PEER REVIEW OF WILBERG *ET AL*, "EVALUATION OF ACCEPTABLE BIOLOGICAL CATCH (ABC) CONTROL RULES FOR MID-ATLANTIC STOCKS"

Prepared for:
Mid-Atlantic Fishery Management Council

by
Dr. Douglas S. Vaughan
Chairman, Review Panel

<u>Review Panel</u>
Dr. André E. Punt
Dr. Chris Legault
Dr. James Berkson

January 20, 2012

EXECUTIVE SUMMARY

The Mid-Atlantic Fishery Management Council (MAFMC) entered into a contractual agreement with the University of Maryland (UMD) to conduct a management strategy evaluation (MSE) of various ABC control rule approaches that were under consideration by the MAFMC for its managed species. The MAFMC established the peer review procedure to obtain an independent peer review of the MSE Final Report submitted by UMD to the MAFMC in August 2011. The panel consists of three reviewers plus the chair. Three terms of reference (TOR) were provided to them for the review (see section headers below). As noted by Dr. Legault (see Appendix), the report was not structured to reflect the TOR since the TOR were developed later for this formal review. The three independent reviews were provided to the Review Panel Chair for coordination and summarization. Their complete reviews are found in the Appendix.

OVERVIEW BY PANEL

There was considerable agreement by the Review Panel as to their concerns about the Wilberg et al. report. In general they believed that it provided much useful information, but that there were weaknesses in the analyses that limited the ability to draw definitive conclusions useful for management purposes. Specific comments by the Panel Reviewers are quoted below:

Dr. Punt begins with

> "*The evaluation of control rules is fairly thorough and could be helpful for Council decision making. A range of data-poor methods was considered. However, in my opinion, some aspects of the analyses weaken the ability to derive definitive conclusions regarding ABC control rules for the MAFMC."*

Dr. Legault begins his review comments with the following:

> *"Overall, the authors have conducted a wide range of analyses to address their contractual obligations. In my opinion, the authors have taken the first step in a long process of evaluating ABC control rules. This work will provide a base for future work to build upon and provide guidance for future evaluations of ABC control rules."*

and concludes with his final comment under TOR #3.

> *"So while the modeling approach is sound and the statistical treatment of results is generally appropriate, for those results which were considered, the lack of evaluation of trade-offs prevents me from concluding that this work can be used for management purposes. I think it is a good starting point for discussions with the Council and SSC regarding how to evaluate trade-offs, and which trade-offs are of most interest. However, until the trade-offs are fully integrated into the analysis, it is insufficient for management purposes."*

Dr. Berkson begins his review with the following statement:

> *"Overall I find this study to be an interesting pilot study investigating the potential effects of alternative control rules on simulated stocks. It is far too incomplete, in its methods, results presentation, results interpretation, and discussion to be used for management purposes in its present state. The work would need to be greatly expanded and improved to move beyond being a pilot study to being useful for management."*

and concludes with:

> *"In conclusion, due to time constraints, I have pointed out many of the problems I've noted with this work. I do see much of value. It appears that the work could be developed more fully, incorporating the comments made by myself and the other*

reviewers. This would make the simulations much more effective, realistic, and defensible. This is not a trivial amount of work. I view this project about one-third of the way done, before it could really be useful for management."

In summary, the results presented in Wilberg et al. should be viewed as preliminary in nature. Interpretations based on the incomplete nature of this work (e.g., study "not fully crossed" as stated by **Dr. Punt**) are not sufficiently general to have broad application. There are pieces of the puzzle that may be useful, but only with care, by the Council and SSC. The results should not be considered definitive.

In the material to follow, I have attempted to pull out some common themes from the three panel reviewers organized by the three Terms of Reference. Because their reviews are included as an Appendix, I have not tried to include every comment. Please look at the Appendix for a more complete explanation by each of the panel reviewers. I referred to **Dr. Punt's** outline when referencing his material, while I refer to numbered paragraphs for **Drs. Legault** and **Berkson**, the former within each TOR, and the latter overall.

TOR #1. EVALUATE THE STUDY DESIGN AND METHODOLOGY, WITH PARTICULAR REFERENCE TO MODELS AND MODEL ASSUMPTIONS, SELECTED TO EVALUATE ABC CONTROL RULE PERFORMANCE

All three reviewers felt that the description of the methodology was inadequate. Lack of detail was highlighted by all of the panel reviewers in various places in the report. **Dr. Punt** emphasized this generally (B1) suggesting that there was not "sufficient detail ... to allow a colleague to come close to repeating the analyses." **Dr. Legault** refers to a few places where he found or implies inadequate detail (see paragraphs 2-4 under TOR #1). **Dr. Berkson** does not believe that the "ORCS approach is fully described or applied" (paragraph #4). In particular, he does not believe that the flexibility of this approach was acknowledged. He later suggests that the data-poor results would be different if ORCs were applied differently (paragraph #12). As noted by **Dr. Berkson**, this approach can be highly flexible, but I would also note that such flexibility can provide difficulties in conducting simulations comparing a wide range of approaches. **Dr. Berkson** also states that much of the methods section is not clear (paragraph #5), and asks why the simulation populations had different trajectories in data-poor vs data-rich scenarios (paragraph #6). This last comment seems to relate to a concern by **Dr. Legault** (paragraph #8). He suggests that the data-rich simulations starting at a rebuilt stock size may lack generality for use in other regions besides the Mid-Atlantic. This may be beyond the mandate for the MAFMC, particularly with limited funding. **Dr. Berkson** also felt there was insufficient discussion comparing methods throughout the document, and that results would make more sense if put in proper perspective (paragraph #13).

All three panel reviewers suggest that the testing of methods is not a complete design and as result interpretations are difficult, and it may not be possible to evaluate certain comparisons. **Dr. Punt** makes this point explicitly when listing his major concerns about the report (Executive

Summary #1). He follows up on this under TOR #1 (B2). He was also concerned that different simulation set-ups were used for the data-poor and data-rich methods. After noting there are three parts to the study design, **Dr. Legault** suggests that each part appears to have been done independently with little carryover of what was learned in one part to later parts (first paragraph under TOR #1). This reviewer does note that the data-rich situation considered only a single life history and exploitation history due to inclusion of the stock assessment into the simulation framework, because of greater amount of time required. In a similar vein, **Dr. Berkson** suggests that the study concerning summer flounder was not well developed and did not seem to fit well with the second part of the study (simulations) and should have been a separate report (paragraph #2). This reviewer does not draw quite the distinction between the data-poor and data-rich situations, and notes that even in so-called data-rich situations, data can be uninformative and require a data-poor control rule (paragraph #3). I would re-iterate here that Wilberg et al. provided reasoning for the reduced level of analyses for the data-rich situation; i.e., computational intensity required for these. Clearly more time and funding would allow more work and wider range of situations to be considered.

All three panel reviewers highlighted different aspects of how uncertainty and mis-specification were investigated for summer flounder and simulated data. **Dr. Punt** notes the importance of this for the summer flounder study in Section A.1(a) and (c). He notes that uncertainty in assigning data-poor cases to the appropriate level of productivity and depletion categories need to be considered (B.3.a), that the CV of 0.2 used seem rather low (B.3.b), and that no account was taken of uncertainty in catches (B.3.c). He questions that there are likely to be more sources of uncertainty associated with the assessment than given in Table 4 of the report (B.5). **Dr. Legault** notes that the authors addressed the mismatch in *M* between the assessment and simulation model, but expresses concern that estimates are not necessarily centered on Mbar. **Dr. Berkson** wondered why time-varying *M* was the only mis-specification modeled (paragraph #10).

Dr. Berkson had a number of comments related to model realism. He thought it highly unlikely that 95% of the exploitable biomass would be caught, and not continue for multiple years. The current model would allow this to continue for four years in a row until re-estimation of the ABC (paragraph #9). He also refers to the biased scenarios that assumed status and productivity were overestimated every four years for 20 years, and does not find this realistic (paragraph #9). He also found a situation in the data-rich scenarios that if the stock becomes overfished, the same control rule is applied to be unrealistic (paragraph #11).

Dr. Legault offers the following recommendations for expanding this study:

- changes over time in biological parameters such as M, weights at age, and maturity, especially as a function of stock abundance,
- changes over time in fishery selectivity,
- environmental impacts on recruitment, and
- starting conditions for application of the ABC control rules.

TOR #2. EVALUATE THE STATISTICAL TREATMENT AND INTERPRETATION OF MODELING RESULTS

Again, there were some common threads among the responses of the reviewers to this Term of Reference. Both **Drs. Punt and Legault** raise comments concerning the "foregone yield statistic". **Dr. Punt** suggests this statistic be defined to enhance comparability and the simulations could be run long enough such that foregone yield would be positive (C.1.a), while **Dr. Legault** makes this same point (#4). Meanwhile **Dr. Punt** believes that discounted revenue should have been included among the performance statistics. He is also concerned that the probability of overfishing is based on the perception from the stock assessment, and should have been based on values from the operating model.

Dr. Legault provides some additional points. He believes that P* was incorrectly defined (paragraph #1). At the same time **Dr. Berkson** thinks that P* is used here as a summary statistic, and does not think it is appropriate as an output statistic (paragraph #14). **Dr. Berkson** also wondered why SPR-based reference points were used throughout (paragraph #7). MSY-based reference points could be calculated, since the population was simulated.

Dr. Legault has some concerns about the use of the ΔS metric used in the data-poor simulations because they could be highly influenced by recruitment events.

Dr. Punt found F/F_{lim} to be a more useful statistic than the probability of overfishing in Figures 7-11 (C.2). He did not find any specific conclusions regarding how easily different control rules can be implemented given the standard assessment output (A.2). He suggests an alternative approach would be to "tune" each control rule so that they are all equivalent on one metric, in this respect, he suggests the probability of overfishing for one of the data-poor scenarios be the way the control rules could be tuned (A.3).

TOR #3. DETERMINE IF THE CONCLUSIONS ARE SUPPORTED BY THE STUDY RESULTS AND IF THE INFORMATION IS OF SUFFICIENT SCIENTIFIC RIGOR TO USED FOR MANAGEMENT PURPOSES

In addition to bottom-line comments discussed in the OVERVIEW, **Dr. Punt** raised three particular concerns: 1) it was not clear whether there are any stocks managed by the MAFMC for which a complete catch history is available, given the key conclusion that DB-SRA was best data-poor method (D.1); 2) there is a need for the analyses for P* to consider multiple life histories and current statuses, not to mention levels of data quality (D.2); and 3) the preferred CV value rests on trade-off between long-term yield and the value of rebuilding; however the status of the data-rich stock as start of the first year the ABC control rule is applied is not given (D.3).

Dr. Legault had a major concern that the *"authors did not fully evaluate trade-offs between catch and overfishing. If the only goal is to prevent overfishing, just set catch to zero and you are* *done. Without the evaluation of trade-offs, the most conservative options will always be determined to be the best. There is no mention of the pain of initial cuts required by conservative control rules compared to less conservative ones. Even if mean catch over many years is the same, the industry may not survive the initial cuts required in some of the most conservative approaches, a claim made repeatedly in some regions."* (paragraph #1)

Other concerns by **Dr. Legault** include: 1) there was no discussion regarding additional information that could be used to inform data-poor situations to evaluate whether the control rule was working as expected or not (paragraph #2). 2) the authors conclude from the data-rich simulations that using larger CV's in the Mid-Atlantic ABC control rule always performed better than using smaller CV's, is there is a limit to this conclusion (paragraph #3)? 3) is there a lesson to be learned from the data-poor simulations that jacking up *M* to "fix" an assessment could require more conservative control rules (paragraph #4)? 4) the final paragraph of the paper claims that the Restrepo rule may be useful in highly uncertain conditions because it was the only one to have P* less than 0.5 for the slow life history and overexploited harvest pressure (true statement based on Table 9), however, it ignores the trade-off being made in the mean catch for this control rule, which is typically on the order of one third the other control rules. If avoiding overfishing is the only goal, then the control rule of zero catch will always be preferred. It is the trade-offs between catch and overfishing that the authors describe earlier in the paper that are being ignored here (paragraph #5).

Recommendation (Dr. Legault): *For example, mean length of fish caught, CPUE, and time to reach some fraction of the quota could all be used as additional information* (to inform data-poor situations). *While this would be difficult to do within a simulation framework, this sort of information could be artificially provided through noisy indications of change in direction of the stock size to potentially modify the application of control rules. This might be a fruitful avenue for future simulation work.* (paragraph #2)

APPENDIX

REVIEW OF WILBERG *ET AL*. ("EVALUATION OF ACCEPTABLE BIOLOGICAL CATCH (ABC) CONTROL RULES FOR MID-ATLANTIC STOCKS)

Dr. André E. Punt
University of Washington
Seattle, WA

EXECUTIVE SUMMARY

The evaluation of control rules is fairly thorough and could be helpful for Council decision making. A range of data-poor methods was considered. However, in my opinion, some aspects of the analyses weaken the ability to derive definitive conclusions regarding ABC control rules for the MAFMC. My major concerns with the report are:

1. The testing of methods is not fully-crossed. Applying as many methods as feasible in each test would have made interpreting the results much easier and would provide the Council with much additional information. For example, a 40:10 control rule may perform better than one of the P* rules, but this cannot be evaluated given the simulation results presented. Similarly, different simulation set-ups were used for the data-poor and data-rich methods, which makes comparisons difficult.
2. The analysis based on summer flounder aimed to comment on how easily different control rules can be applied, but the report does not draw any conclusions in this regard.
3. The performance statistics for the simulation study should have included discounted revenue and the biological output statistics should have been based on the true (operating model) values and not those estimated by assessment model in the final year of the assessment.

There were problems with the document submitted to the review panel and my comments are based on the ICES paper version of the report.

A. INTRODUCTION AND GENERAL

A.1 This review was based on the reviewing the document "Evaluation of Acceptable Biological Catch (ABC) Control Rules for Mid-Atlantic Stocks" by Michael J. Wilberg, Thomas J. Miller, and John Wiedenmann. The document evaluated ABC control rules for data-rich and data-poor stocks (with a focus on the latter). The evaluation was based on applying the rules to summer flounder and to simulated data using closed loop simulations. The application to summer flounder aimed to address three questions:

(a) better understand how uncertainty estimates produced from an assessment may be influenced by changing model assumptions;
(b) determine how easily different control rules can be implemented given the standard assessment output; and
(c) quantify the variability in the ABC and its effect on the stock.

The aim of the simulation study was to explore a range of ABC control rules and to evaluate the probability of overfishing (defined as here as the F exceeding MFMT or the catch exceeding the Overfishing Level, OFL), becoming overfished, average catch, inter-annual catch variation, etc.

A.2 Although most of the aims are addressed in the report, I could not see any specific conclusions regarding how easily different control rules can be implemented given the standard assessment output.

A.3. The ABC control rules are applied 'as defined'. An alternative approach would be to 'tune' each control rule so that they are all equivalent on one metric (say the probability of overfishing for one of the data-poor scenarios). This would reduce the difficulties comparing among the ABC control rules.

B. TOR1. EVALUATE THE STUDY DESIGN AND METHODOLOGY, WITH PARTICULAR REFERENCE TO MODELS AND MODEL ASSUMPTIONS, SELECTED TO EVALUATE ABC CONTROL RULE PERFORMANCE

B.1. I generally found the description of what was done to be inadequate. One expects that sufficient detail would be provided to allow a colleague to come close to repeating the analyses. This is not the case here. The lack of detail (e.g. the effective sample sizes for creating the catch- and survey proportions-at-age; the extent of process error, observation error, and variation in M) means that I have several queries on the methods themselves.

B.2 An ideal experimental design would have applied all of the ABC control rules to summer flounder and in the simulations. This was not done, and the reasons for doing so are not clear to me.

(a) The "eyeball" approach was applied to summer flounder, but not in the simulation study. This was unfortunate because making this approach operational (e.g. by testing for stability in trends in abundance indices and catch) is a key difficulty when applying this approach. Similarly, no reason is given for not applying ORCS and Restrepo method to the data for summer flounder.

(b) The production model, 0.75 F_{lim} and 40:10 rules were only applied to the actual summer flounder data. I do not see why they could not have been considered as part of the data-rich simulation.

B.3 To be ideal, simulated control rules should not know anything which would not be known in the real world. Areas where the report was may have failed in this respect were:

(a) In a data-poor case, the probability of assigning a stock to "high", "median" and "low" productivity will be uncertain and may be wrong. It seems that 'inflation' was applied to these methods (their results differ between Tables 6 and 9 for example), but how this was done is not clear. Similarly, there should be error in assigning stocks to depletion categories (Eqns 2.1 and 2.2).

(b) The distributions for DCAC and DB-SRA are centered on the true values for M, F_{lim}/M, S_{targ}/S_0 and Δ for many of the analyses. However, the CVs of 0.2 for these quantities seem rather low given the results of meta-analyses I have seen.

(c) No account was taken of uncertainty in catches for the data-poor methods (Eqn 2.13 only appears for the data-rich case). This is particularly an issue for DB-SRA, which requires a (potentially) a long time-series of catches.

B.4. Using catches up to year t-1 in control rules such as ORCS or the Restrepo rule can lead to the "ratchet effect" where catches are continually decreasing. This has been recognized in the control rules used in Australia, and has led to control rules which are based on catches during a fixed period in time (e.g. Little *et al*. 2011)

B.5. I am not familiar with the assessment of summer flounder. However, I would have thought that there would be more sources of uncertainty associated with this assessment than are given in Table 4. For example, I would have expected that there would be uncertainty regarding (a) M, (c) retrospective patterns (sense Ralston *et al*. 2011), and (c) uncertainty about F_{lim}, noting that one source of scientific uncertainty whether the proxy definition of F_{lim} is correct. Ignoring this latter source of uncertainty will lead to under-estimation of the probability of overfishing.

C. TOR2. EVALUATE THE STATISTICAL TREATMENT AND INTERPRETATION OF MODELING RESULTS

C.1. Most of the performance statistics used to summarize performance are standard when conducting management strategy evaluations. My thoughts regarding the performance statistics were:

(a) The forgone yield statistics should be consistently defined to enhance comparability. The approach applied for the data-poor case would seem most appropriate as it is an approximation to the "expected value of perfect information". By definition, had the simulations been run long enough the summed "forgone yield" would be positive.

(b) The performance statistics related to the probability of overfishing appear to be based on the perception from the stock assessment. This is wrong – this probability should have been based on the values in the operating model. A performance statistic which assesses how often the assessment correctly detected that overfishing had occurred would have been interesting, but it is the probability of actually overfishing which is of primary interest.

(c) Discounted revenue should have been included in the set of performance statistics because strategies such as CV=1.2 may achieve adequate catches over the longer term, but at the cost of lower short-term catches.

C.2 The focus on F-ratio and probability of overfishing in Figures 7-11 was not ideal because logically these measures are highly correlated. I found F/F_{lim} to be more useful than the probability of overfishing in terms of comparing approaches.

D. TOR3. DETERMINE IF THE CONCLUSIONS ARE SUPPORTED BY THE STUDY RESULTS AND IF THE INFORMATION IS OF SUFFICIENT SCIENTIFIC RIGOR TO USED FOR MANAGEMENT PURPOSES

D.1 A key conclusion is that DB-SRA is the best data-poor method. However, it is not clear from the report whether there are any stocks managed by the MAFMC for which a complete catch history is available.

D.2 The results for the data-rich case are only based on one life-history strategy which makes evaluating any conclusions regarding the appropriate CV difficult. I understand that the analyses are fairly computationally intensive, but final conclusions regarding P* need to consider multiple life histories and current statuses, not to mention levels of data quality (which cannot be evaluated because the sampling CVs, and multinomial sample sizes are not given).

D.3 The conclusion that CV=1 / 1.2 is preferable to lower CVs rests on the trade-off between short-term and long-term yield and the value of rebuilding. Unfortunately, the document does not indicate the status of the stock at the start of the first year the ABC control rule is applied. Figure 23 suggests that large changes in biomass (up to a factor of 2.2) occur (this seems inconsistent with Figure 17, but the metrics in the two figures may be different) which seems to suggest that the stock was probably below S_{MSY} when the ABC control rule was first applied.

E. OTHER

E.1 Was steepness estimated when applying the assessment model? Also, were the CV and effective sample sizes assumed known when conducting the assessment? Why was a control rule based on the estimated CV for the OFL not considered?

E.2 It would have been interesting to know how biased / imprecise the results from the assessment were. For example, it is unclear why the average F-ratio exceeds 1 in median terms in some years in Figure 23.

E.3 It was not clear to me why the results for the median catch control rule change when there is imperfect knowledge of stock status and life history (e.g. compare Tables 6 and 9) because median catch should not depend on stock status.

E.4 I would have liked to have seen Table 4 annotated by the estimated OFLs and ABCs.

E.5 Including the OFL in Table 5 would have enhanced understanding. Why in Table 5 is the realized P* 0.29 for the diagonal elements in the last three rows. I expected it to be 0.4 as that is the council-selected P* value.

E.6 Why does DCAC exhibit catch variability? I guess this is because it was updated (i.e. the year to which Δ was assumed to apply to changes over time). This is not how use of DCAC has previously been discussed (e.g. by Alec MacCall).

References

Little, L. R., Wayte, S. E., Tuck, G. N., Smith, A. D. M., Klaer, N., Haddon, M., Punt, A. E., Thomson, R., Day, J., and Fuller, M. 2011. Development and evaluation of a cpue-based harvest control rule for the southern and eastern scalefish and shark fishery of Australia. *ICES Journal of Marine Science* 68: 1699–1705.

Raltson, S., Punt, A.E., Hamel, O.S., DeVore, J.D. and R.J. Conser. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fish. Bull. 109: 217-231.

Review of Wilberg, Miller, and Wiedenmann "Evaluation of Acceptable Biological Catch (ABC) Control Rules for Mid-Atlantic Stocks"

Dr. Chris Legault
NOAA Fisheries NEFSC
Woods Hole, MA

This manuscript was written as part of a contract between the University of Maryland and the Mid-Atlantic Fishery Management Council to conduct a management strategy evaluation to evaluate various acceptable biological catch (ABC) control rule approaches. The terms of reference (TOR) for this review were created after the report had been completed, and thus the report is not structured to explicitly address these terms of reference. Overall, the authors have conducted a wide range of analyses to address their contractual obligations. In my opinion, the authors have taken the first step in a long process of evaluating ABC control rules. This work will provide a base for future work to build upon and provide guidance for future evaluations of ABC control rules. I note a number of places in my responses to the TOR for this review where future work could be helpful. I was surprised by the numerous typographical errors found throughout the manuscript and hope this is just a reflection of the shortened time period of the study due to funding limitations.

TOR 1: Evaluate the study design and methodology, with particular reference to models and model assumptions, selected to evaluate ABC control rule performance.

The study design consisted of three parts: 1) an evaluation of applying a range of control rules to a specific stock assessment, summer flounder; 2) simulation analysis for data-poor situations; and 3) simulation analysis of data-rich situations. Each part appears to have been conducted independently, with little evaluation of lessons learned from one part informing the other parts. In fact, the conclusions from parts 1 and 2 regarding the use of data-poor control rules appear contradictory, with part 1 concluding they are typically associated with high probabilities of overfishing and part 2 concluding that the most-conservative data-poor control rules should prevent overfishing in nearly all situations.

The evaluation of applying a range of control rules to the summer flounder assessment first evaluated a range of model configurations to examine the robustness of the terminal year F and SSB estimates. These were found to be quite robust, as in the actual stock assessment, due to the strong signals in the data. The authors next estimated deterministic biological reference points and applied thirteen ABC control rules. Details were not provided in the text on how these control rules were applied to generate Table 5. My specific questions are in which year were the catch and OFL predicted in the data-rich approaches and what, if any, assumptions were made for intervening years between the end of the assessment and the catch year? I found the production model results in this section particularly informative given some recent claims by others that simple models perform better.

The same simulator was used for both the data-poor and data-rich simulations. The age structured simulation model used standard population dynamics equations, except for not using a plus group for the oldest age. However, the use of relatively large maximum ages for each scenario should mitigate against this causing any problems. All scenarios considered had the fishery selectivity ogive to the right of the maturity ogive, so that the populations could withstand high fishing mortality rates. The parameter settings for the three life histories used in the data-poor simulations appear reasonable, although I did not see where sigmaR was defined.

The simulation analysis for data-poor situations was set up as a factorial design to examine three life histories, three levels of exploitation, three levels of catch data collection, and two scenarios of expert judgment for those control rules which required it. This produced a lot of data that the authors summarized in Tables 6-11 and Figures 7-15. I was confused by the bold font used in the tables. At first I thought it denoted situations where the overfishing probability was greater than 0.5, as stated in the text, but there were numerous examples where this criteria was met but the results was not bolded and others which were bolded that did not meet the criteria. Am I missing something?

Constant F for the three exploitation scenarios was applied for twenty years before applying the ABC control rules in the data-poor simulations. This had the interesting effect of causing the catch for the over-exploited and under-exploited scenarios to be approximately equal in both the fast and medium life histories. The slow life history had not reached equilibrium after 20 years of constant F. If the desire was to start all three life histories in equilibrium, then a longer period of constant F is required for the slow life history. It is not clear to me whether this use of constant F in these simulations is the cause of the difference in conclusion between parts 1 and 2.

The simulation analysis for the data-rich situation considered only a single life history and exploitation history due to the much greater amount of time required by the inclusion of stock assessments within the simulation framework. This inclusion of the assessment model is good because it allows for both bias and imprecision to enter during application of the control rules. However, due to the manner in which the simulation was conducted, I did not see any way for bias to enter except when M was misspecified. This means that only imprecision would be expected to result from the stock assessment. In which case, why not skip the stock assessment and just add noise to the simulated stock size? This would allow quicker run times and evaluation of more scenarios. It is only through mismatches between the assumptions made in the assessment model and those used in the simulation that the use of the simulation model is warranted. There are certainly many types of mismatch that could be explored, such as the shape of the selectivity curve, the form of the stock recruitment relationship, and changes in biological parameters such as weight at age over time. Formulating scenarios based on hypotheses regarding potential mismatches in Mid-Atlantic stock assessments would be beneficial in the future.

The authors did address the mismatch in M between the assessment and simulation model.

However, equation 2.14 is not necessarily centered on Mbar, in fact all the values in the time series can be either above or below the value of Mbar. The runs when the average M of the time series is far away from the value used in the assessment would be expected to be biased, not just imprecise. This is a harder test of robustness than just imprecise input and should be characterized as such in the text.

Starting the data-rich simulations at a rebuilt stock size may be appropriate for many Mid-Atlantic stocks, but is a limitation if these results are to be used by other regions. The good starting stock size means there is a lot of room for overshooting OFLs without driving the stock to an overfished or depleted level. This means the trade-offs will be skewed towards catch and away from stock size considerations. This also allows the conclusion that similar catch levels can be achieved over a range of control rules because there is little chance of driving the stock so low that recruitment fails. My guess is that starting the population in a much worse condition would produce different results and may cause a different impression of the trade-offs.

Areas for future research:

- changes over time in biological parameters such as M, weights at age, and maturity, especially as a function of stock abundance
- changes over time in fishery selectivity
- environmental impacts on recruitment
- starting conditions for application of the ABC control rules

TOR 2: Evaluate the statistical treatment and interpretation of modeling results.

The desired distribution of probabilities of overfishing under a given P* was incorrectly defined, in my opinion. A given P*, say 40%, should result in a distribution of F with 40% of the probability mass above the reference F and 60% below. The authors appear to be using a different metric whereby a distribution will all of its probability mass for F below the reference F is declared a good control rule under the assumption that preventing overfishing is the main goal. This does not seem logical to me.

The authors define overfishing two different ways and are not always clear about which definition is being applied. On page 3 first paragraph, overfishing is defined as F exceeding a reference F, while on page 3 second paragraph, overfishing is defined as catch greater than the overfishing limit (OFL). These two are not necessarily the same thing when evaluated at different times. For example, an assessment can be used to set a future OFL according to a specified reference F, but later assessments including more data can conclude that an actual catch of less than the OFL did in fact result in F exceeding the reference F. Did overfishing occur in this situation? This question is still being resolved in other areas, so the authors should not be admonished for this dichotomy. However, it would help in the future to explicitly define which type of overfishing is evaluated in every situation.

It seems to me that the delta S metric used in the data-poor simulations could be highly influenced by recruitment events at the start and end of the time series because it only compares the SSB in the last year to the SSB in the first year. Using an average SSB over a small number, say three to five, years for both the start and end of the time series would reduce this dependency of results on recruitment pulses.

The metric perceived forgone yield is an important one for a single forecast, but it is not clear to me the meaning of this metric when it is accumulated over time because forgone yield this year should allow additional population growth this year and thus higher OFL next year, all other things held equal. My confusion is most clearly seen in the second Figure 21 where forgone yield increases dramatically with increasing CV while mean catch is almost constant over all CV's in Figure 18. If forgone yield is a "pain" metric to the industry, then the large increase as CV increases would argue against using higher CVs, yet this is not discussed.

TOR 3: Determine if the conclusions are supported by study results and if the information is of sufficient scientific rigor to be used for management purposes.

The authors did not fully evaluate trade-offs between catch and overfishing. If the only goal is to prevent overfishing, just set catch to zero and you are done. Without the evaluation of trade-offs, the most conservative options will always be determined to be the best. There is no mention of the pain of initial cuts required by conservative control rules compared to less conservative ones. Even if mean catch over many years is the same, the industry may not survive the initial cuts required in some of the most conservative approaches, a claim made repeatedly in some regions.

There was no discussion regarding additional information that could be used to inform data-poor situations to evaluate whether the control rule was working as expected or not. For example, mean length of fish caught, CPUE, and time to reach some fraction of the quota could all be used as additional information. While this would be difficult to do within a simulation framework, this sort of information could be artificially provided through noisy indications of change in direction of the stock size to potentially modify the application of control rules. This might be a fruitful avenue for future simulation work.

The authors conclude from the data-rich simulations that using larger CV's in the Mid-Atlantic ABC control rule always performed better than using smaller CV's. It is not clear to me if there is a limit to this conclusion. Meaning, would a CV of 5 be better than a CV of 2 which is better than a CV of 1.2? This would seem to imply that lower catch advice is always better, with the extreme case of zero catch having the best "performance" when only overfishing is considered. Or is this a particular feature of the Mid-Atlantic ABC control rule whereby using P* of 0.40

means that there is always some catch no matter how larger the CV is assumed about the OFL (see Figure 1). If P* was set at 0.49 or 0.20, how different would the results be for the simulations examined? Perhaps plotting CV on the x-axis and proportion of OFL buffer on the y-axis for a given P* value would demonstrate this point.

Is there a lesson to be learned from the data-poor simulations that jacking up M to “fix” an assessment could require more conservative control rules?

The final paragraph of the paper claims that the Restrepo rule may be useful in highly uncertain conditions because it was the only one to have P* less than 0.5 for the slow life history and overexploited harvest pressure. This is a true statement based on Table 9. However, it ignores the trade-off being made in the mean catch for this control rule, which is typically on the order of one third the other control rules. If avoiding overfishing is the only goal, then the control rule of zero catch will always be preferred. It is the trade-offs between catch and overfishing that the authors describe earlier in the paper that are being ignored here.

So while the modeling approach is sound and the statistical treatment of results is generally appropriate, for those results which were considered, the lack of evaluation of trade-offs prevents me from concluding that this work can be used for management purposes. I think it is a good starting point for discussions with the Council and SSC regarding how to evaluate trade-offs, and which trade-offs are of most interest. However, until the trade-offs are fully integrated into the analysis, it is insufficient for management purposes.

Review of Wilberg, Miller, and Wiedenmann: Evaluation of ABC Control Rules for Mid-Atlantic Stocks

Dr. James Berkson
NOAA Fisheries SEFSC
Gainesville, FL

Overall I find this study to be an interesting pilot study investigating the potential effects of alternative control rules on simulated stocks. It is far too incomplete, in its methods, results presentation, results interpretation, and discussion to be used for management purposes in its present state. The work would need to be greatly expanded and improved to move beyond being a pilot study to being useful for management.

The first part of the study involved summer flounder and had three components: (1) a study investigating how uncertainty estimates produced in the summer flounder assessment are influenced by changing model assumptions, (2) investigating the application of control rules to the stock, and (3) quantifying the variability in the ABC and its effect on the stock. The purpose of this part of the study was not well developed, and it didn't seem to fit in with the second part of the study. This likely should have appeared in a separate report.

The second part of the study involved simulation studies to explore the impacts of various ABC control rules. The authors divided their study into those used in data-poor situations and those used in data-rich situations. I personally don't know that I agree with such a sharp line of distinction between the two groups. In some cases a stock that is supposedly data-rich may have uninformative data and may require a data-poor control rule. DB-SRA methods may appear data-rich for many stocks in the Southeastern U.S.

The ORCS approach is not fully described or applied, and because of this, it's utility is limited. The ORCS approach is meant to be highly flexible, allowing it to be modified as needed by individual SSCs. It uses an evidence-based scoring system (not discussed whatsoever in this report) to evaluate whether the stock should be classified as lightly, moderately, or heavily exploited, where the attributes, criteria, and scores can be modified as needed. It was created in such a way in an attempt to minimize the chance of misclassifying a stock as lightly or moderately exploited, if that wasn't the case. Wilberg et al. refer to a table from the ORCS Technical Memo that shows examples of how to assign a scalar based on risk for multiplication to the OFL to get an ABC, using two columns for their simulations. In doing so, they fail to recognize or even comment that these are example values and that this is another aspect of the flexibility of this method, which can be made more liberal or more conservative. The report is so abbreviated that key points such as this are not included. In leaving these points out, the methods appear more formulaic than they are, and the results of the simulations appear set in stone.

Much of the methods section is not clear. For example, it is unclear to me how the recent stable catches were calculated in the Restrepo method, despite the brief explanation. For DCAC, if I understand it correctly, the method was used when M was greater than 0.2, even though it is not supposed to be. In those cases the method was constrained, using incorrect values for M. This sounds like forcing a round peg in a square hole to get the simulation to work, which wouldn't be done in practice. In doing so, the simulation results were impacted. It's not explained why the 40^{th} percentile of the OFL distribution was used as the ABC when using DB-SRA. A different value would have given different results.

I don't understand why the simulated population had different initial trajectories in the data-poor and data-rich scenarios. In the data-poor scenario, the F-values increased throughout the initialization. In the data-rich scenario, the F-values increased then decreased. If there was a reason for this difference, it wasn't explained. A similar initial trajectory would have allowed for some comparison of data-poor to data-rich control rules, which was definitely lacking from this study.

It also was not clear to me why SPR-based reference points were used throughout the simulations. Given that the population was simulated, actual values of Fmsy and Bmsy could be calculated. They could also be calculated in the assessment for the data-rich scenarios.

In the biased scenarios, assumed stock status and productivity were overestimated every four years for 20 years. This doesn't seem realistic, as ABCs would be overestimated, the population would likely respond and decrease and eventually, you would think, someone would recognize the problem. Instead the problem continues to get worse over time. This makes for a very exaggerated scenario. It's no wonder that virtually none of the methods work well in this case. Perhaps it would have been more useful to look at the impacts of overestimating once or twice over twenty years or the impacts over a shorter time frame. This seems like something one would try in a pilot study, then refine to be something more realistic given additional time. Conclusions should not be made about which control rule to use based on this as the only "biased" scenario modeled.

In the real world, if an ABC exceeded the exploitable biomass in a given year, or not necessarily even that large, it is highly unlikely that 95% of the exploitable biomass for most stocks would be caught. More importantly, this would likely not continue for multiple years. As set up for the data-poor scenarios, this could continue for four years in a row before there would be a new re-estimation of the ABC. It would be far more likely that a smaller catch would occur and that the ABC would be re-estimated sooner. Model runs where the ABC exceeded the exploitable biomass should have either modified their control rules more severely to reflect something closer to reality or should have ended at that time step, keeping track of the number of model runs in each scenario when this happened. This also seems like a quick and easy way of handling this problem in a pilot study (the coding is far simpler) but requiring refining to actually be useful.

How were BRPs calculated in the data-rich scenarios beyond the first year? Why was time-varying M the only mis-specification modeled? Given more time, surely more mis-specificiations and sensitivities should have been run. These are needed to better understand the system before the model is applied for management.

In the data-rich scenarios, if the stock becomes overfished, the same control rule is applied. This again is a shortcut for the sake of the programming, but completely unrealistic in the real world. This should be handled differently, either by changing the control rule or by discontinuing that specific model run.

The data poor results seem to indicated that the more conservative the rule, the fewer populations experience overfishing. Results of the ORCS methods were an artifact of the way in which it was used in the simulations. As the method is highly flexible, alternative decisions regarding its implementation would result in different results. Once again, this what not discussed in the document.

There was insufficient discussion comparing the methods throughout the document. For instance, many of the results would make more sense if put in proper perspective given the similarities/differences of the methods.

In the results section, the term "P*" seems to be used as a summary statistic, as in the probability of overfishing in the simulation results. P* is traditionally used as a parameter input to the management process. Given a P* selected by management reflecting risk and a pdf of OFL, an ABC can be calculated. I don't think it's an appropriate use of the term to use it as an output statistic.

In conclusion, due to time constraints, I have pointed out many of the problems I've noted with this work. I do see much of value. It appears that the work could be developed more fully, incorporating the comments made by myself and the other reviewers. This would make the simulations much more effective, realistic, and defensible. This is not a trivial amount of work. I view this project about one-third of the way done, before it could really be useful for management.