accurately estimate effort for of the MRIP index (reduce risk of hyperstability)".* The Panel noted the importance of the MRIP-

CPUE index in model fitting and suggested this be given a high priority. I agree and would see this as higher priority than any of (i) through (iii), above.

v) *Moderate Priority: Explore age- and time-varying natural mortality from, for example, predator prey relationships; quantify effects of age- and time-varying natural mortality in the assessment model.* The SAW noted that *"This addresses the issue of predation on bluefish by, for example, coastal sharks and/or limited prey resources (top down effects, bottom up effects, and/or environmental effects)".* The Panel noted that this would warrant a high priority in an ecosystem context, but lower priority for assessment per se, as only in the longer term is it likely to inform the estimate of M. In my view, the priority rating for assessment needs, including reference point setting, could be addressed through targeted sensitivity testing using the new benchmark as a basis. I note the recommendation is not for data collection and analysis but is rather for model-based exploration.

vi) *Moderate Priority: Continue to evaluate the spatial, temporal, and sector-specific trends in bluefish growth and quantify their effects in the assessment model.* The SAW noted that this *Addresses appropriateness of WG pooling age data spatially (and temporally) for potential changes regarding the efficiency of the biological collection program.* The Panel noted that this item addresses a concern over incomplete mixing of the population, differential growth and/or possible regional differences in selectivity, and results are likely to improve the results produced by the assessment model. I agree.

vii) Moderate Priority: Continue to examine alternative models that take advantage of length-based assessment frameworks. Evaluate the source of bimodal length frequency in the catch (e.g., migration, differential growth rates). The SAW noted that "*This item would address a source of uncertainty in the assessment with age data from different hard parts & provide means to examine the appearance of bimodal length frequency in the catch data".* The Panel noted that this item addresses two independent issues that should be considered separately. First, whether there are potential benefits (and costs) of moving to a length-based selectivity model given that (a) the predominant gear is not particularly size-selective, (b) growth is very rapid, and (c) migration patterns appear to be age rather than length dependent. Given these considerations, the Panel considered the recommendation to be of low priority. Second, whether or not it is important to conduct research to determine the process that leads to bimodality of length compositions. The Panel does not understand the concern over bimodality of length compositions as these may well be caused by age modalities. Sufficient data should exist to allow this to be addressed now. My understanding is that there is already sufficient confidence in the scale-based ageing and the use of the six plus group reduces any likely conflicts caused by using scale- and otolith-based ageing in different datasets. Unless there is a major concern about ageing, I can see no benefit of moving to a length-based model for a stock with strong age- rather than length-dependent life-history and exploitation processes, In general, where data permit, age-based models should be preferred. I do not see any problem with the bimodal length frequencies and agree with the Panel view.

viii) Moderate Priority: Modify thermal niche model to incorporate water temperature data more appropriate for bluefish in a timelier manner [e.g., sea surface temperature

data & temperature data that cover the full range of bluefish habitat (SAB and estuaries)]. The SAW noted that *"This addresses the current limitations of the habitat suitability model for bluefish (limited to hindcast bottom temps, in the MAB)"*. The Panel noted that given how little effect the current habitat model has on bluefish, this research recommendation would appear to be of low priority for this species but may be of higher priority for other species. I agree with this view.

Bluefish Conclusions and Recommendations

The SAW has made a thorough job of compiling and processing data as inputs to the assessment model. These have been applied methodically and carefully when fitting the existing model to new data and in development of a new benchmark assessment.

Model development has been systematic and thorough. The new model is a major improvement over the existing benchmark, bringing in new data sources, disaggregating data, attending to developments in best practice, and exploring more thoroughly selectivity patterns. Overall, the model, reference point definition, stock status determination, and advice implicit in projections are reliable and robust. (Note that at the time of this report, the final projections have yet to be made available.)

In my view, the SAW has provided scientific assessments, which are adequate to serve as a basis for developing fishery management advice.

APPENDIX 1

BIBLIOGRAPHY

Prior to the Workshop, extensive materials were provided *via* a weblink (http://www.nefsc.noaa.gov/SARC/SARC-60-pdfs/). Materials included the draft SAW reports and draft stock summaries, previous assessment reports and reviews, and background documents. No presentations were made available in advance. During the workshop, an internal network filing system and IT support were provided.

During the workshop a number presentations were given, and some additional materials were provided on request, including further background documents and presentations as well as responses to Panel requests. All files were made available through the internal network. The access was generally adequate.

Further references:

Lapp, Almeida and Cadrin, 2015, Commercial Fishermen's Perspectives on the Scup Fishery and Resource. Written submission to SARC 60 Review

APPENDIX 2

Attachment A: Statement of Work for Dr. Kevin Stokes

60th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for scup and bluefish

Statement of Work (SOW) for CIE Panelists
(including a description of SARC Chairman's duties)

BACKGROUND

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are independently selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description: The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC peer review is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development and report preparation (which is done by SAW Working Groups or ASMFC technical committees), assessment peer review (by the SARC), public presentations, and document publication. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fisheries within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO).

The purpose of this meeting will be to provide an external peer review of benchmark stock assessments for **scup** and **bluefish**.

OBJECTIVES

The SARC review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the SSC of the New England or Mid-Atlantic Fishery Management Council. The SARC panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Duties of reviewers are explained below in the "**Requirements for CIE Reviewers**", in the "**Charge to the SARC Panel**" and in the "**Statement of Tasks**". The draft stock assessment Terms of Reference (ToRs) which are carried out by the SAW WGs are attached in **Annex 2**. The draft agenda of the panel review meeting is attached in **Annex 3**. The SARC Summary Report format is described in **Annex 4.**

Requirements for the reviewers: Three reviewers shall conduct an impartial and independent peer review of the **scup** and **bluefish** stock assessments, and this review should be in accordance with this SoW and stock assessment ToRs herein. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include statistical catch-at-age, state-space and index models. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points that includes an appreciation for the varying quality and quantity of data available to support estimation of Biological Reference Points. SARC 59 will address fishery stock assessments of **scup** and **bluefish**. For both species, experience in assessing pelagic stocks and in incorporating environmental factors into assessments would be desirable. For bluefish, experience in the use of recreational fisheries data would also be desirable.

PERIOD OF PERFORMANCE

The contractor shall complete the tasks and deliverables as specified in the schedule of milestones within this statement of work. Each reviewer's duties shall not exceed a maximum of 16 days to complete all work tasks of the peer review described herein.

Not covered by the CIE, the SARC chair's duties should not exceed a maximum of 16 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation).

PLACE OF PERFORMANCE AND TRAVEL

Each reviewer shall conduct an independent peer review during the panel review meeting scheduled in Woods Hole, Massachusetts during June 2-5, 2015.

STATEMENT OF TASKS

Charge to SARC panel: During the SARC meeting, the panel is to determine and write down whether each stock assessment Term of Reference (ToR) of the SAW (see **Annex 2**) was or was not completed successfully. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. **If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted.** Where possible, the SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment Term of Reference of the SAW.

If the panel rejects any of the current BRP or BRP proxies (for B_{MSY} and F_{MSY} and $MSY_{)}$, the panel should explain why those particular BRPs or proxies are not suitable, and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs or BRP proxies are the best available at this time.

Each reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review of stock assessments prepared by SAW WGs or ASMFC Technical Committees in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, FAX number, and CV suitable for public distribution) to the COR, who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports for review, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Foreign National Security Clearance: The reviewers shall participate during a panel review meeting at a government facility, and the NMFS Project Contact will be responsible for obtaining the Foreign National Security Clearance approval for the reviewers who are non-US citizens. For this reason, the reviewers shall provide by FAX or by email the following

requested information (e.g., 1.name [first, middle, and last], 2.contact information, 3.gender, 4.country of birth, 5.country of citizenship, 6.country of permanent residence, 7.whether there is dual citizenship, 8.country of current residence, 9.birth date [mo, day, year], 10.passport number, 11.country of passport) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/.

Pre-review Background Documents and Working Papers: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the SARC chair and CIE reviewers the necessary background information and reports (i.e., working papers) for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review of the stock assessments in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

(SARC chair)
Act as chairperson, where duties include control of the meeting, coordination of presentations and discussions, making sure all stock assessment Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For each assessment, review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)
For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. From a reviewer's point of view, determine whether each stock assessment Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point or BRP proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist. Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

Tasks after the panel review meeting:

SARC CIE reviewers:
Each CIE reviewer shall prepare an Independent CIE Report (see **Annex 1**). This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified above in the "Charge to SARC panel" statement.

If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report produced by each reviewer.

The Independent CIE Report can also be used to provide greater detail than the SARC Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

SARC chair:
The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was

adequate to complete the stock assessment Terms of Reference of the SAW. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report (see **Annex 4**).

SARC chair and CIE reviewers:
The SARC Chair, with the assistance from the CIE reviewers, will prepare the SARC Summary Report. Each CIE reviewer and the chair will discuss whether they hold similar views on each stock assessment Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this SARC Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion.

The SARC Summary Report (please see **Annex 4** for information on contents) should address whether each stock assessment Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

DELIVERY

Each reviewer shall complete an independent peer review report in accordance with the SoW. Each reviewer shall complete the independent peer review according to required format and

content as described in **Annex 1**. Each reviewer shall complete the independent peer review addressing each stock assessment ToR listed in **Annex 2**.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the Woods Hole, Massachusetts scheduled during the tentative dates of June 2-5, 2015.
3) Conduct an independent peer review in accordance with this SoW and the assessment ToRs (listed in **Annex 2**).
4) No later than June 19, 2015, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to mshivlani@ntvifederal.com, and to Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each assessment ToR in **Annex 2**.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

April 24, 2015	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
May 19, 2015	NMFS Project Contact will attempt to provide reviewers the pre-review documents
June 2-5, 2015	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
June 5, 2015	SARC Chair and CIE reviewers work at drafting reports during meeting at Woods Hole, MA, USA
June 19, 2015	Reviewers submit draft independent peer review reports to the contractor's technical team for independent review
June 19, 2015	Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair *
June 26, 2015	SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)

July 2, 2015	Contractor submits independent peer review reports to the COR who reviews for compliance with the contract requirements
July 10, 2015	The COR distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on substitutions. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: The deliverables shall be the final peer review report from each reviewer that satisfies the requirements and terms of reference of this SoW. The contract shall be successfully completed upon the acceptance of the contract deliverables by the COR based on three performance standards:

(1) each report shall be completed with the format and content in accordance with **Annex 1**,
(2) each report shall address each stock assessment ToR listed in **Annex 2**,
(3) each report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Upon the acceptance of each independent peer review report by the COR, the reports will be distributed to the NMFS Project Contact and pertinent NMFS science director, at which time the reports will be made publicly available through the government's website.

The contractor shall send the final reports in PDF format to the COR, designated to be Allen Shimada, via email allen.shimada@noaa.gov

Support Personnel:

Allen Shimada, COR
NMFS Office of Science and Technology

1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
allen.shimada@noaa.gov Phone: 301-427-8174

Manoj Shivlani, CIE Lead Coordinator
NTVI Communications, Inc.
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136

Key Personnel:

Dr. James Weinberg, NEFSC SAW Chairman, NMFS Project Contact
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov (Phone: 508-495-2352) (FAX: 508-495-2230)

Dr. William Karp, NEFSC Science Director
Northeast Fisheries Science Center
166 Water St., Woods Hole, MA 02543
william.karp@noaa.gov Phone: 508-495-2233

Annex 1: Format and Contents of Independent Peer Review Report

1. The independent peer review report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).

2. The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs. For each assessment reviewed, the report should address whether each ToR of the SAW was completed successfully. For each ToR, the Independent Review Report should state why that ToR was or was not completed successfully. To make this determination, the SARC chair and reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the SARC Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not others read the SARC Summary Report. The independent report shall be an independent peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: 60th SAW/SARC Stock Assessment Terms of Reference (file vers.: 10/162014)

8. Scup

1. Estimate catch from all sources including landings and discards. Include recreational discards, as appropriate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

2. Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty and any bias in these sources of data.

3. Describe the thermal habitat and its influence on the distribution and abundance of scup, and attempt to integrate the results into the stock assessment.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review.

a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to SAW TORs for definitions).

a. Provide numerical annual projections (3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. Review, evaluate and report on the status of the SARC, SSC, and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

Annex 2: (cont)

9. Bluefish

a. Estimate catch from all sources including landings and discards. Evaluate and if necessary update the discard mortality estimate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

b. Present and evaluate data and trends on life history information including, age, growth, natural mortality, food habits, and maturity.

c. Present the survey data available for use in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.), evaluate the utility of the age-length key for use in stock assessment, and explore standardization of fishery-independent indices. Investigate the utility of recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data, including exploring environmentally driven changes in availability and related changes in size structure. Explore the spatial distribution of the stock over time, and whether there are consistent distributional shifts.

d. Estimate relative fishing mortality, annual fishing mortality, recruitment, total abundance, and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Explore inclusion of multiple fleets in the model. Include both internal and historical retrospective analyses to allow a comparison with previous assessment results and previous projections. Explore alternative modeling approaches if feasible.

e. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

f. Evaluate stock status with respect to the existing model (from previous peer review accepted assessment) and with respect to a new model developed for this peer review.
   i. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
   ii. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

g. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level; see Appendix to the SAW TORs).
   i. Provide annual projections (3 years). For given catches, each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
   ii. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
   iii. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

h. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports, as well as MAFMC SSC model recommendations from 2005 and the research recommendations contained in its 23 September 2013 report to the MAFMC. Identify new research recommendations.

Annex 2: (cont)

Appendix to the SAW Assessment TORs:
Clarification of Terms used in the SAW/SARC Terms of Reference

On "Overfishing Limit" and Acceptable Biological Catch" (DOC Nat. Stand. Guidel. Fed. Reg., v. 74, no. 11, 1-16-2009):

Acceptable biological catch (ABC) is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of [overfishing limit] OFL and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*

ABC for overfished stocks. For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect annual catch that is consistent with schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*

NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)

ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, [optimal yield] OY does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

On "Vulnerability" (DOC Natl. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):

"Vulnerability. A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce MSY and to recover if the population is depleted, and susceptibility is the

potential for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Interactions among members of a SAW Assessment Working Group:

Anyone participating in SAW assessment working group meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

One model or alternative models:

The preferred outcome of the SAW/SARC is to identify a single "best" model and an accompanying set of assessment results and a stock status determination. If selection of a "best" model is not possible, present alternative models in detail, and summarize the relative utility each model, including a comparison of results.

Annex 3: Draft Agenda

60th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for A. scup and **B. bluefish**

June 2-5, 2015

Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

DRAFT AGENDA* (version: Dec. 1, 2014)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Tuesday, June 2

10 – 10:30 AM

Welcome	**James Weinberg**, SAW Chair
Introduction	**TBD,** SARC Chair
Agenda	
Conduct of Meeting	

10:30 – 12:30 PM Assessment Presentation (A. Scup)
Mark Terceiro TBD **TBD**

12:30 – 1:30 PM Lunch

1:30 – 3:30 PM Assesssment Presentation (A. Scup)
Mark Terceiro TBD **TBD**

3:30 – 3:45 PM Break

3:45 – 5:45 PM SARC Discussion w/ Presenters (A. Scup)
TBD, SARC Chair **TBD**

5:45 – 6 PM Public Comments

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Wednesday, June 3

8:30 – 10:30 AM	Assessment Presentation (B. bluefish)		
	Tony Wood	TBD	**TBD**
10:30 – 10:45 AM	Break		
10:45 – 12:30 PM	(cont.) Assessment Presentation (B. bluefish)		
	Tony Wood	TBD	**TBD**
12:30 – 1:30 PM	Lunch		
1:30 – 3:30 PM	SARC Discussion w/presenters (B. bluefish)		
	TBD, SARC Chair		**TBD**
3:30 – 3:45 PM	Public Comments		
3:45 -4 PM	Break		
4 – 6 PM	Revisit with presenters (A. Scup)		
	TBD, SARC Chair		**TBD**
7 PM	**(**Social Gathering **)**		

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Thursday, June 4

8:30 – 10:30	Revisit with presenters (B. bluefish)		
	TBD, SARC Chair		**TBD**
10:30 – 10:45	Break		
10:45 – 12:15	Review/edit Assessment Summary Report (A. Scup)		
	TBD, SARC Chair		**TBD**
12:15 – 1:15 PM	Lunch		
1:15 – 2:45 PM	(cont.) edit Assessment Summary Report (A. Scup)		
	TBD, SARC Chair		**TBD**
2:45 – 3 PM	Break		
3 – 6 PM	Review/edit Assessment Summary Report (B. bluefish)		
	TBD, SARC Chair		**TBD**

Friday, June 5

9:00 AM – 5:00 PM	SARC Report writing.

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public. The public should not engage in discussion with the SARC during SARC report writing, which is scheduled for June 5.

***The NMFS Project contact will provide the final agenda about four weeks before meeting. Reviewers must attend the entire meeting.**

Annex 4: Contents of SARC Summary Report

1.

The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

To make this determination, the SARC chair and CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If the CIE reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

The report may include recommendations on how to improve future assessments.

2.

If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3.

The report shall also include the bibliography of all materials provided during the SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

The report shall also include as a separate appendix the assessment Terms of Reference used for the SAW, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

APPENDIX 3
PERTINENT INFORMATION FROM THE REVIEW

NAME	AFFILIATION
Jim Weinberg	NEFSC
Paul Rago	NEFSC
Mike Simpkins	NEFSC
Sheena Steiner	NEFSC
Chris Legault	NEFSC
Gary Shepherd	NEFSC
Mark Terceiro	NEFSC
Tony Wood	NEFSC
Kirby Rootes-Murdy	ASMFC
Katie Drew	ASMFC
Mike Celestino	NJ DFW
Joey Ballenger	SCDNR
Julia Beaty	MAFMC
Jocelyn Runnebaum	Univ. of Maine
Nicole Lengyel	RI DEM
DFWJason McNamee	RIDFW/ASMFC
Steve Cadrin	SMAST
Wendy Gabriel	NEFSC/MAFMC
Chuck Adams	NEFSC
David McElroy	NEFSC
John Manderson	NEFSC
Brian Linton	NEFSC
Mike Palmer	NEFSC
Susan Wigley	NEFSC
Alicia Miller	NEFSC
Kiersten Curti	NEFSC
Larry Alade	NEFSC
Jon Deroba	NEFSC
Loretta O'Brien	NEFSC
Paul Nitschke	NEFSC