Scientific and Statistical Committee Sub-Panel Peer Review of:

Kerr et al. Evaluating the Impact of Inaccurate Catch Information on New England Groundfish Management

Review prepared by

John Wiedenmann
Assistant Professor
Department of Ecology, Evolution and Natural Resources
Rutgers University

Christopher M. Legault
Supervisory Research Fishery Biologist
Population Dynamics Branch
Northeast Fisheries Science Center

Michael J. Wilberg
Professor
Chesapeake Biological Laboratory
University of Maryland Center for Environmental Science

September 2nd, 2020

1. **Introduction**

 The review panel met virtually on August 21st, 2020. The panel was composed of three scientists: John Wiedenmann (chair of the committee and current member of the New England Fisheries Management Council Scientific and Statistical Committee), Chris Legault (former New England SSC member), and Mike Wilberg (current member and vice chair of the Mid-Atlantic SSC). A couple of weeks prior to the meeting, the report (Kerr et al.) was made available to the panel. The virtual review opened on the morning of Friday, August 21st, with welcoming remarks and comments on the agenda by panel chair John Wiedenmann, followed by introductions of the panel, and a listing of audience members. Dr. Lisa Kerr then gave a presentation summarizing the work, followed by a series of questions from the panel, as well as from the audience. The remainder of the day was devoted to going through Terms of Reference (ToR) 1-6:

 1. Are the methods adequately described and based on sound analytic techniques and statistical principles?
 2. Are important uncertainties identified, and are the impacts of these uncertainties on the analyses adequately described?
 3. What are the strengths and weaknesses of the approach?
 4. Are the conclusions supported by the results?
 5. Are there recommendations for improvement?
 6. Do the conclusions provide information that is relevant for the Council to consider?

 The review panel agreed that the report was well-written, and the analyses conducted by Kerr et al. were of high quality, and that conclusions were generally supported by the results. The review panel has a number of comments in response to each ToR below.

2. Review of the Work for Terms of Reference

Terms of Reference (ToR)

1. Are the methods adequately described and based on sound analytic techniques and statistical principles?

This ToR was successfully completed. The peer review panel agrees that the methods are based on sound analytic techniques and statistical principles. The panel recommends that methods description (A) needs some additions/clarifications and (B) some explanation as to why certain results were selected to be presented as the base scenario in the report.

 A. The peer review panel identified several aspects of the study that would benefit from additional description. In particular, the details of how management was implemented (the calculation of the annual catch limit (ACL) through its

implementation in a realized fishing mortality rate). The peer review panel recommends providing this information in equation form (perhaps starting with the equations that were provided in the peer review meeting), and to also include time subscripts so the lags are clearly described. The calculation of the SSB reference point using the truncated recruitment time series is not consistent with current practice for this stock. Lastly, how the results are summarized could use some additional explanation, in particular how the assessment bias was an average over time rather than the bias in just the terminal year of the assessment. The latter quantity is likely more relevant for management than the average.

B.   The way stock size and fishing mortality information feeds into management was modeled differently than is currently done for New England stocks. These differences include a lack of projections in the simulation model, no retrospective pattern adjustments, and harvest control rules that differ from the previously realized management performance. While the justification provided for the simplifications in the simulation study are valid, the consequences of these differences for the results and conclusions should be described.

There were also several minor issues with the equations that should be checked and updated:
- The Ms in the equations should have t subscripts
- There are two different symbols for fishery selectivity (eqs. 1 and 4)
- Table 4 eq. 5 has an error (the phi should be an N)
- Two variables are used for proportion mature at age (P and theta)
- Omega appears to be defined differently in the equations and the text
- Table 5 eq. 9 has the numerator and denominator reversed

2. Are important uncertainties identified, and are the impacts of these uncertainties on the analyses adequately described?

This ToR was successfully completed. The review panel considered that overall the report does a good job of identifying important uncertainties and describing the impacts of these uncertainties on the analyses. This is mostly accomplished by clearly defining what the model does and does not do.

The review panel identified some uncertainties that could be further described in the report.

A.   While a wide range of catch misreporting was examined in the report and the derivation of the values used was reported, a stronger link between the values used and what could be actually accomplished through Amendment 23 would be useful. This would entail not only defining the range of misreporting examined, but also consider an improvement in reporting sometime in the future. The plausibility of the different catch misreporting scenarios examined could also be addressed in the report. Currently, all the values explored are treated as equally likely. Some basis for why one value might be preferred over another could be

attempted or else a stronger justification for why the extreme values are plausible (as opposed to needing much wider or narrower range) would help the reader understand the consequence of this uncertainty better.

B. The current formulation has catch misreporting beginning at the same time as the assessments begin. This means that the first few assessments will not exhibit retrospective patterns or bias in results. Because the Gulf of Maine cod assessment has exhibited a retrospective pattern for many years, allowing the impact of misreporting catch to begin before the assessment period begins would make the model more consistent with the actual assessment and remove some of the odd results of no bias in early assessments.

C. Related to point B, the panel notes that there is a disconnect between using the most recent assessment estimates as the "true" historical stock dynamics, but also allowing catch to be misreported in the past. This issue was discussed during the review, but should also be made explicit in the document.

D. The model assumed a hockey-stick stock recruitment relationship, whereby expected recruitment did not increase above a given stock size. This assumption has the effect of making the expected catches similar for a wide range of exploitation rates, meaning the same catch is generated by fishing hard on a small population or fishing more lightly on a larger population. This assumption also makes the calculation of maximum sustainable yield reference points nonsensical because the Fmsy value will either be Fmax or Fcrash depending on where along the hockey-stick stock recruitment relationship the replacement line intersects. Since Fmax is well known to be a poor estimator for Fmsy, and clearly Fcrash would be a bad choice for Fmsy, the Fmsy values reported should not be considered reliable. Assuming a Beverton-Holt stock recruitment relationship with steepness less than one would allow a more realistic examination of the consequences of fishing at different intensities and allow calculation of a reasonable Fmsy value (although the latter would depend heavily on the value assumed for steepness).

E. The use of 100 realization in the simulations is fine for defining the mean or median outcomes, but is not sufficient for defining the tails of the distributions. Either the number of realizations could be increased, although this would take some time, or it could be noted in the discussion that the uncertainty ranges are not well defined due to the limited number of realizations conducted. The relatively low number of realizations may also have led to some of the non-intuitive results associated with the inconsistent separation of the results for the different catch misreporting rates.

F. There are some inconsistencies during the pre-management period that arise due to the manner in which catch misreporting was modeled. For example, the very high F in the last year before assessments begin has to be ascribed to implementation error even when there is no catch misreporting occurring yet. This would be more of a problem if other management measures were considered during the feedback period, but still could be more clearly addressed in the discussion.

G. A common uncertainty in providing catch advice from actual assessments arises due to the need to project the population into the future. This aspect was not

modeled in the simulations. The authors noted that this aspect was under development for future work. The review panel encourages this development. In the meantime, the report could be more clear about how catch advice is generated in these simulations.

3. What are the strengths and weaknesses of the approach?

This ToR was successfully completed. The review panel considered the simulation modeling approach used as one of its major strengths. Using the actual ASAP stock assessment model for Gulf of Maine cod within the simulation, as opposed to just assuming some sort of biased response in the simulations, made for much stronger conclusions about how misreporting impacts both the assessment and catch advice. The systematic approach used of addressing a number of factors allowed easy comparisons to be made to examine the impact of a single variable on the results. Another notable strength of this work was the close working relationship between the authors and the Groundfish Plan Development Team. The feedback provided by the PDT helped improve the scenarios examined and made them more relevant to management.

The major weakness identified by the review panel was that this work did not directly address the actions proposed as part of Amendment 23, specifically the improvement of catch reporting after a period of catch misreporting. It was explained to the review panel at the end of the review meeting that this topic was not the focus of the requested work. The review panel feels this was a missed opportunity because the simulation framework developed is well-suited to address the issue of Amendment 23 directly. A less important weakness was the emphasis on the sliding harvest control rule, which the review panel did not agree was more similar to the actual harvest control rule for Gulf of Maine cod than the constant one. Similarly, the Mramp model results were distracting in the report because the report emphasized the challenges associated with reference points for this model, which are independent of catch misreporting. The authors could assign the Mramp results to an appendix for completeness and simply note that the results related to catch misreporting were consistent between the Mramp and M=0.2 models. Removal of the reference point distraction associated with the Mramp model would allow more focus on the catch misreporting issue with the M=0.2 model.

4. Are the conclusions supported by the results?

This ToR was successfully completed. The peer review panel largely agrees with the conclusions in the report but recommends additional caveats for several conclusions. The primary conclusion of the study is that negatively biased catch data (caused by unreported discards or other sources) produces a negative bias in estimated biomass, but less of a bias in the estimated fishing mortality rates. The pattern of bias becomes more complicated when there is a change in bias of the catch data during the time series, but the ultimate effect is similar. The bias in estimated biomass is directly related to the bias in catch.

The peer review panel found that several of the conclusions require additional caveats.

A. The study concludes that correcting for the retrospective pattern in their study would correct in the wrong direction. While this is true specifically for SSB and F, it is not necessarily true for the ABC or for stock status. Similarly, the performance of adjusting estimates for retrospective pattern prior to calculating the ABC was not tested in this study. Lastly, this conclusion is also affected by how bias in SSB was summarized - the results may be different if only the performance metric was SSB in just the terminal year of the assessment.

B. The study implies that having better catch data will improve management. However, a return to unbiased (or less biased) catch data after a period of substantial bias was not simulated. From this study it is unclear how long it will take for assessment accuracy to improve.

C. The peer review panel was concerned with the conclusion that <50% underreporting had a small effect on management performance. While it is true that this level of underreporting resulted in better outcomes than higher levels of underreporting, it still resulted in unintended overfishing and relatively substantial discarding.

D. The results are sensitive to using a harvest control rule that is very conservative relative to the history of management. Because of the low fishing mortality rates used in the control rule, the 50% bias scenario seems to have a relatively minor effect on management performance. However, a control rule that achieved fishing mortality rates similar to recent ones would likely see much worse management performance (in terms of stock size and overfishing).

5. Are there recommendations for improvement?

The panel had a number of recommendations, both in the short- and long-term. Short-term recommendations could be addressed without additional reruns or modifications to the model using existing model output . Long-term recommendations would require modifications to the existing model structure and reruns of the model. The review panel acknowledges that the project period is at or near its end, and long-term recommendations are provided as things that could be explored in the future. Some of the recommendations listed below have already been discussed in the panel's responses to ToRs 1-4, and are included here for completeness.

Short-term recommendations

- Consider providing some additional clarification about the motivation for this work, being more explicit about what this work was intended to inform, but also about what it is not exploring (i.e., improvements in catch bias; see comment in ToR 6 for more).
- In the Introduction or in the Discussion, expand on the current body of literature that has explored misreporting of catch and how to deal with it. Currently the only study cited is Rudd and Branch (2016), which used a production model. There are a number of recent studies that have used simulations to explore the bias due to misreporting in age-based assessments, and some that have incorporated the bias in

a management feedback loop as was done here (see below for a list of potentially relevant papers).

- There was some issue amongst the public about the magnitude of catch bias explored in the analysis. It might be helpful to provide some additional detail for the levels chosen, and also to note that the levels you explored do not imply that that is the level of underreporting that is occurring. Alternatively, the portion of the PDT documents provided to the panel could be included as an appendix to the report.
- Consider moving discussion of the M-ramp model to an additional Appendix. While relevant to GOM cod, the M-ramp discussion in the current document is limited, and distracts from the overall message of the work.
- Calculate the relative error in assessment estimates (REE) of bias using the terminal year only. Currently REE is estimated using all years of assessment estimates, but the panel noted that the terminal estimates are the basis for management advice, and are subject to retrospective patterns, particularly in the change point bias scenario (see Figure 1 here).
- Include a Figure that illustrates how the magnitude of bias relates to the magnitude of REE. This could be done having catch bias on the x-axis and median REE (calculated for all years and terminal years only) on the y-axis, and would allow the reader to extrapolate the impact of levels of catch bias not explored in the analysis.
- Consider calculating additional performance measures, such as the probability of biomass dropping below some threshold, probability of overfishing, proportion of years when the total catch exceeds the exploitable biomass, median rebuilding time, etc. The panel understands that such metrics would not change the current conclusions about the impact of catch bias on assessment estimates, but could illustrate how such effects ultimately impact the population.
- Check in text citations and references for accuracies. Some misspellings were noted and some references were missing from the list.

Long-term recommendations
- Increase the number of model iterations to better characterize the uncertainty in performance measures.
- Include projections and multiyear intervals between stock assessments.
- Include retrospective adjustments when setting catch advice.
- Explore additional change points in catch bias where the bias is reduced in the future by different magnitudes, mimicking the impact that additional observer coverage would have on assessment and management performance. Doing so would provide insight into how long it takes for assessment estimates to become less biased. Such an analysis would likely require exploring different length time periods of biased catch data (e.g., 5, 10, 15 years) as well as different magnitudes of improvement in the catch reporting. The panel agrees that this would be a very informative extension, but also a very big undertaking.
- Explore an earlier start date for the initial change point (currently 2015) in catch data.

- Explore cases of model misspecification. For example, how does catch bias impact assessment estimates when M is assumed to be 0.2 but it is really 0.4, and vice-versa. There may be some non-intuitive interactions among misspecifications that could cause bias to be worse or better than presented in the current study.

6. Do the conclusions provide information that is relevant for the Council consider?

This ToR was successfully completed. There was some confusion amongst the review panel about the scope of work and how the analysis was intended to inform management. Because different levels of increased observer coverage are being considered for Amendment 23, the review panel initially thought that to be useful to management, the analysis should have included different levels of improvement in the biased catch data. However, Executive Director Tom Nies clarified that this was not the intent of the work, and the focus was on a basic understanding of how biased catch data impacts assessment and management performance. The review panel agrees that the work done is relevant and useful for management advice in this context. However, the panel feels that in the future the scope of work should be provided by the Council to the panel as part of the review materials (perhaps by including the RFP as a document) to avoid confusion.
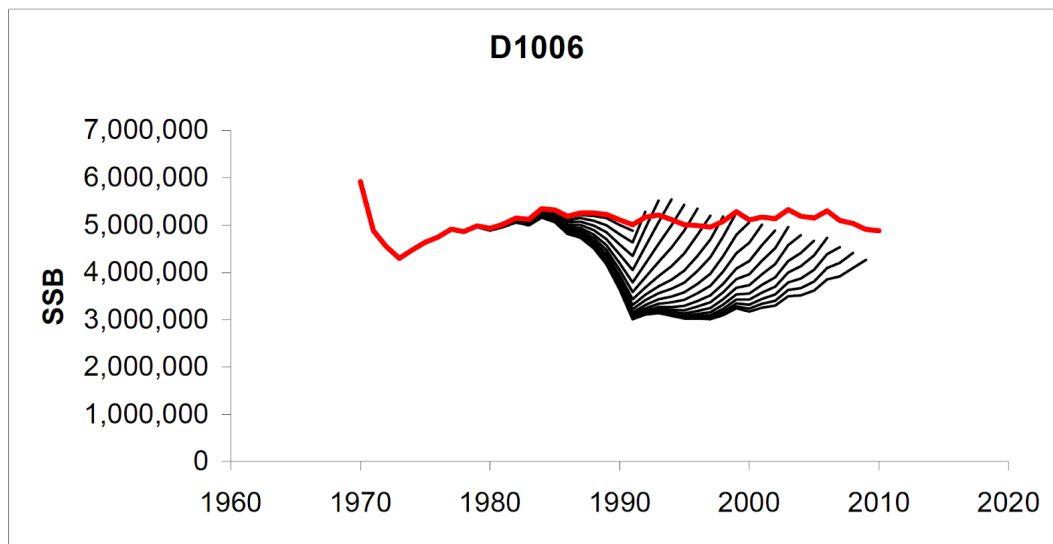

Figure 1. True (red) and estimated (black) SSB from the sequential assessment model fits from the retrospective analysis with underreported catch of Legault (2009; this panel is part of Figure 7 in the report).

Relevant Papers

Cadigan, N. G., 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences, 73: 296-308.

Hordyk, A. R., Huynh, Q. C., and Carruthers, T. R. 2019. Misspecification in stock assessments: Common uncertainties and asymmetric risks. Fish and Fisheries, 20: 888-902.

ICES. 2020. Workshop on Catch Forecast from Biased Assessments (WKFORBIAS; outputs from 2019 meeting). ICES Scientific Reports. 2:28. 38 pp. http://doi.org/10.17895/ices.pub.5997

Legault CM, Chair. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. Northeast Fisheries Science Center Reference Document. 2009; 09-01. 30 p. Available: http://www.nefsc.noaa.gov/nefsc/publications/s-a/

Perretti, C.T., Deroba, J. J., and Legault, C. M. 2020. Simulation testing methods for estimating misreported catch in a state-space stock assessment model. ICES Journal of Marine Science 77: 911-920.

Van Beveren, E., Duplisea, D. E., Castonguay, M., Doniol-Valcroze, T., Plourde, S., and Cadigan, N.. 2017. How catch underreporting can bias stock assessment of and advice for northwest Atlantic mackerel and a possible resolution using censored catch. Fisheries Research, 194: 146-154

Van Beveren, E., Duplisea, D. E., Smith, A., and Castonguay, M. 2020. An example of how catch uncertainty hinders effective stock management and rebuilding. Fisheries Research, 224: 105473.

AGENDA

Scientific and Statistical Committee Sub-Panel Peer Review of

Evaluating the Impact of Inaccurate Catch Information on New England Groundfish Management

Friday, August 21, 2020

Via webinar

9:00 am   Opening remarks and agenda review................................ (*Wiedenmann*)

9:10 am   Presentation – Overview of report, *Evaluating the Impact of Inaccurate Catch Information on New England Groundfish Management* (*Kerr)*

9:50 am   Questions on presentation..

10:50 am   Public comment...

11:00 am   Review panel discussion... (*Review panel)*

12:30 pm   *Lunch break*

1:30 pm   Review panel discussion continued.................................... (*Review panel)*

2:45 pm   Wrap up/plan for sub-panel report writing............................ (*Review panel chair*)

3:30 pm   *Adjourn*