UNITED STATES DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
NATIONAL MARINE FISHERIES SERVICE
Northeast Fisheries Science Center
166 Water Street
Woods Hole, MA 02543-1026

January 17, 2012

Capt. Paul J. Howard
Executive Director
New England Fishery Management Council
50 Water Street
Newburyport, MA 01950

Dear Paul:

Thank you for your letter of January 11 requesting the Center to review the document prepared by Mr. David Goethel on the SARC 53 Gulf of Maine cod assessment. My staff has reviewed this document, and enclosed are our comments for consideration by the SSC at its January 25 meeting.

If you have any questions, please contact me.

Sincerely,

William A. Karp, Ph.D.
Acting Science and Research Director

Enclosure

2010 Gulf of Maine Cod Working Group Assessment Notes

The following notes were prepared by NOAA's Northeast Fisheries Science Center at the request of the New England Fishery Management Council to assist its Scientific and Statistical Committee in reviewing the recent assessment of Gulf of Maine cod. The issues addressed were raised by a council member in written form, and that document will also be considered by the Scientific and Statistical Committee.

Biology

- *Stock identification is incorrect for cod in New England*
    - *Tagging evidence suggests that cod stock boundaries should be separated into: eastern Gulf of Maine-Eastern Georges Bank, and western Gulf of Maine-Cape Cod-Southern New England.*
    - *Information regarding stock structure and discussions regarding proper stock boundaries were explicitly avoided during the assessment meetings for Gulf of Maine cod even though the current boundaries are highly questionable in light of the last decade of scientific research.*

    Reconsidering cod stock boundaries requires a comprehensive evaluation for all stock components and available information, including any new tagging data, rather than one component in isolation. This kind of review is beyond the normal SAW/SARC benchmark assessment process, and was not a term of reference for the SARC 53 Gulf of Maine cod assessment. Reconsideration of boundaries also affects management measures that are based on them, such as annual allocations of cod to the fishery, and fishery data collection and monitoring requirements. This kind of thorough biological and management review is best conducted in collaboration with Canadians and all stakeholders well in advance of any benchmark assessment for a single stock.

    It is widely believed that the recent expansion of cod into Southern New England (a region with historically low cod abundance in recent decades) is due to a 'spillover' migration effect of cod from the Gulf of Maine. This suggests that the Gulf of Maine cod stock is actually expanding and contradicts the stock contraction hypothesis being presented by NEFSC.

Based on NEFSC survey data for cod, the contraction of cod to the western Gulf of Maine is a statement of fact. Within the Gulf of Maine region cod do not exhibit a wide spatial distribution as they did prior to 1980; cod are not showing up in areas where they have been historically plentiful (central and eastern Gulf of Maine). We are aware that fishermen are reporting cod in high densities in certain areas of southern New England. However, preliminary explorations of biomass trends in the southern New England waters suggests that overall biomass in these areas has declined over the past forty years with little to no evidence of recent increases (more details on these analyses are presented later). Such trends would appear to run counter to a 'spillover' hypothesis.

- *The length-weight relation and catch weight-at-age matrix are unreliable*
    - *The length-weight relation is based on survey catch and not on the commercial catch*

      The rationale for how the length-weight relationship was derived was provided on page 15 of SARC 53 WP#1:

      "Currently in the Northeast Region, fishery surveys are the only source of individual length-weight sampling."

      Had other sources of individual lengths and weights been available and presented during the Gulf of Maine Cod Data Meeting, these could have been evaluated and compared to the LW relationships generated from the survey data. The length-weight relationship is used not only to estimate catch weights, but to also derive spawning and stock weights. It is more appropriate to use a survey-based length-weight relationship to derive spawning and stock weights to avoid bias caused any size-at-age selectivity associated with gears used in the fishery.

    - *Catch weight-at-age matrix is averaged over the recreational and commercial fisheries and over discard and landed catch*
        - *This acts to blur the signals in the catch because the weight of recreationally caught fish are lower than that of commercial fish thereby decreasing the weight of 'caught' fish in the model*

      This is precisely why the weights-at-age used in the previous (2008) assessment were biased high; the previous assessment did not fully account for the lighter fish-at-age that characterize the discard component of the catch. An average weight-at-age from all catch sources is the correct method to estimate true catch weights-at-age. Additionally, the average weights-at-age were not simple arithmetic averages across all catch sources, but rather were weighted proportionally by the numbers in each of the catch sources. A numbers-weighted average approach preserves the proportional contribution of each catch source to the weight-at-age. This is clearly documented on page 28 of SARC 53 WP#1:

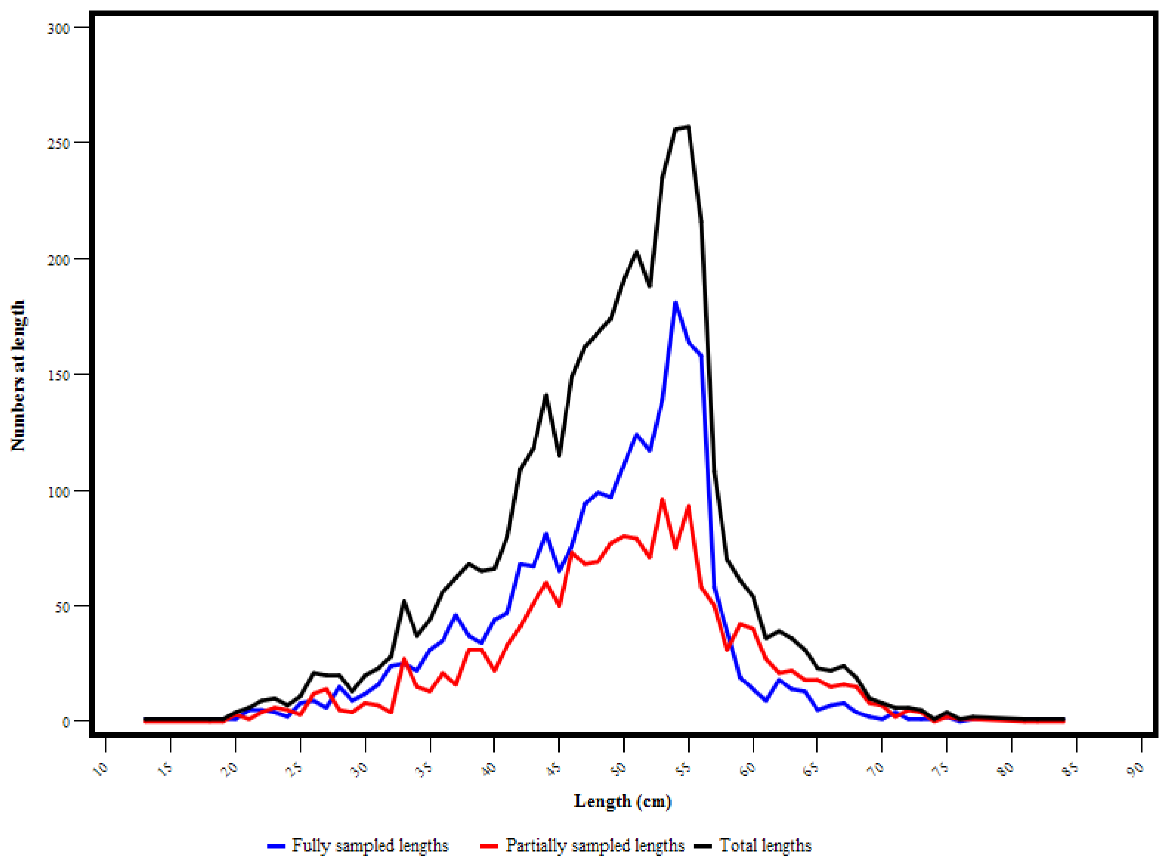
      "Mean catch weights-at-age were estimated by using a numbers weighted average of the individual catch component's mean weight-at-age."

- *There is an apparent under-sampling of older fish in the catch-at-age, which gives the appearance of a truncated age-structure and increased F-at-age*

      The basis or evidence for this statement is not clear. Sampling coverage rates, by market category (Table A.10), reveal a nearly 10-fold increase in sampling coverage since the 1990s. Moreover, in recent years, the number of length samples of "large" cod has often exceeded the number of samples of "market" size cod.

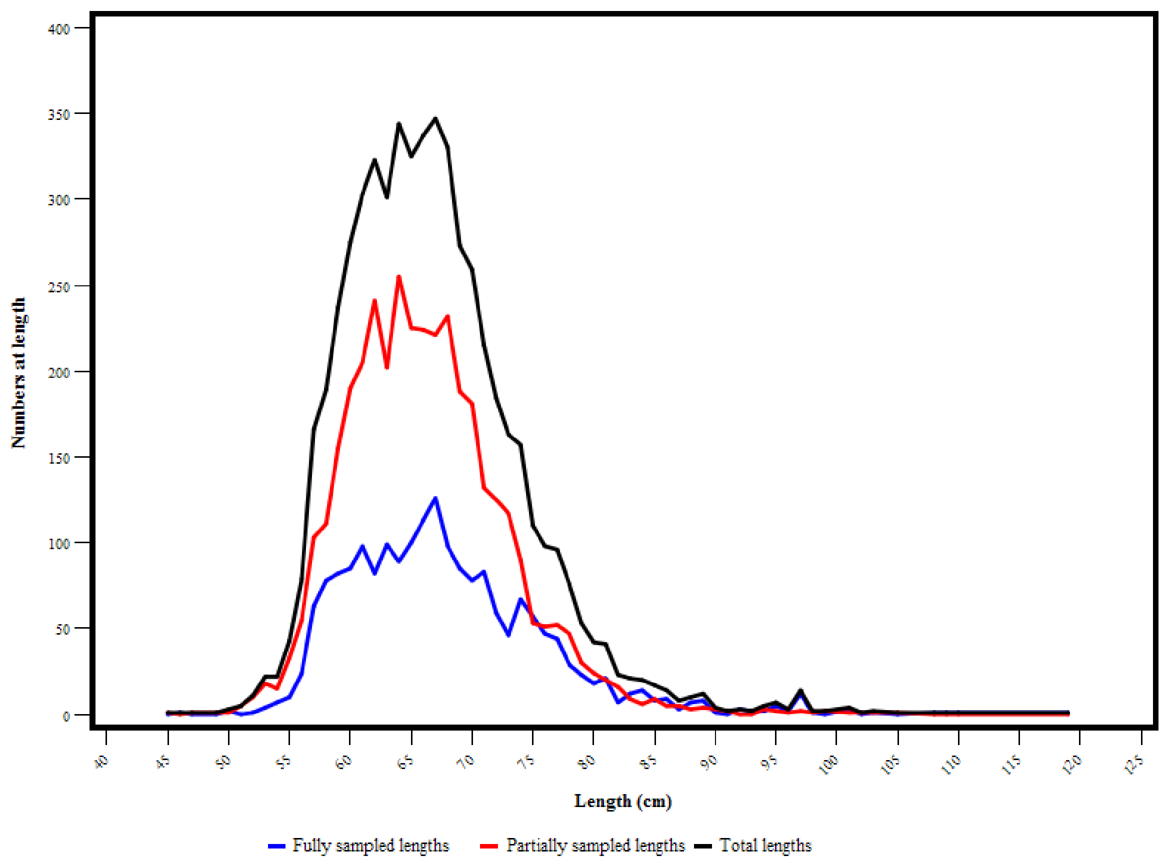
- *Observer samples do not accurately reflect the actual catch, perhaps due to focusing on measuring smaller discarded fish.*

  The available data do not support this statement. As an example, the figure depicts the length frequency distribution of Gulf of Maine cod discarded by the otter trawl fleet in 2008, based on two types of length frequency samples taken by observers: the blue line represents the length frequency distribution when all discarded cod in a particular tow were measured (n=2,305 lengths); the red line represents the length frequency distribution when only a fraction of the discarded cod could be measured (n=1,671 lengths) (and the black line is the combined length frequency distribution using all the sampling data). Partial sampling occurs when an observer cannot measure the entire catch due to time constraints, or the volume of discards is so great that a representative subsample is taken to adequately characterize the size distribution of the catch. If observers preferentially measured small fish in the discards when they sampled only a subset of the fish, the red line should be shifted to the left. However, it is not. The length frequencies from the two types of samples are nearly identical indicating that when observers take subsamples, they do so in a manner that provides an accurate representation of the length composition of the catch.



- *Large, older fish are not being properly sampled and this has led to a lack of old (age-7+) fish being 'seen'.*

  Assuming that this refers to observer catch samples, the available data do not support this statement. As most older cod are the larger cod that are typically retained and landed, the analysis shown above for discarded cod was repeated, except using the observed length frequencies of the <u>retained</u> catch (year=2009, gear=otter trawl). In the figure below, the blue line is the length frequency distribution when all retained cod in a particular tow were measured (n=1,923 lengths), and the red line is the length frequency distribution when only a fraction of the retained cod could be measured (n=3,691 lengths). Again, the length frequency distributions are nearly identical implying that all lengths (and ages) are representatively sampled.



- *Very few otoliths of fish greater than age -7 are being taken during surveys (which is used for determining the length-weight equation).*

  Large cod taken in the survey have the same sampling intensity as any other size group of cod. The survey samples otoliths in proportion to the length composition of the survey catches. The sampling protocol for cod on surveys throughout most of the assessment time series has been to obtain 3 otolith samples per 3 cm length frequency interval.

- *It is likely that port samples of cod otoliths are also biased towards smaller fish (reported samples are divided by market category and not age; it is likely that samplers are taking a majority of otoliths from smaller fish within each market category and sampling at times of year and ports where large fish are not being landed, thereby violating the random stratified sampling design).*

  The basis for this statement is not presented. Table A.10 in the assessment report shows the thoroughness of the sampling effort with respect to market categories and seasonal sampling. This table also indicates that substantial improvements have occurred in commercial sampling efforts during the past 10 years.

- *Use of external data sources could help to verify age-structure information (e.g., gear studies and tagging studies that have information on age or length structure could be used as an exploratory check).*

  Agreed. However, no validation data were brought forward for consideration by the Data Working Group

- *The age-9+ formulation is invalid and underutilizes the flexibility of the Age-Structured Assessment Program (ASAP) framework.*

  It is unclear how use of an age 9^+ group underutilizes the flexibility of the ASAP framework, or why the age 9+ formulation is thought to be "invalid". Clearly, considerable uncertainty exists in the catch-at-age beyond age 9, and even at younger ages earlier in the assessment time period (see Table A.14 in SARC 53 WP#1). More importantly, for 13 years in the time period covered in the assessment, no fish older than 11 years old occurred in the catch samples. Moving the plus age group beyond age 9 was explored in several of the ASAP sensitivity runs, as well as in the base VPA model examined by the Models Working Group and the SARC Panel. Models with extended age structure produced highly uncertain estimates of selectivity on the older ages, and did not alter the perception of the resource.

  - *The length-weight relation clearly shows that fish continue to grow past age-16, yet the age-structure used implicitly assumes no growth after age-9 by assuming a plus group at this age and an associated average weight of fish in this group.*

    - *Considering cod's ability to put on significant weight after age-9, this formulation inherently underestimates SSB and biomass if even a few older fish survive out to ages greater than age-9*

      The mean weight-at-age of the age 9^+ group is neither the weight-at-age of the age 9 fish, nor a straight average of the age 9 – 11^+ fish. It is a numbers-weighted average of all age groups included in the plus group, and thus explicitly accounts for growth of fish out to the maximum age in the catch in any given year. This is illustrated by examining Table A.40 in SARC 53 WP#1.

- *Constant natural mortality (M=0.2) assumption is biologically unrealistic*
    - *This estimate is based on a maximum age of 15-17 years, yet the length-weight relation indicates cod continue growing past this age so it seems unlikely that the fish would [not] continue to grow up until the maximum age.*

      See the previous response with regard to growth. The use of M=0.2 in the assessment is not based on a maximum age of 15-17 years. M=0.2 is the precedent used in previous Gulf of Maine cod assessments. This assumption was reviewed as part of the SARC 53 benchmark assessment and evaluated using meta-analyses to determine the validity of this assumption. The Working Group examined alternate values of M that were based on a suite of life history parameters, such as maximum observed age, growth, and energy allocation strategies. All approaches suggested that the assumption of M=0.2 remained appropriate for this stock (that is, it is biologically realistic).

    - *Additionally, changes in the ecosystem over the last 3 decades would indicate the necessity for a time-varying natural mortality rate and also an age-varying natural mortality*

        - *Juveniles are well documented to inhabit different habitats from adults and predation is much heavier on juveniles (e.g., seal and dog fish predation)*
        - *Lack of 2007 year class recruiting to fishery as predicted in GARM III might be an indication of high predation on age-1 fish meaning a higher M is supported for juvenile cod.*

          It is assumed that the statement above refers to the 2005 year class, not the 2007 year class. These types of ecosystem issues were discussed by the Working Group, but no empirical evidence (i.e., data) was available to support or evaluate their validity. An additional hypothesis, other than predation, variable/different natural mortality, or sampling anomalies (i.e. the cohort was never really that large), is unreported discarding of undersized fish. Unfortunately, there are no data to evaluate this hypothesis either.

Catch

- *Observed catch is split by recreational/commercial and landed/discarded but models fitted to these more 'complex' data sets were deemed too unstable and showed results "similar to the simple (lumped catch) model"*
    - *Although the model might be more statistically stable, it is much less biologically realistic due to the severe differences in selectivity and weight of the commercial and recreational catch.*

      While biological realism in any model is a desired goal, model stability is critical if the sensitivity of the model to alternate assumptions is to be explored. The flexibility to perform these types of sensitivity runs is precisely why the more complex models were abandoned; they did not provide the stability that would allow a full exploration of model sensitivity to alternate assumptions. Nonetheless, the results from the more complex model configurations were very similar to the final BASE model put forward at the SARC. These results support the robustness of the BASE model output and indicate that the catch and weight components were aggregated in a technically appropriate way.

- *Tradeoffs between biological realism and statistical assumptions must be made, however this assessment always errs on the side of statistics instead of actual, proven biology***.**

  The basis for this statement is unclear and not supported. The assessment modeling process placed a premium on making objective evidence (including biological data and all statistical and modeling assumptions) available for examination and comment by all.

- *Fishery selectivity was broken down into two time blocks (pre and post 1991) based on statistical fit, yet no management actions or fishery changes support this choice.*

  This statement is not correct. The decision about where to split the two selectivity blocks was informed by the major changes that occurred in the fishery during the early 1990s. This is clearly documented on page 37 of SARC 53 WP#1:

  > *"An additional selectivity block was introduced beginning in 1989 and several intermediate models were run exploring splits from 1989 to 1994. The period from 1989 to 1994 encompassed major changes in data availability, reporting sources and fisheries management. The model with the 1990/91 split had the lowest objective function and offered improved fit to the age composition in the way of the reduced residual patterning."*

- *Due to lumping of fishery catch across recreational and commercial fleets it is impossible to gain any biological insight into what the estimated selectivity patterns indicate (i.e., is one fleet fishing more heavily on older fish, etc…) and it is impossible to determine the individual effects of each fleet (i.e., is the fishing mortality greater from the recreational or commercial component, which is an important facet when determining possible future management scenarios)*

  - *This is another indication that degradation of the data in order to simply increase statistical fit at the cost of biological insight is inappropriate*

    Model configurations that split the catch into individual components were explored as documented in SARC 53 WP#1 (page 36-37), and had very little impact on the overall assessment results. These earlier models had severe diagnostic problems and did not fit the data very well (e.g. consistent overestimation/underestimation of some catch inputs). A model with a poor statistical fit raises serious concerns about how accurately it is interpreting the data, and whether such a model presents a solid framework on which to base management advice. Models with poor statistical fits are almost invariably rejected by peer-review panels.

    Furthermore, stability is desirable in terms of reducing retrospective patterns. The BASE model has a much more stable basis and a reduced retrospective error compared to alternative, more complex model configurations.

- *Marine Recreational Fishery Statistic Survey (MRFSS) data is used to estimate recreational catch-at-age by imputation based on MRFSS estimates of numbers caught at length and applying the NEFSC survey length-weight equations*
    - *Uncertainty in MRFSS data is well known and estimates in recent years counter data from other sources and common sense*
        - *Vessel trip reports (VTR) from recreational head boats indicate catch estimates 75% lower than MRFSS data.*

          VTR data are only collected from federally permitted vessels. On average, 55% of the recreational catch comes from private vessels with no VTR reporting requirements (SARC WP#1, Table A.25). Additionally, the VTR data provide almost no information on the magnitude of recreational discards.

        - *It is difficult to believe that recreational vessels accounted for the same level of catch (~5500mt) as commercial vessels in 2010.*

          There is uncertainty in the MRFSS numbers, particularly the 2010 estimate, and this is noted in the assessment (page 26 of SARC 53 WP#1). Preliminary MRIP numbers (which were not available at the time of the assessment) suggest that recreational catch may be less than the MRFSS estimate. Sensitivity of the ASAP model to lower estimates of recreational catch have been conducted and —while the modeling results are similar—the net effect is that the MRIP-adjusted ASAP run estimate of the 2010 spawning stock biomass (11,033 mt) is lower relative to the estimate from the base ASAP model (11,868 mt).

- *This is the first time that recreational discard levels have been included in the assessment, however estimates are basically guesses with extremely high and ever increasing levels (~2300mt in 2010).*

  Recreational discards have been reported in previous assessments, but never formally incorporated into the assessment model. The inclusion of all fishery catch components in the updated assessment represents an improvement over past assessments. Estimates of recreational discards are about as certain as type B1 recreational landings. Type B1 landings have been included in previous assessments and, over the course of the assessment period, typically have accounted for more than half of the recreational landings.

- *Discard mortality is assumed 100% for all fisheries because the literature does not provide a comprehensive estimate of mortality rates for all gear type and seasonal combinations*
    - *Most literature proves that discard mortality is less than 100%, which considering levels of assumed discard rates could provide substantial sources of biomass that are being falsely accounted as mortality within the model, yet no sensitivity runs were undertaken to look at the affect of the assumed discard mortality rate.*

      Mortality estimates provided in the literature typically only consider short term survival (<72 hours) in the absence of post-release predation. There is work to suggest that short-term survival may underestimate true post-capture mortality by as much as 50%. Additionally, literature studies indicate a compromised ability of discarded fish to avoid predators post-release (e.g., inability to dive due to swelling of the air bladder, decreased schooling ability, compromised swimming ability, and fatigue). Although it is reasonable to assume that some fraction of the discarded fish survives, the literature available is insufficient to accurately quantify the extent of this survival. While sensitivity runs evaluating alternate assumptions of discard mortality were not explicitly performed as part of the SARC 53 (the VPA bridge building process allowed some evaluation of the impacts), subsequent sensitivity runs have been conducted. The net impact of assuming lower discard mortality lowers estimates of both spawning stock biomass and fishing mortality. Even under an assumption of 0% discard mortality, the Gulf of Maine cod stock remains overfished (virtually no change in 2010 SSB) and overfishing is still occurring at a fishing mortality rate 3 times higher than F_{MSY}.

Surveys

- *Inshore strata of the NEFSC surveys were excluded from the assessment due to inconsistent sampling even though they provide indications of higher age-0 to 2 indices of abundance.*

  Inshore survey data were not used because the inshore areas have been inconsistently sampled. Therefore inclusion of these data would add noise, rather than a signal, to the age 0-2 survey indices used in the assessment. Had the survey captured more age 0-2 fish, recruitment estimates would not necessarily have been have been higher. The model estimates a selectivity ogive for each survey that provides information on the relative selectivity by age in that survey. Had inshore indices been included, this would likely have resulted in higher selectivity-at-age for the NEFSC survey similar to that observed for the MADMF survey. The higher selectivity would have generated the same basic age 0-2 signal in the survey, as observed in the offshore survey data. The best surveys are those that provide precise estimates of population trend, regardless of scale. Higher, but more variable indices degrade the performance of a model and decrease the likelihood of a model providing accurate results.

- *Massachusetts Division of Marine Fisheries (MADMF) surveys are the only reliable estimators of juvenile fish abundance because they survey inshore juvenile habitat, however they are consistently down weighted and the MADMF fall survey was completely removed from the final model.*

  The MADMF spring survey data were not down-weighted. The MADMF spring survey was given the same treatment as the two NMFS surveys. The MADMF fall survey was removed from the final model because: (a) this survey primarily catches only age 0 and 1 fish, and catches of these age groups are highly variable and have been shown to be poor indicators of incoming recruitment (see Fig. A.102 in SARC 53 WP#1); and (b) the removal of the MADMF fall survey did not affect the assessment results.

- *Surveys supposedly cover all areas of major cod catch and accurately represent abundance trends, yet years with high catch rates are consistently considered outliers.*

  The last phrase in this statement is not correct. Years with high catch rates are NOT considered outliers. The ASAP model uses all data, both high and low survey indices, and uses the information contained in the survey data itself (estimates of precision) to determine how well to fit any given year of survey data.

  - *Over the entire time series the NEFSC may cover all areas of major cod catches, but not on a consistent year to year basis (i.e., major areas of cod concentration are sampled sporadically over the last 15 years, however on a year to year basis many concentrations are missed which is likely one contribution to seeing large tows dominate survey catch and cause jumps in catch from year to year).*

    The NEFSC surveys use a random stratified sampling design. Sampling only cod concentrations would impart a positive bias to the fisheries-independent estimates of abundance. The intent of the survey is to achieve a consistent and random sampling of the entire region, not just areas of high fish concentration. This is the primary difference between fisheries-independent estimates of abundance (survey indices) and fisheries-dependent CPUE estimates. If only areas of high cod concentrations were sampled, and the numbers from these data then expanded by assuming that the density of fish in these high concentration areas was the same throughout the stock area, this would imply that cod are plentiful throughout the entire Gulf of Maine. This is obviously not the case, and is why a stratified random sampling scheme in the NEFSC surveys is appropriate.

  - *It is entirely possible that years with high catch rates are actually representative of the population and that the low catches are outliers due to survey locations in areas where cod are not found (e.g., due to the surveys avoiding hard-bottom habitats which cod often inhabit).*

    See the previous response concerning 'outliers.'

- *NEFSC survey catchability is approaching 1.0 and back-transformed catchabilities for R/V Bigelow are above 1.0, indicating that the two research boats are approaching or above 100% efficiency even though almost no catch of fish older than age 7 are reported and area swept estimates of stock biomass approach model estimates of biomass for the entire stock*
  - *Regardless of statistical arguments provided by NEFSC these values indicate poor model performance and should not be treated lightly.*

    A comparison of model-independent and model-based estimates of stock biomass was presented at the SARC and showed close agreement between the two approaches**.** It is incorrect to assume that the high values of survey catchability (q) are indicative of poor model performance. The reason for the high q values was clearly shown during SARC 53 to be a by-product of the expansion scalar used to convert the raw survey indices to area-swept indices of abundance. Alternate expansion factors (which do not alter the assessment results) generate much lower estimates of q. The absolute values of q are a byproduct of expansions performed outside of the model and should not be over-interpreted.

- *Survey selectivity is flat topped and fixed at 100% for ages 6+*
    - *Assessment claims "little biological evidence" for domed selectivity, however allowing for domed selectivity increases SSB by 21%.*

      There is no empirical evidence or statistical basis in the form of model fit to support an assumption of domed-survey selectivity. Nonetheless, several domed survey selectivity sensitivity runs of the ASAP model were explored, and gave results that were generally within the confidence intervals of the base model.

    - *Tagging evidence indicates that shorter tows allow older, larger fish to more easily escape the net than younger fish*
        - *Survey tow times average between 20-30 minutes and therefore present a very high probability that older fish are able to out swim the net and escape.*

          The NEFSC surveys catch a higher proportion of older age fish compared to the fishery, in which tow durations are much longer.

        - *It is therefore more likely that the survey selectivity is heavily domed and that is why few fish older than age-7 are seen in it, as opposed to the current assumption that fishery selectivity is domed (where commercial tows are often upwards of 3-5 hours) and survey selectivity is flat-topped.*

          See previous comment that explains why this statement is not supported by data.

    - *In combination with the survey catchability estimates around 1.0, it appears that there is an issue within the model with the survey time series*
        - *The assumptions used to fit this data consistently err on the side of a pessimistic instead of optimistic stock status (e.g., allowing for domed survey selectivity and bounding catchabilities around .7 would greatly increase abundance estimates).*

          With respect to survey catchability, see the previous response on this subject.

          Bounding catchability at 0.7 would likely increase biomass. However, this would increase biomass beyond the model-independent estimates of biomass derived from the survey data. This result would not be biologically realistic.

<u>Catch-per-Unit Effort</u>

- *NEFSC claim that incorporating CPUE data is not possible due to problems standardizing effort statistics, however the final GARM III model used a CPUE data set and initial ASAP runs made use of this same data set until it was determined that the model was insensitive to its inclusion.*

  Earlier ASAP models, as well as the VPA, did include a LPUE (landing per unit effort) time series covering the 1982 – 1993 period, as this index had been used in previous assessments. The rationale to not extend the LPUE time series beyond 1993 is clearly documented on page 32 of SARC 53 WP#1:

  > *"Given the uncertainty in VTR reported fishing effort since 1994 and the impact of DAS, rolling closures and trip limits on the comparability of LPUEs estimated from 1994 onward with the earlier time series, the time series has not been extended beyond 1993."*

  Since the time series has not been updated since 1993, removal of the time series had no influence on the assessment model results.

- *Recently calculated CPUE data from NEFSC scientists indicate that CPUE has been consistently and drastically increasing since 2000 with large decreases in effort and increases in cod landings, however NEFSC refuses to attempt any exploratory runs with this data set due to the 'difficulty' in incorporating CPUE data.*

  The NEFSC did not refuse to attempt any exploratory runs with this data set. The Data Working Group did not have the opportunity to make any such attempt as the data only became available during the SARC 53 meeting. The CPUE data represent <u>nominal</u> CPUE, which does not account for the major changes that have occurred in the fishery over the last ten years (see comments below for examples), and therefore the data set needs to be standardized prior to inclusion in an assessment model. Cod undergo hyperaggregation when reduced to low stock sizes and a fishery-dependent CPUE index obtained in this situation may not accurately represent overall stock abundance (e.g., Canadian northern cod*)*. Nominal CPUE indices have been constructed for the commercial and recreational fisheries and compared to biomass estimates, but have been shown to be poor indicators of stock abundance.

  - *If old CPUE data sets were possible to incorporate there should be no reason that new data cannot be used.*

    Major input controls (i.e., designed to curtail fishery efficiency) were implemented in the groundfish fishery in 1994 and have since changed frequently. In 1999, for example, trip limits were reduced to 30 lb/day at sea, and then were gradually increased to 800 lb/day over the next five years. Beginning in 1994, marked reductions in the days at sea have occurred including the 2:1 accounting of DAS in western GoM beginning in 2006. There has also been a high rate of exit from the fishery of less profitable vessels, leaving more efficient vessels in the fishery. All of these changes make interpretation of nominal CPUE indices extremely difficult, and standardization of any CPUE time series data for Gulf of Maine cod a daunting challenge.

- *The data shows that the increasing CPUE trend is robust to multiple effort statistics and greatly contradicts the notion that the stock is decreasing as demonstrated by recent survey data.*

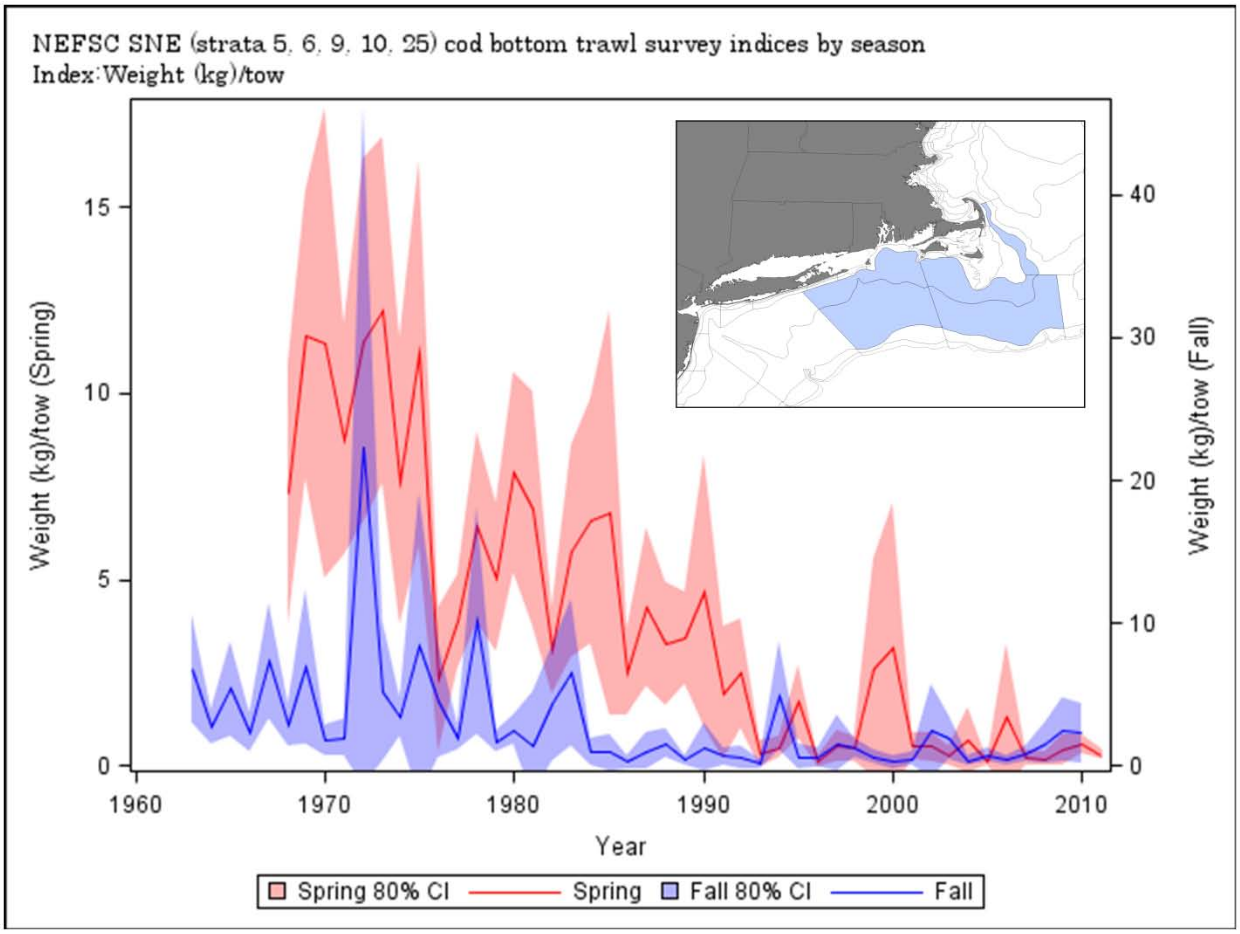
  At face value, the increasing trend in CPUE indicates increases in abundance (if the other concerns with interpreting fisheries-dependent CPUE indices are ignored) in the area where the fishery is occurring. However, because the fisheries occur almost exclusively in the western Gulf of Maine, where the stock is now also concentrated, these CPUE indices provide no information on the abundance of cod in central and eastern Gulf of Maine. If cod were abundant in these latter two areas, this would be reflected in high survey catches in these areas as occurred in the 1970s and/or the fisheries would be operating in these areas. Neither is currently happening.

- *NEFSC argue that this data supports the stock contraction theory (because CPUE will increase as fish concentrate together at smaller population sizes making them easier to catch) and thus do not want to include it because it would inherently force the model to estimate higher biomass.*

  See the previous response explaining why the recent CPUE has not been incorporated in the assessment model.

- *However, taken in context with observations from around New England that cod are being caught in locations that they have not been seen for decades, it indicates the opposite of what the NEFSC is portraying; cod appear to be expanding and higher CPUE is due to an enormous cod biomass throughout the region and not just at small, concentrated locales.*

  In the Gulf of Maine, cod are not showing up in areas where they have been historically abundant such as the central and eastern Gulf of Maine. Fishermen are now reporting cod in high densities in certain areas of southern New England. However, preliminary explorations of biomass trends in the southern New England waters suggests that, like the central and eastern Gulf of Maine, overall biomass in these areas has declined over the past forty years. The region included in these preliminary investigations is shaded in blue in the next figure.



Historical VPA Bridge Assessment

- *Updated data used from the previous assessment (i.e., new length-weight equation, updated weight-at-age, updated catch-at-age, inclusion of discards-at-age, and a revised maturity schedule) have caused a complete change in stock status from the GARM III assessment without changing any of the model formulation or adding new data since 2007 (i.e., the change in historic data since 2007 has changed stock status without adding the last 3 years of data or changing any of the model framework)*
  - *F in 2007 has increased by .1 (21.7%) to F=.56 and SSB has decreased by 14,428mt (42.6%) to SSB=19,445.*

    The major change that affected stock status was how weights-at-age are now estimated. The methods used in GARM III did not fully account for the true weights-at-age of the population. All other data changes resulted in only minor modifications of terminal SSB and F.

Final ASAP Model Results

- *Current estimates of fishing mortality and spawning stock biomass go against all information from the fishery (decreasing effort and increasing CPUE) and management actions (increasingly stringent measures over the last 2 decades).*

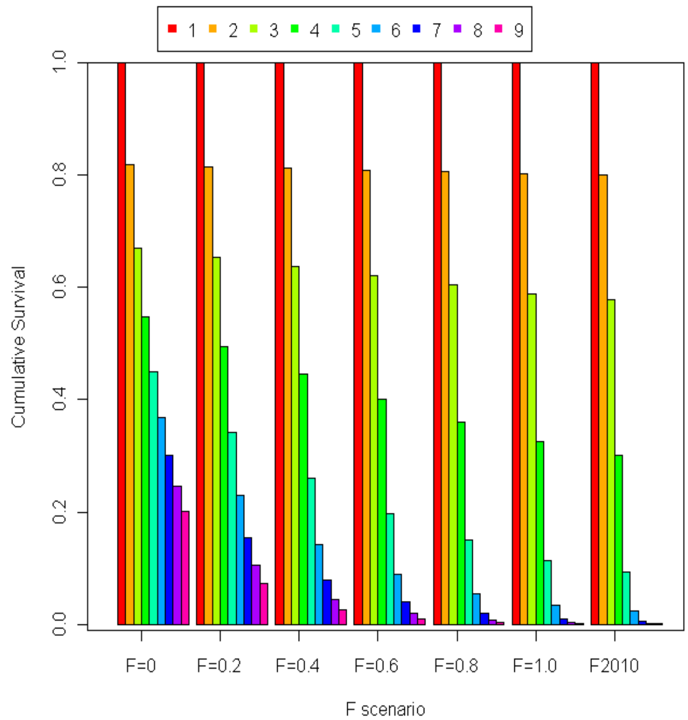
  See previous comments related to CPUE.

- *Current instantaneous fishing mortality rates on fully selected fish of 1.14 indicates that 68% of these age classes are harvested in a given year and total mortality (i.e., including a natural mortality of 0.2) indicates that almost 74% of these age classes die*
  - *Such estimates are absurd and if correct this stock should have collapsed long ago.*

    Spawning stock biomass is only about half of the total biomass, and stocks can be subjected to very high Fs for short period of time (a few years). Existing management measures regulating minimum mesh sizes, minimum retention sizes, and area closures have resulted in a fishery selectivity pattern that allows Gulf of Maine cod to spawn one to two times (on average) prior to capture in the fishery. These spawning opportunities—prior to recruitment in the fisheries—have allowed the stock to withstand high fishing pressure.

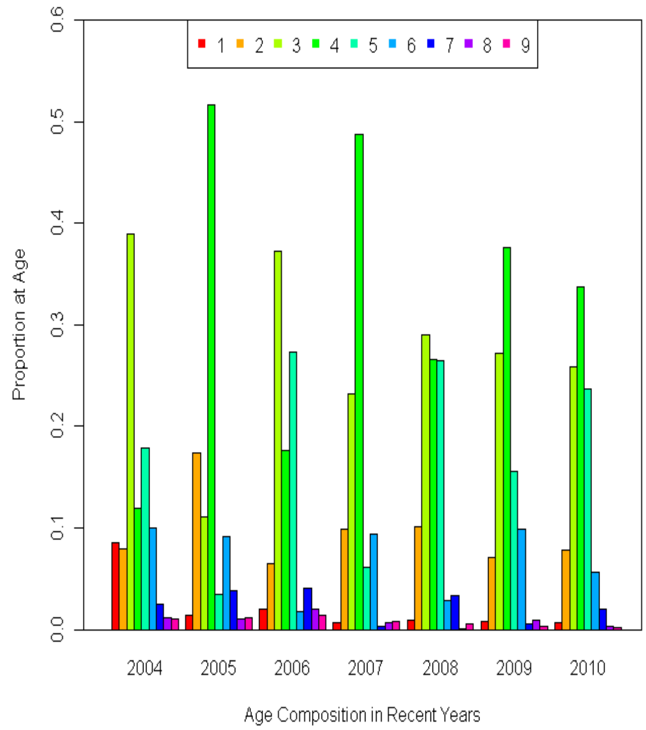
  - *Under this mortality regime only .056% of fish live to age-9, which means that the 2010 age-class of 4.286 million fish would yield only 2418 age-9 fish*

    Although this statement is true, it is important to examine the cumulative survival to ages less than 9, under a variety of F levels, and mindful of the proportion of Gulf of Maine cod mature at age: age 1=9.4%; age 2=28.7%; age 3=61%; age 4=85.9%; age 5=95.9%; age 6=98.9%; age 7=99.7%; age 8=99.9%; age 9+=1.0. The calculations of cumulative survival at age for F=0, 0.2 (F_{MSY}), 0.4, 0.6, 0.8, 1.0, and 1.14 (F_{2010}) indicate that young cod have opportunities to spawn prior to their full selection to the fishery (see the figure below). Due to low fishery selectivity at the youngest (undersize) ages, the difference in cumulative survival to age 3 differs little over a wide range of F values. Cumulative survival is about 30% at age 4 and 10% at age 5 given the estimated F in 2010; however, the decline in survival is precipitous for ages 6 and older.



- *Such results are difficult to believe in the face of current catch compositions and catch rates.*

  A cumulative survival of 0.056% would only be realized if fishing occurred at an F=1.14 for 9 consecutive years. This has not been the case at any point in the modeled time series, and clearly is a circumstance is to be avoided. The figure below depicts the age composition of the Gulf of Maine cod catches for the last 7 years. Ages 3, 4, and 5 dominate the catch in numbers, consistent with the stock age composition results from the assessment.



- *Lack of diagnostics (coefficient of variations) for all model parameters makes it impossible to objectively assess model fit and performance; only CVs are given for selectivity parameters and indicate the model is poorly estimating these parameters.*

  The model input files and software were provided to the SARC 53 Panel to allow the Panel to run the models and evaluate model parameters as deemed appropriate. The Panel requested several additional analyses of model diagnostics, which were all provided to the Panel during the SARC meeting. The CVs on the selectivity parameters are generally less than 0.30, which does not indicate poor estimation. The CVs on the estimates of selectivity on some of the older ages are high, reflective of the limited information at these older ages to precisely estimate selectivity.

- *The use of incremental sensitivity analysis to look at how changing a single assumption at a time affects stock status does not necessarily portray these affects accurately.*

  Without performing sensitivity runs independently, it is impossible to really comprehend how each change to a model (or to the input data) affects the model results.

  - *In reality the base assessment has a number of assumptions that go against the basic biology of the fishery and results should be given showing the effects of changing multiple assumptions simultaneously*

- *For example, what is the effect of allowing domed survey selectivity, bounding catchability at reasonable levels, calculating age-structure out to ~age-16, splitting commercial and recreational catch, incorporating CPUE data, decreasing discard mortality, and decreasing unrealistic recreational catch levels?*

  The available data are insufficient to conduct analyses out to age 16. In most years in the assessment times series, no age information is available beyond age 12 for either the catches or the surveys. As mentioned previously, using a model formulation that splits commercial and recreational catches results in an unstable model that does not allow these sorts of sensitivities to be evaluated. These types of issues were openly, publically, and thoroughly discussed at both the Data Working Group and Models Working Group Meetings.

  Two sensitivity models, which addressed four of the seven concerns noted above, were presented at SARC 53 and gave results that were within the confidence intervals of the base model results.

- *No single change will greatly alter the output of a model, however when numerous assumptions do not reflect reality it makes sense to change all simultaneously and see how the model responds, something that was never considered in the development of the Gulf of Maine cod assessment.*

  No evidence has been presented (i.e., data) to indicate that the model assumptions are not reflective of reality, or that the base model is incorrectly specified. When multiple changes are made to a model, these must be done in a careful and methodical manner so that the impacts of each change can be evaluated. Indeed, such a process was followed in this assessment. When moving from the previous assessment to the current benchmark assessment, many changes were made that better reflected reality. This bridge from the old to the new was not conducted in a haphazard manner by making several changes all at once. Rather, changes were made incrementally so that the impacts of each change could be evaluated. The cumulative effect of these changes was substantial and culminated in a much more realistic final base model.

- *Biological Reference Points are based on an ASAP run back to 1970 (longer time series than the actual assessment) assuming a Beverton and Holt stock-recruit function*
  - *However, analysis by Butterworth and Raddenmeyer (2011) demonstrate that if the model is extended into the late 1960s a decline in recruitment at extremely high stock sizes is present (possibly due to cannibalism on juveniles by adult cod, etc…) indicating that a Ricker style stock-recruit curve is more appropriate and model estimates indicate that GoM cod is NOT overfished.*

    The biological reference points approved by the SARC Panel are NOT based on a spawner-recruit relationship, but rather on long-term projections at $F_{40\%}$ (consistent with the methods used to establish biological reference points in the previous assessment). The methodology proposed by

Butterworth and Rademeyer were not supported by the Models Working Group (of which Butterworth was a member) because (a) age composition data for the fishery are not available prior to 1982 leading to high uncertainty in recruitment estimates; and (b) the stock-recruit function, even when estimated through an extended model time series, is poorly defined. Hence, the biological reference points estimated from such models are uninformative. The decision of the Models Working Group not to use data prior to 1982 in the modeling work was supported by the SARC Panel. Additionally, in a sensitivity exercise, the Models Working Group actually used a SR relationship from a model with a 1970 start date to justify the use of $F_{35\%}$ as opposed to $F_{40\%}$. This decision of the Models Working Group was not supported by the SARC Panel.

With respect to the Butterworth and Rademeyer work using a Ricker-style stock-recruit fit; their model had substantial diagnostic problems, most notably a very strong retrospective pattern which can be problematic for determining stock abundance and making catch advice.

Summary

- *Observations throughout New England indicate cod are expanding their range and not contracting as NEFSC hypothesize.*

  No evidence exists suggesting that Gulf of Maine cod are currently occupying a larger range than historically documented. In fact, the available evidence from the survey and extent of the existing fishery suggests the opposite. Gulf of Maine cod no longer appear to be present in the central and eastern Gulf of Maine, areas where historically they were abundant. Survey data also suggest that cod biomass in southern New England waters has declined substantially over the last 40 years

- *Under-sampling of catch has led to a perceived age-structure truncation that does not match large numbers of old, large cod being caught by commercial fishermen.*

  The available data collected by fishery observers do not support this statement.

- Recreational catch is highly overestimated by MRFSS data.

  There is some validity to this statement, but sensitivity runs conducted to adjust for potential overestimation of recreational catch in the MRFSS surveys show that the assessment results are robust to assumptions about lower recreational catches.

- *Flat topped survey selectivity is unrealistic and allowing the model to estimate domed selectivity causes a large increase in biomass and SSB.*

  A domed survey sensitivity run was conducted and presented in the final report. The assumption of a domed survey selectivity pattern results in slightly higher estimates of SSB and lower estimates of F, but these estimates are within the confidence intervals of the base model results, and do not alter the stock status determination. It should be noted that there was no model support for domed selectivity (i.e., allowing a domed survey selectivity did NOT improve the overall fit of the model to the data). More importantly, there are no data to support a domed survey selectivity, and what data do exist support higher selectivity in the survey relative to the fishery.

- *The purposeful avoidance of exploratory analysis of recent NEFSC CPUE data within the assessment indicates a lack of objectivity by the assessment scientists as this data clearly counters recent trends in NEFSC survey abundance and indicates an expansion of cod biomass in the Gulf of Maine.*

  There was no purposeful avoidance of an exploratory analysis of CPUE data. No source of CPUE data was brought forward to either the Data or Models Working Group. The first mention of recent CPUE data was brought up during the SARC Meeting, and the SARC Panel (not NEFSC scientists) determined that it was not appropriate to introduce these data at that time. The basis for this determination was that (a) the CPUE data had not been standardized to account for changes in fishery efficiency over time, and (b) the CPUE data had not undergone the same amount of review that all other input data received during the Data and Models Working Group Meetings

- *Model results go against all recent management actions and observed biology and are based solely on noisy, unreliable surveys (since catch trends do not reflect the biomass under a hard total allowable catch system, they simply reflect management expectations regarding stock abundance assuming the TAC is fully harvested; CPUE is the only real indication of biomass levels that can be garnered from catch data in this instance).*

  These statements have been addressed and rebutted elsewhere in this document.

- *Tradeoffs between biological realism and statistical assumptions must be made, however this assessment always errs on the side of statistical fit instead of actual, proven biology resulting in many biologically unrealistic modeling assumptions often causing a more pessimistic view of stock status.*

  To imply that the modeling assumptions led to a more pessimistic view of stock status ignores the breadth of sensitivity runs presented during the SARC 53 Meeting. There were a total of 14 ASAP sensitivity runs presented during the SARC. Exactly half of the sensitivity runs provided more optimistic perceptions of stock status relative to the base model. The results from only two of the sensitivity runs fell outside the confidence limits of the base model; one was above, one was below. The consistency of the sensitivity runs provides additional confidence in the assessment results.