REPORT ON THE GROUNDFISH SCIENCE PEER REVIEW MEETING, FEBRUARY 3 – 8, 2003

New England Center
University of New Hampshire
Durham, NH

summarized
by

Andrew I. L. Payne
Chair (Appointed by CIE)
CEFAS
Pakefield Road
Lowestoft
Suffolk NR33 0HT
United Kingdom

GROUNDFISH SCIENCE PEER REVIEW MEETING

Summary Report Contents

Executive Summary and Recommendations

The Groundfish Science Peer Review meeting was held in Durham, New Hampshire, USA, from 3 to 8 February 2003, the first three days in public meeting and the final three days in executive session. The independent reviewers were Dr.s Bell and Payne (CEFAS, Lowestoft, England), Cook (Fisheries Research Services Marine Laboratory, Aberdeen, Scotland), McAllister (Imperial College, London, England) and Mohn (DFO, Halifax, NS, Canada).

The public meeting, led by a moderator, consisted of solicited and offered presentations, mainly by scientists and technical advisers, followed by in-depth debate on the research survey warp offset and other fishing problems, biological reference points, and stock projections, with one day given to each topic. The executive session of the panel was held privately with only the independent reviewers present; no consensus or unanimity of advice was sought. Arrangements for both meetings were excellent.

Documentation was comprehensive and fully relevant, but was voluminous and a little late in coming to the reviewers. The Terms of Reference were so comprehensive that they were very difficult to understand; there were also a mix of scientific and policy issues. Answering policy issues was deemed to be beyond the mandate of the reviewers.

The recommendations listed below (some technical, some generic) are drawn directly from the reports of four independent panellists, rationalizing where possible (to avoid duplication), but allowing for differences between opinions. No priority of urgency is intended by the numbering, but an attempt has been made to list the recommendations under each of the main terms of reference. Each recommendation is a concise summary of more extensive discussion in the four individual reports (to which the reader is directed if more detail is required), in the body of this summary report, or in the chair's own notes from the meetings. Other valid suggestions, as opposed to recommendations *per se*, are contained in the body of the report. The reader is urged to consider those suggestions as well as the recommendations listed below.

Warp issues

1. The NEFSC should continue to use the trawl survey data for New England groundfish stock assessment unadjusted, because such use is scientifically justifiable. The sensitivity tests carried out to evaluate the implications of the trawl warp offsets for the evaluation of stock status and rebuilding plans adequately bounded the range of potential introduced biases.
2. If the various survey gear protocol issues remain controversial, then an experiment specifically designed to detect these effects should be undertaken.
3. Survey gear should perform as consistently as possible. Further, the commercial gear in the experiment yielded less sampling variability, and the reasons for this should be investigated.

4. An evaluation of the ability of the survey to detect population signal above the inherent noise should be conducted for those assessments most dependent on survey indices of abundance.
5. Conversion coefficients, for use in assessments, should be determined for stocks for which the catch results from optimal and suboptimal gear settings differed most.
6. If the NEFSC surveys are subjected to redesign with the involvement of stakeholders (all reviewers consider this measure to be unnecessary), independent scientists with a knowledge of survey design must be included to ensure that scientific standards and data-series continuity are not compromised.
7. Consideration should be given to using estimates of survey CV to supplement current attempts to minimize the influence of survey variability, in survey-data-only stock assessments, through the use of running averages, thus producing a more risk-averse strategy.
8. Because estimation error can trigger a stock rebuilding response even when there is no need to do so, consideration should be given as to how assessment error should be handled within the management process.

Biological reference points

9. If the current B_{MSY} reference points are adopted, the potential costs of adopting an erroneous value need to be evaluated.
10. Alternative means of modelling groundfish stock dynamics should be evaluated, and their results compared with present procedures, although the currently used ADAPT model is deemed by the reviewers to be scientifically sound. Reasons for differences in the outputs of the different models need to be sought.
11. Most methodologies used by the NEFSC to compute F_{MSY} and B_{MSY} are adequate, but the protocol used to evaluate the goodness of fit of alternative stock/recruit functions to the data, and to select alternative models to determine F_{MSY} and B_{MSY} for the purposes of fisheries management, needs to be revised.
12. A simulation study, using age-structured operating models, should be undertaken to validate the survey index method in a management context, to investigate its sensitivity to *ad hoc* assumptions and to evaluate the potential biases and imprecision in the results.
13. The methods used to derive non-parametric stock/recruit functions to approximate B_{MSY} should be simulation tested with a variety of underlying operating models for stock/recruit processes, to test the robustness and accuracy of the methodology.
14. If the issue of model selection (Beverton-holt vs. Ricker) remains on the table, then model validation needs to be addressed more fully, and more divergent models ought to be tested (e.g. non-parametric deterministic and probabilistic).
15. If NEFSC adopts a Bayesian statistical approach to select a stock-recruit model for reference point determination, the center should first decide (Beverton-Holt or Ricker) the baseline set of priors deemed the most appropriate reflection of existing knowledge of model parameters for each stock-recruit model form, and then evaluate the alternative functional forms using Bayes' factor.

16. More than one diagnostic tool should be applied to evaluate convergence on posterior distributions, rather than simply relying on the MCMC software for statistical estimation.
17. There may be other objective criteria that are more appropriate for deciding when to include autocorrelation in models to determine F_{MSY} and B_{MSY}, and in stock projection models. Such criteria should be investigated
18. Use of a probabilistic methodology for estimating the intrinsic rate of increase of r, using demographic data or meta-analysis and a projection methodology that probabilistically accounts for uncertainty in r, should be considered since the ASPIC surplus production estimation and projection methodology has serious methodological limitations.
19. An adaptive approach to using harvest management to control biomass is eminently sensible in the light of uncertainty, and should be pursued.
20. Can all 19 stocks be moved simultaneously towards B_{MSY} through single-species management? Given current knowledge of the complexity of biological interactions and the ecosystem in which the stocks exist, the reviewers doubted whether the question could be answered satisfactorily. Single-species MSYs are not good indicators of multispecies MSY, so caution will be needed in the choice of B_{MSY} target. There could well be value in assessing interactions in multispecies fisheries, using spatially heterogenous models such as Iceland's BORMICON.

Stock projections

21. The stock/recruit models used in projections should be validated against historical observations of stock dynamics.
22. Consideration should be given to implementing a phased reduction of fishing mortality, with consideration given to the possibly greater risk to the stock if this reduction is implemented.
23. From a purely scientific perspective, the setting of an intermediate biomass target could be defended. In such a case, once a rebuilding F and trajectories were in place, this intermediate target could be used as a signpost to see if stock rebuilding was on track. If the rebuilding target were not on track, an assessment would be needed to determine what the causes are (e.g., were the original projections in error, was recruitment atypical, or was recruitment as expected but F not achieved).
24. An evaluation of technical interactions in the mixed fishery should be undertaken to investigate the consistency of multiple MSY targets.

Background, preliminaries and documentation

The CIE issued the call for the names of independent experts to be considered as participants in the Groundfish Science Peer Review meeting in early January 2003. Given the large volume of documentation subsequently received, this call was very late, though perfectly understandable given the background information we received. All reviewers stressed the lack of time allocated to read the documentation. The initial call was for three independent experts to be contracted by the CIE, and two others to be nominated by those involved in the review, but all five were ultimately contracted by the CIE. By mid January 2003, the selection of panellists had been. The panellists themselves did not know the names of the other participants until just prior to the meetings. Simultaneously to the selection of panellists, the first draft of the Terms of Reference (in the form of Statements of Work) was sent to the five panellists, and travel and other arrangements were initiated. Some 10 days before the meeting, a revised (much expanded) Statement of Work (Appendix 1) was submitted to the five panellists, and the NEFSC contact (Dr. Murawski) arranged website access to the voluminous documentation that formed the backdrop to the meeting.

The final Terms of Reference are contained in Appendix 2 and the documentation, along with all material subsequently received and/or consulted by reviewers, in Appendix 4. Documentation continued to be added to the website right up until the day of departure of the panellists to the USA, and Dr. Murawski and his NEFSC colleagues, as well as the various other stakeholders who also contributed to the wealth of information listed on the website (subsequently provided in both hard copy and on CD), deserve the gratitude of the panellists for their thoroughness. The NEFSC also went out of their way to ensure that relevant scientific literature requested by the panellists during their executive session was made immediately available. The list of reviewers and the agenda of the meeting (the latter only "finalized" the day before it started) are contained in Appendix 3. Appendix 4 lists not only the formal and informal material provided before and during the meeting, but also details some of the relevant literature consulted by individual panel members subsequent to the meeting, as well as the handout and summary presentations of those variously mandated to address the issues on the table. The latter material was very useful to the review panel, as were the formal and informal discussions the panel had with those participants who expressed their willingness to do so (virtually everyone present).

Prior to and during the meeting, in discussions around the fringes and in formal debate, the final terms of reference were subject to some debate. The problem revolved not around the detail provided in the TORs, which was particularly useful, but around the mix of generalities and specifics that they contained, which allowed for different interpretation. There was also a mix of directives for advice and comment on the science of the assessments, management targets and survey techniques, for which all panellists felt competent and to which the TOR introduction clearly referred, and advice on policy process, for which panellists felt they were not qualified. Ultimately, the report that here summarizes the individual evaluations of the five panellists, skirts those TORs (and specific questions subsumed in them) for which the panel felt unqualified to address in other than general scientific terms. A further concern to panellists was that, when clarity

was asked of the various participants about what specific TORs really meant, there were often several views from the floor, and the panel was left having to make its own (often individual and again varying) interpretation. This situation may well have arisen from two confounding factors: the short notice with which those drafting the TORs had to deal (fully understood to be out of their hands), and the fact that the drafters represented different constituencies with differing requirements of this independent review. Even at the end of their work, more than one reviewer (panellist) was left questioning who the customer for this review was.

A further concern of panellists was the process through which this summary report would have to pass to meet the deadline of February 21, 2003. The only means of meeting the deadline (debated and agreed by all five panellists) entailed them having lengthy discussions during the three-day executive session following the public workshop, commencing the drafting of their reports in each others' company, while discussing issues that required clarity, and then relying on the summarizer to interpret the main points of their evaluations in a summary report without the luxury of time to iterate the report adequately between them. Indeed, some parts of all the sections of this report (including this one) are subject to the same limitation, so it is hoped that those receiving the report will make allowances for this fact. In essence, time constraints made it virtually impossible for the reviewers to consult as much as they would have liked on the final content of the summary report.

Overall, the content of material for review and the requirements of the panel for the six days of meeting were clearly understood well before the meeting. However, although there was some indication one week before the meeting that a moderator was being sought for the public meeting, the subject was not formally raised with the panel chair until two days before his departure. This comment is not meant to be a criticism of the fact that a moderator was present. Indeed, as shown in the next section, the moderator's presence and enthusiasm contributed to the success of the public meeting. The moderator’s knowledge of the fishing community, and associated contentious issues, was extremely valuable in ensuring continuity of debate when there was a risk of clashes between various constituencies. The subject is only raised here to highlight the fact that time was so short in planning the meeting that some independent reviewers, specifically the chair, were rather unsure of the role that they would be playing. For instance, the chair’s Statement of Work chair made no mention of the scientific review report expected of the other four panellists. Ultimately, having a moderator present assisted the chair in making copious notes that allowed him to formulate his own views about the scientific merits of the groundfish management science being practiced in the north-eastern USA. Some of these views are encapsulated in the main body of text of this report along with those of the other four panellists.

Towards the end of the public meeting, discussion between the moderator and the chair generated an idea to allow floor participants a final chance of ensuring that their voices and views were heard and noted by the panel. During the last day of the public meeting the moderator called for participants to raise pertinent questions, or to make brief comment, if they felt that their viewpoints may not have been adequately entertained or

understood. The idea was that the review panel would then, where appropriate, ensure that an answer or comment covering that issue was contained in the body of their report. These floor-generated questions and comments are listed in Appendix 5. The questions are not specifically answered in the main report, but answers to most may be found in both the summary report and the individual reports of the four other reviewers.

Conduct of the meeting and production of the Report

The public meeting was held from February 3 to 5 in the Berkshire Room of the New England Conference Center. Immediately prior to the meeting, a briefing session was held between the review panel and the moderator. At that meeting, the subject of a record of the public meeting was raised; the panel felt that a meeting without a record would not be remembered as a meeting at all. The moderator therefore undertook to arrange that such an *aide memoire* be compiled. For the first day, he took notes himself; for the second and third days of the public meeting, Ms. Togue Brawn handled the responsibility on his behalf. This valuable record was made available to the panel, which considered it in their individual scientific evaluations. The record is not appended to this report however, because it was never meant to be a formal part of the proceedings; participants interested in reading it are encouraged to request a copy from the moderator. The briefing session also ironed out any misconceptions either moderator or panellists may have had about the manner in which the public meeting was to be held.

The structure of the public meeting was sound. Adequate presentation time was allocated to each speaker, and wherever possible about 30 minutes were allowed immediately after the presentation for discussion and questions, initially by the panel and then by floor participants. Predictably, the 30-minute planned discussion time was sometimes inadequate. To cover this eventuality, the moderator allocated time at the end of morning and afternoon sessions for a catchall debate, during which pressing issues could be aired more fully. Where this additional time also proved inadequate, discussion rolled over into the corridors of the meeting and finally in the session where floor participants could make their point for a last time. In general, the spirit of debate was positive and constructive throughout, although all panellists were aware that there were major differences of opinion between representatives from different constituencies on some of the subject material.

The purpose of the public meeting was obviously to showcase as many views as possible relating to the survey warp issue and subsequent experiment, the biological reference point targets and evaluation, and the stock rebuilding and projection methods. The moderator handled the meeting professionally, allowing all parties to make their point, but cracking down firmly on repetitive or long-winded comment. Panel members were grateful for the opportunity afforded them to clarify issues relating to presentations early, and not to get sidelined by lengthy discussion that may have been of limited relevance. The review panel also thanks all participants for sharing their views honestly and willingly whenever called upon to do so (particularly Mr. Eric Smith of the New England Fisheries Management Council and Dr. Steven Murawski of the Northeast Fisheries

Science Center), and for the positive and sympathetic manner in which they stimulated interaction and received the manifold requests for help in understanding some of the voluminous documentation. The two gentlemen mentioned were also (presumably) responsible for making the arrangements for the meeting, and for ensuring that all possible courtesies were extended to the review panel (including during the executive session mentioned below) and other participants. Grateful thanks are therefore extended to them on behalf of us all.

The final three days of the meeting were held in executive session and involved the independent reviewers only, supported by Ms. Karina Jolles of New Hampshire Fish and Game. The meeting was held behind closed doors in the Narrangansett Room of the New England Center, which included a multi-user telephone system provided by the Center for communication with those few participants at the public meeting the panel felt they had to consult again for purposes of clarity. The executive session took the form of scientific debate about the issues raised, ironing out differences in interpretation of the science-based management issues between panellists, and in starting to draft the individual panel reports. In this respect, the chair's personal report was limited to main points and recommendations only, to allow incorporation in the summary report of the individual views of the other four panellists once their own reports had been finalized. In many cases, scientific opinion of the panel converged, but no attempt was made to influence individual panel members to reach consensus or unanimity, so the report that follows sometimes contains several (occasionally diverging) opinions. It also needs to be stressed that each panellist did not take it upon himself to address every aspect mentioned in the terms of reference provided, rather limiting himself to those aspects which he felt both most competent and most moved to address. Overall, though, virtually every issue was covered in the individual reports and hopefully also in this summary.

The Summary Report

For reasons of clarity and ease of understanding by the reader, this summary report is arranged in the same order as the major terms of reference. However, as the questions posed and comments made in the Terms of Reference are inextricably intertwined, no attempt has been made to respond in strict order of question listing. Nor is any priority alluded to by the order in which the comments are listed or numbered. The titles and objective statements listed at the head of each section are drawn from Appendix 2, and readers are referred to that Appendix where the report refers to a specific question. Additionally, intricate technical detail is avoided in this summary, given the breadth of stakeholder expertise in the review process; readers requiring such technical detail are referred to the individual (separate) reviewers' reports of Dr.s Bell, Cook, McAllister and Mohn.

TOR 1 – TRAWL SURVEY ISSUES AND INFLUENCE ON MANAGEMENT ADVICE

Considering the results of the Groundfish Assessment Review Meeting (GARM), subsequent results from experimental trawl comparisons, and other appropriate information, provide an evaluation of the significance of potential differences in trawl survey catchability resulting from recently discovered survey gear problems on management advice for groundfish stocks managed under the Northeast Multispecies Fishery Management Plan.

1. The trawl surveys conducted by the NEFSC are an important source of data used in the stock assessment of groundfish in New England, and are crucial to the management advice given by scientists. The data are used both as direct estimates of relative stock biomass and relative fishing mortality, and as relative abundance estimates to "tune" catch-at-age analyses from which total biomass and fishing mortality values are derived. From a scientific perspective, the relative power of a survey is not important provided the species being surveyed are sampled adequately, and more effort needs to be made to convey this fact to stakeholders. However, the power of a set of gear could be influential where species (or sizes of certain species) are caught rarely by the survey.
2. Eight surveys on board RV *Albatross* between 2000 and 2002 were made with incorrectly marked warps, possibly affecting the efficiency of the gear and contributing to both bias and noise in the survey estimates. Incorrectly marked warps were not the only issue concerning industry participants at the meeting; they were also dissatisfied with the gear itself (outdated!), had concerns about the rigging of the net, questioned the lack of quality control in operating the net in a standard fashion, and felt that the gear was being towed at suboptimal speed. If valid, these concerns could affect not only the cruises where the warp-offset problem was known but also call into question the relevance of the whole time-series of survey-based abundance indices as well as potentially undermine the basis of many of the calculations of biological reference points and the ability to judge the status of certain stocks. Industry concerns about the poor catchability of the survey gear underpinned much of their criticism, so it is clear that there has to be more interaction between scientists and fishers in designing and perhaps participating in future surveys (but see below).
3. The warp offset issue has been intensively analysed and studied, and the review panel overall supported the conclusion of the Groundfish Assessment Review Meeting (GARM) that there is no detectable systematic reduction in trawl survey catching efficiency attributable to trawl warp offsets. The GARM had also concluded that there were no size frequency effects. Analyses made since the GARM were concluded to be generally unable to detect or quantify any warp offset effect, although this does not mean that there was no effect, but rather that the natural variability of the data is such that no effect is detectable. The recent experiment conducted on board RV *Albatross* comparing an "optimum" net with a sub optimally rigged net was, in the opinion of the panel, inconclusive in that it was generally impossible to attribute any differences in catch rate between the optimally and sub optimally rigged nets to any particular fault or characteristic of the rigging, perhaps

because the individual faults could be compensating for (masking) each other. Fogarty (2003) made a similar statement in his report, stressing that the "design of the experiment does not permit separation of the effects of the trawl warp offset from other gear characteristics." Therefore, if the various survey gear protocol issues remain controversial, an experiment specifically designed to detect individual effects of suboptimal rigging will need to be undertaken. However, the fact that an effect was only detected for skates, herring, and scallops lends support to the general conclusion that the effect of the warp offset problem is not large.

4. The panel reviewed the design of the experiment comparing commercial and survey trawls in three areas, and generally found it to have been appropriate (though rather short, just three days), both in terms of design and protocols used. The statistical analyses applied also appear to have been reasonable for the purpose intended (specifically those of Fogarty and Brodziak). Not unexpectedly, however, there were differences between the catches of survey and commercial gear, in terms of size of catch, its species composition and the variance in the mean catch per tow. The differences were obvious despite the expected noise in the data anticipated from such a short experiment (realistically a gear trial), which made the drawing of any useful conclusions from the differences extremely difficult. Overall, the panel found the design of the experiment limited only in that the number of questions it was expected to answer in such a short time was too many. In terms of testing the control against the offset gear, the experiment performed well, but if the experiment was expected to ascertain causality among perturbations, the data were too few. Future experiments of the same nature, which would be valuable, need clear objectives and sufficient time to evaluate objectives individually, not to be undermined by more and more perturbations being added to the initial experimental objective (in this case the effect of warp offset).
5. Because commercial gear catches more of a certain species, it does not follow that an assessment based on such an index implies a larger stock. The NEFSC survey indices are used as a measure of population change only and not as an absolute measure of total abundance, so the efficiency of the research survey is not crucial. Further, a difference in species composition shows that the gears sample species with different efficiency, which is of consequence if comparisons are made between the relative abundance of different species sampled in a single year but not when tracking the relative abundance of individual species over time. The panel was concerned, however, that the separate analyses of Fogarty and Starr revealed lower sample variance for the commercial than for the survey gear. The two gears are so different that they seem to be sampling different populations, which means that the drawing of firm conclusions from the analyses is tenuous. However, a crucial requirement of a research survey is that it performs as consistently as possible. There is therefore value in investigating exactly how and why the commercial gear reduces sampling variability. Starr's analysis was categorized by the panel as crude (obviously owing to time constraints), but credible.
6. The panel carefully evaluated the findings in the GARM report. In their opinion, the sensitivity analyses presented adequately bounded the range of potential effects inferred from analyses of historical and comparative data. However, if the fishing power on yellowtail flounder really did increase by 50% with the offset warps, as the

analysis in the GARM report seemed to show, then further sensitivity evaluations of the implications for the VPA stock assessment for the species of such an increase in fishing power in the survey need to be carried out. One of the panelists, Dr McAllister, carefully evaluated evidence for any intervention in NEFSC trawl survey data associated with the use of offset warps. He concluded that the statistical tests applied were rigorous and fair, and that the only mild criticism he could level was that the analyses addressed just a few species, certainly not all the target ones. Details of these evaluations are contained in Dr McAllister's own report. Overall, the panel was happy that the GARM had adequately characterized uncertainties in estimated stock sizes and rebuilding mortality rates potentially arising from unequal warp offsets, except perhaps for yellowtail flounder, as mentioned above.

7. In summary, the panel believes that the warp offset did not undermine of the validity of the survey time-series of abundance indices for target species. However, industry concerns about the ability of the NEFSC survey to track with precision any real population changes for those species assessed on the basis of survey data only, deserve consideration. Where sampling efficiency of the survey gear for a species is low, or the population being sampled is small, there is danger that the precision of any abundance estimate is not sufficient to make meaningful judgements about stock trends, particularly over short time-scales. Therefore, it would be useful to analyse the ability of the survey to detect a population signal above the inherent noise in the survey for those assessments most dependent on survey indices of abundance. It is therefore recommended that detailed simulation modeling be undertaken to address the point at which the trawl survey no longer serves as a reliable index of abundance for low catchability species, so that more appropriate means of assessment and advice for such species can be sought. Further, conversion coefficients can be determined for stocks for which the catch results from optimal and suboptimal gear settings differed most; such coefficients can be applied to the assessments.
8. Attempts to minimize the influence of survey variability in survey-only assessments have been made through the use of running averages and smoothers. This technique trades reduced survey variability against the ability to detect rapid change. The risk of inducing stock collapse through over predicting stock size is much greater at low than at high stock size, something currently not taken into account. A precautionary approach to management might use estimates of survey CV to produce a more risk-averse strategy.
9. The panel also identified another use of the survey time-series of "other species", namely that anomalously low catch rates of species such as scallops and skates in certain years may be indicative of poor fishing practice by the research vessel in those years. This issue could be investigated through the current time-series.
10. The review panel was unanimous in stressing the fundamentally different objectives of commercial fishing and scientific surveying. Simply, there is no drive in research surveys to increase catchability, whereas this aim is uppermost in the minds of commercial fishers. This difference in itself does not provide reason to reject the time-series of abundance or to modify survey design. A long data series is of inestimable value to scientific endeavour, and the coherent NEFSC series is one of the longest of its type in the world. Therefore, if survey design needs to be reviewed in light of recent experience, and the panel concluded that there was not conclusive

evidence to do so, it must be done with the involvement of stakeholders. Further, independent scientists with a knowledge of survey design should be included at an early stage if redesign is being countenanced, to ensure that scientific standards are not compromised and that the results of any research survey conducted with amended design are comparable with those of the current design.

11. TOR 1D seeks comment on the precision of assessments in relation to the trawl survey, a request worthy of review in its own right. First, it must be emphasized that the precision of model-based calculations of stock size and fishing mortality rates reported in the GARM report are comparable with the precision in many other fisheries stock assessments, e.g. at ICES and ICCAT. Second, the error variability in trawl survey indices did not increase systematically with the trawl warp offset. Third, the sensitivity analyses conducted on assessments assuming bias in the survey indicate that the determination of stock status is not changed by any likely effect of warp offset. However, as the differences between B_{2001} and F_{2001} and their respective reference points are usually quite large, this result is not surprising. Perhaps what is more important are the distances between current estimates of B and F and their reference points, because these will determine the strength of a management response; this situation needs investigation. A particular concern is the question of the extent to which estimation errors in current F and B trigger the wrong management response. This is a problem not confined to the warp-offset issue, but relates more generally as to how assessment error should be taken into account when making judgements about management action. It does appear that estimation error can trigger a stock-rebuilding response even when there is no need to do so. Given the potential economic impact of this, some thought should be given as to how assessment error can be handled within the management process. Overall and very importantly, though, the panel has no reason to doubt that the precision of model-based calculations of stock size and fishing mortality rate use appropriate methodology to take into account variability in trawl survey results.
12. TOR 1D also questions whether the methods for incorporating uncertainty in survey results into management advice are adequate. The methods used by the NEFSC include stock assessment modelling of alternative scenarios for the key assessment assumptions, and retrospective analysis, the latter providing a useful method of interpreting both the degree of reliability and uncertainty in assessment results. Both are useful and in this case properly applied. The AGEPRO modelling software applied by NEFSC also provides a methodologically consistent means of incorporating the uncertainties modelled in the ADAPT VPA and of projecting the estimates into the future. It may also, however, be possible to develop it further so that implementation uncertainty is catered for, something currently not explicitly done by the software. Despite the above, however, the current management framework arising from interpretation of the law demands, at face value, far more precision and knowledge of the stocks than is deliverable, so the question could just as easily be asked whether the management system is adequately tailored to the degree of uncertainty inherent in current knowledge. As total biomass is the quantity least under the control of managers, because transient population change is influenced more by recruitment than mortality, it must be queried whether a rebuilding strategy

should be so heavily predicated on a biomass target determined with substantial uncertainty and on a quantity over which managers have so little direct control.

TOR 2 – BIOLOGICAL REFERENCE POINTS

Review the fishing mortality and biomass targets and thresholds established for the 20 groundfish stocks included in the Northeast Multispecies FMP. Consider the adequacy of technical analyses supporting estimates of F_{MSY}, B_{MSY} or their proxies, as provided in the* Report of the Working Group on Re-Estimation of Biological Reference Points for New England Groundfish Stocks *(the "Report"). Comment on issues related to the simultaneous achievement of B_{MSY} values for the groundfish complex.

1. The latest guidelines for developing reference points to manage New England groundfish recommend that age-structured models be applied if there is sufficient information to do so. Simply, the management framework is based on the concept of maximum sustainable yield (MSY) and the requirement to rebuild to B_{MSY}. In its purest form, this theory requires knowledge of the stock/recruit relationship, determination of which, for many fish stocks, is highly uncertain. An additional problem is that many stock-size observations have been made only during periods of high exploitation, when biomass is likely to have been restricted to the lower left hand side of the relationship. Thus, even if there are adequate data to fit a stock/recruit curve (albeit over a restricted biomass range), the biomass and recruitment values associated with MSY probably lie outside the range of the observations. Hence, calculation of MSY conditioned on such a relationship effectively forces extrapolation beyond the range of the data, with consequent loss of credibility. Solving these problems requires a realistic interpretation of MSY in the light of incomplete data and knowledge.
2. In addition, most stock/recruit functions use spawning stock biomass (SSB) as a proxy for stock reproductive potential, but there is increasing evidence that use of this proxy can overestimate a stock's resilience to fishing pressure. High fishing pressure can result in a truncated age structure, resulting in the spawning potential at low biomass falling below that predicted by a stock/recruit curve and longer-than-anticipated rebuilding times. The stock/recruit function that results in the lowest B_{MSY} will likely result in the least short-term "pain" for industry, but it may lead to underutilization of the resource and actually increase the risk to the stock. The function resulting in the highest B_{MSY}, while minimizing the risk to the stock, may in fact be unrealistic, resulting in ever more stringent management action while the target stays unattainable. One advantage of aiming for a higher B_{MSY} is that the true value is likely to be found en route, though doing this could be viewed as implementing "precautionary science" rather than "precautionary management".
3. In order to address the stock/recruit problem, considerable effort has been directed towards selecting models to fit existing data for as many stocks as possible. The models in use in the US are essentially variants of the Beverton-Holt (asymptotic) and Ricker (over-compensatory) types. "The Report" acknowledges the advantages and disadvantages of both. NEFSC model selection has been based partly on statistical diagnostics and partly on the credibility of MSY-related values in relation

to other values, such as F_{max} and the range of observed landings, and overall the basis for model selection seems sensible. However, the estimated values for B_{MSY} are generally higher than the largest biomass observed historically, meaning that there is little information in the stock/recruit time-series to inform the calculation of MSY. This applies also to the use of $F_{40\%}$ to estimate B_{MSY}, where it has been used in conjunction to model recruitment. As a result, estimates of B_{MSY} and F_{MSY} are subject to considerable uncertainty.

4. From a management perspective, the implications of uncertainty in F and B are not the same. Common sense can be used to judge whether F_{MSY} is obtainable because fishing mortality is entirely man-induced. The main principle behind the choice of F_{MSY} is to achieve a productive fishery with low risk to the resource. Therefore, a value of F can be selected that approximates to F_{MSY} without recourse to stock/recruit data, as the $F_{40\%}$ proxy does, and there is limited scope for concern about its uncertainty. In stark contrast, however, where B_{MSY} estimates are beyond the range of observations, there is the possibility that not only is the target not where it is thought to be, but also that it may simply be unobtainable. Further, it is obvious that, if B_{MSY} is not well estimated, then any associated $F_{rebuild}$ will be inappropriate. From this perspective, if the current B_{MSY} reference points are adopted/retained, it is crucial that the potential costs of adopting erroneous values be evaluated rigorously. Moreover, a pragmatic solution to covering the substantial uncertainty in estimates of B_{MSY}, where for example they fall outside the observation range, could be to set them at a value closer to the range of actual observations. For example, a B_{MSY} proxy could be set provisionally as the largest observed value in the historical time-series or some conventionally chosen percentage above the largest value. Then, once more years of data have been obtained, the proxy can be re-evaluated (this is a form of adaptive management advice, a protocol strongly supported below).
5. In light of the above, the question of choosing between Beverton-Holt or Ricker curves is somewhat redundant, although the panel found the selection analysis carried out by NEFSC and in the GARM report to be well done. Simply, while choice of curve has a substantial impact on the calculation of B_{MSY}, neither can be supported by the data in the region of predicted MSY. However, beyond the basic aspect of model choice, it must be stressed that there are two principal assumptions regarding the biological attributes underlying the models, that they are well estimated and stationary. Neither issue was investigated in detail, either in terms of the uncertainty in the underlying attribute or in their impact on the determination of biological reference points. Notwithstanding the above, if the issue of model selection is to remain on the table, then model validation will need to be addressed more fully and more divergent models tested, e.g. non-parametric deterministic and probabilistic.
6. To address some technical issues specifically, it is the opinion of the panel that most but not all of the methodologies currently used by the NEFSC to compute F_{MSY} and B_{MSY} provide an adequate scientific basis for fisheries management. However, AIC (or BIC) should not be used as a model selection criterion for the Bayesian statistical models used. Bayes' factor or Bayes' posterior is preferable. Further, where Bayes' factor was used, the method of implementation was incorrect. The AIC value was incorrectly applied to compute the marginal posterior and Bayes' factor for each

alternative model. Instead, the marginal probability of the data given each stock/recruit function should be used to compute Bayes' factors. This is a goodness-of-fit measure integrated across the entire parameter space of the model, rather than only at the posterior mode, as produced by AIC.

7. In certain analyses conducted by the NEFSC, MCMC software was applied for statistical estimation. Unless the results for different types of diagnostic method are satisfactory, then there is an inadequate basis to know whether the MCMC results obtained have converged on the posterior distribution and are therefore reliable. No convergence diagnostics were reported, so it is suggested that more than one diagnostic tool be applied to evaluate convergence, and that those provided in the WinBugs software be used.
8. Models with autocorrelation in recruitment should not be tested statistically against models without such correlation using the current time-series of stock/recruit data, because the time-series are too short to permit reliable statistical detection of autocorrelation. Therefore, consideration needs to be given to the derivation of other objective criteria that might be appropriate for determining when to include autocorrelation in models to determine F_{MSY} and B_{MSY}, and in stock projection models.
9. Most of the stock assessment and projection methodologies currently applied by the NEFSC provide an adequate scientific basis for fisheries management. The ADAPT VPA and AGEPRO methodologies provide a rigorous and adequate basis for assessing stock biomass and fishing mortality rate, making projections, evaluating the differences in potential consequences of alternative possible fisheries management policies, and for taking into account parameter and important model structure uncertainties.
10. It is necessary to point out here that the ASPIC surplus production estimation and projection methodology has some serious methodological limitations. The software cannot incorporate a prior probability distribution for the intrinsic rate of increase (r) that could be derived from meta-analysis or demographic analysis, as can other more-recent methods of surplus production modelling. This is a serious problem because of the relatively small store of information in relative abundance time-series typically available for jointly estimating r and carrying capacity (K). As a result, the imprecision in parameter estimates could lead to poor assessments of stock status and seriously biased predictions of the stock's response to current and new management regulations. The use of a probabilistic methodology for estimating r, using demographic data or meta-analysis and a projection methodology that accounts probabilistically for uncertainty in r, should therefore be considered as a potential improvement to the current ASPIC methodology.
11. During the meeting, Butterworth and his co-workers presented an alternative modelling approach for two cod stocks, using what they termed an "age-structured stock production model" (ASPM) to reconstruct the population history and to estimate MSY reference points. The analyses were not intended to be definitive, and some sample runs were made during the meeting for illustrative purposes. Indeed, some panel members felt that the analyses had not been produced with the necessary rigour and attention to detail to warrant serious consideration as an alternative to current methodology, although they concede that they clearly provide an alternative

perspective on the calculation of reference points. One point of concern to the panel was that the ASPM seems to be overly sensitive to the assumptions made. For instance, the value for the steepness parameter h was one issue that could not be resolved between participants at the meeting. Nevertheless, although important differences about the error structure of the data are made in Butterworth's analyses, the underlying population dynamics model and data used are essentially the same as in the NEFSC assessments. Therefore, if configured similarly, the two approaches would be expected intuitively to yield at least similar results. Differences may well be caused by an insufficiency of the basic observed information to estimate the parameters of the population model well, so there would definitely be value in investigating the ASPM and ADAPT-based approaches to better understand differences between them. At present, there are no obvious reasons to choose one method over the other, to deviate from the current approach, or indeed to select a completely different model/method from the two presented. However, consistency in using one model to estimate reference points and the same one to judge stock status was deemed by the panel to be crucial to the provision of good management advice.

12. Some emphasis was placed in the meeting on the relevance of natural mortality (M) in the assessment and calculation of MSY reference points. MSY can be sensitive to the choice of M, but great care is required in altering M in assessments where no real information exists on its magnitude. Specifically, the fact that the ASPM method is capable of estimating M from the data does not make its estimate reliable. The majority view of the panel was that it is often better to fix M externally than to attempt to estimate it, and that, because the main problem with reference points is the extrapolation of biomass values beyond the range of the data, any attempts to improve the estimates of M further may be inappropriate, given other uncertainties.
13. As well as more conventional age-based assessment methods, "the Report" introduces a novel approach to the use of survey indices for stocks where VPA is not available. The method is innovative, elegant, and shows promise. Interestingly, in implementing the method, flexible judgements have been made about the choice of B_{MSY} proxies. Validation is now vital and the method must be rigorously tested. Necessary but arbitrary choices are made, so it is essential to demonstrate that implementation would drive the stock in the direction desired. A simulation study is therefore required to validate the survey index method in a management context, and to investigate its sensitivity to *ad hoc* assumptions. The fact too that the survey method is rooted in empirical data and does not make external assumptions about stock and recruitment means that it would be interesting to see if it can be applied to all stocks, even those managed on the basis of VPA. If this is done, it could provide a consistent framework across most of the stocks managed according to NEFSC advice and avoid some of the weaknesses of methods that make assumptions about stock and recruitment. The method should be aired through the scientific literature.
14. If the intention of New England fishery management is to achieve a long-term, sustainable MSY, then fishing at F_{MSY} should achieve that. Fishing at a value of F higher than F_{MSY} will not allow the stock to build to a level capable of supporting MSY. With this in mind, the change in the overfishing threshold from $F_{20\%}$ to $F_{40\%}$ appears to be related to a change of emphasis from a limit reference point ($F_{20\%}$), i.e. an area to be avoided, to a target one, i.e. an area to attain, which is primarily a

change in management protocol, rather than one driven by the science. $F_{20\%}$ is traditionally regarded as a proxy for deriving minimum spawning biomass, whereas $F_{40\%}$ is regarded as a value to derive B_{MSY}, and is therefore more consistent with identifying a target. Scientifically, and taking only biological risk into account, it is considered more important to be above a limit than at a target and this fact should be used to inform the strength of management response to the proposed reference points. Nevertheless, the extensive simulations made and on which the decision to change from $F_{20\%}$ to $F_{40\%}$ was founded appeared to the panellists to have been scientifically sound. Further, the choice of a conservative $F_{50\%}$ basis for the reference points for Acadian redfish is totally appropriate given the species' slow growth rate and inherently slow rate of increase. In summary, therefore, the panel agreed that the proxies for F_{MSY} based on both $F_{40\%}$ and $F_{50\%}$ for the various groundfish species appear to have sufficient scientific basis for the purposes of New England fisheries management, judged on the basis of considerable previous modelling and empirical work**.**

15. At a practical level, more is known about the region of undesirable stock size than about the region of desirable (and attainable) stock size for New England fish stocks. This presents both scientists and managers with considerable challenges if the law requires that stocks be rebuilt to a B_{MSY} target within a specified time frame. The current state of knowledge means that it is easier to know the direction in which stock biomass should be driven than when or if B_{MSY} can be attained, or even perhaps if such a target is overly conservative. Given this, the proposal by the working group to adopt an adaptive approach to biomass management seems eminently sensible. Provided target F values are achieved, biomass levels should evolve naturally to a state of B_{MSY} or through a process of rebuilding not dependent on theoretical levels of biomass.
16. What is not clear now is whether it is possible to move all 19 stocks simultaneously to B_{MSY} through single-species management. From the evidence presented during the meeting, there do not appear to be strong trophic linkages within most of the species assemblages analysed, although extension of some of the analyses to include known prey species such as shrimp and sandlance could be informative in determining the sensitivity of the system and the analyses to strong biological interactions. Competition is likely to have a more potent effect at large stock size, but the data series provided to the panel do not extend back far enough to demonstrate definitively that the system can sustain all stocks at B_{MSY}, or even much above current levels. As each calculation for every stock assumes fixed schedules of growth, maturity and natural mortality that may well be adequate when interpolating but may not be totally realistic extrapolating outside the range of observations, answering the oft-put question of whether the desired targets are attainable in practice (i.e. in nature) is extremely difficult. Given the complexity of biological interactions and the ecosystem in which the stocks exist, one doubts whether the question can indeed be answered satisfactorily with present data and knowledge. However, experience in Europe suggests that single-species MSYs are not good indicators of multispecies MSY, so caution is certainly needed in the choice of B_{MSY} targets. Also, there could be good mileage in assessing the interactions in

multispecies fisheries, for which spatially heterogenous models (e.g. Iceland's BORMICON) will be required.

TOR 3 – STOCK REBUILDING AND RELATED PROJECTIONS

The Sustainable Fisheries Act requires that various resources be rebuilt to B_{MSY} in no more than 10 years, unless life history attributes of individual stocks dictate a longer rebuilding period (e.g. Georges Bank cod, Acadian redfish). Considering the uncertainty in stock dynamics and the ability to achieve target rebuilding fishing mortality rates for all stocks in the complex simultaneously, comment on the stock projection methodology used to advise on management strategies intended to achieve stock rebuilding goals.

1. The AGEPRO software used to forecast rebuilding strategies by NEFSC is well documented and rigorously constructed, and it interfaces well with the ADAPT VPA stock assessment model and the bootstrapping output produced by ADAPT VPA to take into account parameter uncertainty. It certainly allows for uncertainty in the estimates of initial population size, natural mortality and future recruitment, and has additional potential for examining autocorrelation in recruitment. While this covers some of the uncertainty likely in projections, there could well be uncertainty associated with variations in weight, maturity and fishing mortality that widen the confidence limits further. Of course, provided there is no bias to the missing uncertainties, the mean trajectory will remain unaffected, but the confidence in the rebuilding protocol would then tend to be overestimated. Of more concern is the inability of the software to cope with parameter trends. Systematic changes in parameters such as weight at age and maturity at age can cause significant bias in stock projections, e.g., a declining trend in weight at age, if unaccounted for, would result in spawning stock biomass being overestimated and an overoptimistic probability of recovery. No information on historical age-based parameters is given in the GARM report, so the likely impact of these factors is impossible to assess.
2. The stock projection methodology used is designed to answer the question "what constant *fishing mortality* is required to rebuild the stock to B_{MSY} with 50% probability". This question is not the same as "what *management action* is required to rebuild the stock to B_{MSY} with 50% probability". To answer the latter question it is necessary to model stock assessment and management implementation errors, and these features are not built into the projection methodology. It must be stressed here that the NEFSC calculations indicate the fishing mortality required *provided it could actually be achieved*, regardless of assessment and implementation error. If the two latter sources of error are truly random, then their effect is probably on the width of the envelope of uncertainty and will tend to have a small effect on the 50% probability chosen. Unfortunately, the errors are almost certain to have a tendency to be in one direction (e.g. to overestimate stock size and to underestimate the effectiveness of management action). This would mean that such projections run the risk of being overly optimistic.
3. Some comment is necessary here on the uncertainty in projections. Four types of uncertainty may be considered, measurement, process, model and implementation.

The measurement error is well considered in the projections that have full VPA treatment, both parametric stock/recruit and empirical non-parametric. Such error is not obvious in those assessments that rely on catch survey data, but it may be possible to cover the issue with bootstrapping. Although measurement error is estimated in the production-model based projections, it is probably insignificant when compared with model assumptions regarding processes. Process error is an explicit parameter in the projection software, but details were not available on how it is estimated and used. However, it seemed not to include time or density-dependent effects, which could be included if identified. As mentioned above, neither model error nor implementation error have been adequately covered yet.

4. Notwithstanding the above, it is the considered opinion of the panel that, although the projection methodology does have limitations with respect to assessment and implementation error, it is probably good enough for the purpose of estimating a conventional value for $F_{rebuild}$. Attempting to model implementation error involves too many assumptions about largely unknown processes, because management of the stocks does not follow a defined management procedure. Thus, calculation of $F_{rebuild}$ should be seen as a conventional calculation based on an agreed recipe rather than one that attempts to capture all elements of reality.
5. Three technical aspects of the method are worth comment. First, randomization of the starting values of stock size for the projection appears to consider estimation error only. It is common for assessments to show bias, so affecting the value of $F_{rebuild}$. One way of dealing with this would be to examine the retrospective pattern of the assessment and to try to include non-random behaviour in the process. Second, natural mortality M is assumed to vary randomly and independently, but because M is dominated by predation, it could show autocorrelation as a result of predator populations changing gradually over time. As the current fishery management plan is to increase predator populations, there may be merit in considering simulations that take into account autocorrelation in M. Such effects could be large over periods of a decade. Finally, the stock/recruit model used in the projections was the same as that used to estimate the reference points, which in itself is desirable. However, the model was chosen partly on criteria assigned with the purpose of estimating MSY, which involves the right hand side of the curve. Rebuilding projections use more of the left hand limb of the relationship, and it is important that the choice of model be consistent with the dynamics of the stock at lower stock sizes. Given this, validation of the models against historical observations of stock dynamics assumes huge importance.
6. Butterworth showed the potential (economic) value of rebuilding strategies in which F reductions are phased-in over the rebuilding period (customarily 10 years, though the actual rebuild starting date and duration was never clarified at the meeting). There is no scientific reason why such an approach could not be adopted, but the choice of strategy (and the starting date and duration) would have to be a management decision, not a scientific one. Nevertheless, if such a strategy stages the F reductions so that the initial drop is small, then the prevailing risk to the stock will likely be higher than a strategy where all the "pain" is taken at the start of the rebuilding period. In addition, and purely from a scientific point of view, the setting of an intermediate biomass target could be defended. In such a case, once a rebuilding F and trajectories were in

place, this intermediate target could be used as a signpost to see if stock rebuilding was on track. If it is deemed not on track, an assessment would be needed to determine why, e.g., were the original projections in error, was recruitment atypical, or was recruitment as expected but F not achieved?

7. Whatever rebuilding strategy is set, and adaptive strategies again seem to be the way to proceed, its performance needs to be assessed regularly in light of information on recruitment level, discarding, etc. Of course, whether the rebuilding target is being achieved or not, institutional and legal constraints are invoked, although this is not strictly an issue of scientific concern. What would be of scientific concern, however, is the regularity of such assessment updates; too frequent could mean disruption to the natural scientific process, but too infrequent could mean loss of yield or irrevocable downturn in the stock.
8. Finally, all analyses discussed at the meeting were strictly single species, so the implication is that such analyses, especially those identifying F_{target} or $F_{rebuild}$, would require single-species management. NEFSC scientists acknowledge correctly that, in practice in the generally mixed species fishery of New England, such a requirement is not achievable because of the mix of species taken in differing quantities by different fleets. Therefore, while the target F for individual stocks can be calculated, such targets may be impossible to achieve simultaneously, because differing fleet activity constrains management. As a consequence, the limitations on achieving multiple targets in a mixed fishery need to be quantified clearly for decision-makers. This could be done by modelling Fs by fleet to evaluate the technical interactions between them and hence identifying the constraints on achieving multiple targets. Evaluating technical interactions in the mixed fishery would therefore be a valuable exercise while fisheries management continues trying to attain a multitude of MSY targets.

Final comments

It is not for me to judge how well the process succeeded in meeting its objectives, though it was certainly within my power to try to steer it in the right direction. I do believe, however, that the review succeeded in meeting most of the objectives it was set, and hopefully many of the suggestions made in both this summary and in the reports of individual reviewers will receive the consideration they deserve. Hopefully too, several lessons will have been learned from the experience of this review, not least of which is the absolute necessity to provide independent reviewers access to relevant documentation in good time and to agree between all customers and stakeholders and with the reviewers themselves clear Terms of Reference *before* the meeting. It was extremely frustrating to have to admit to stakeholders and interested parties several times during the meeting that we were not that clear about what we (the independent reviewers) were being asked to advise on, although there was gratifyingly total acceptance that, as scientists, we were not competent or indeed expected to advise on matters of policy.

The appointment of a trained moderator for the public meeting was, in retrospect, a masterstroke. Don Perkins converted what could easily have been an uncomfortable clash between people with different expectations and constituencies into a forum for interactive

debate that was notable most for its rarity in similar-tasked organizations around the world. His knowledge of the issues was commendable, seeing that he is not a marine scientist by training, and personally I was reassured by his attention to detail and overall fairness to all participants. He ran the meeting efficiently and to time, and I enjoyed the experience of "chairing" a meeting that someone else ran.

Writing the summary was one of the most difficult tasks I have ever had to undertake in fisheries management science, not because of a lack of understanding of the processes involved, but because the nature of scientific evaluation necessarily takes many sides. Allied to this of course was the virtual impossibility of absorbing the voluminous documentation in the time available before and at the meeting. In this summary, I have therefore erred on the side of being liberal with comment, in an attempt at incorporating as many of the independent views expressed as possible. However, time pressures also meant that, to meet the very tight deadline for production of the summary, I may have shown bias towards the report contents and views of the reviewer's report that was submitted to me earliest, namely Dr Cook's, and perhaps given less weight to the last and technically most voluminous, Dr McAllister's. If this comes out in the contents of the summary report, then I apologize to both of them. It was certainly not deliberate on my part, and I specifically guarded against omitting from this summary valuable input contained in the individual reports or, more important, placing too much emphasis on my personal scientific perspective of any situation. The summary report should be read with these conditions in mind; further (largely technical) detail can be found in the four individual reports.

In my capacity as Chair of the Groundfish Science Peer Review panel, I thank Drs Bell, Cook, McAllister and Mohn for their positive attitude towards the task in hand and for being so willing to share their views and, in certain cases, to delve deep into their scientific knowledge and experience to explain the presented science in terms clear to the rest of us. The quality of their reports is clear for all to see, and working with all of them was a highlight. Chairing the panel itself was an interesting and humbling experience, and I thank Dr David Die for persuading me to even consider taking on the task as well as the CIE for their confidence in affording me what turned out to be a privilege.

APPENDIX 1

STATEMENT OF WORK

Consulting Agreement between the University of Miami and Dr Andrew Payne

January 7, 2003

Introduction

This document presents terms of reference for peer review of the stock assessment and population dynamics science supporting the New England Fishery Management Council's (NEFMC's) Northeast Multispecies Fishery Management Plan (FMP). Specifically, the review will focus on three major terms of reference:

- Effects on the accuracy and present usefulness of trawl survey data due to uneven trawl warps and other recently discovered gear-related trawl survey problems. These evaluations will be based on gear testing cruises and related workshops conducted during autumn 2002 as well as any other information available to the reviewers.
- Estimates of stock biomass and fishing mortality targets and thresholds for the complex of stocks comprising the groundfish resource, and,
- The adequacy of projections of stock rebuilding to achieve the biomass targets, consistent with time frames as mandated under the Sustainable Fisheries Act.

The staff of the NEFMC originally proposed these three focus areas. Specific comments appropriate to three terms of reference provide guidance to the review committee recognizing that reviewers are likely to be unfamiliar with the specifics of the Northeast Multispecies FMP and, the provisions of the Sustainable Fisheries Act. Overall, the terms of reference generally concentrate on the adequacy of the science currently available to support fishery management plan development.

For each subject area, a brief objective statement is provided to give an overall context for the terms of reference to the reviewers. Within these subject areas, specific questions are provided with the intent of providing a minimum set of questions to consider in formulating the group's responses. It is envisioned that the reviewers' responses will take the form of detailed reviews of the information and conclusions reached in the various supporting documents and verbal presentations made to the group, along with their own summaries and opinions regarding the adequacy of existing science in supporting fishery management decisions. The reviewers are encouraged to pay particular attention to alternative methods presented by independent experts, if any, in concluding whether the conclusions of the Report, or other approaches, represent the best science available.

1. Trawl survey issues and influence on management advice

Considering the results of the Groundfish Assessment Review Meeting (GARM), subsequent results from experimental trawl comparisons, and other appropriate information, provide an evaluation of the significance of potential differences in trawl survey catchability resulting from recently-discovered survey gear problems on management advice for groundfish stocks managed under the Northeast Multispecies Fishery Management Plan.

In responding, reviewers should consider the following:

A. Are conclusions regarding use of 2000-2002 trawl survey data adequately supported by analyses reported by the GARM? Were analyses sufficient to detect differences in survey catches arising from unequal warps and other survey problems? Did the sensitivity analyses presented in the GARM report adequately bound the range of potential effects inferred from analyses of historical and comparative data? Did the GARM adequately characterize the uncertainties in estimated stock sizes and rebuilding mortality rates potentially arising from unequal warp offsets?

B. Was the design and analysis of data from experimental trawl comparisons adequate to estimate the magnitude of differences resulting from the use of unequal trawl warps and other experimental treatments? Were estimates of the power of these experiments to detect statistical differences in fish catches between treatment and control survey configurations adequately described?

C. Advise on the significance of differences in species composition and relative catch rates resulting from side-by-side tows performed by commercial and government vessels in the recent trawl experiment with respect to model- and index-based estimates of stock size and fishing mortality rates.

D. Comment on the precision of model-based calculations of stock size and fishing mortality rates in relation to variability in trawl survey catches and other sources of information included in assessments. Are the methods used for incorporating uncertainty into management advice sufficient? How should other sources of uncertainty (e.g. model selection, estimates of total removals) be incorporated?

2. Biological reference points

Review the fishing mortality and biomass targets and thresholds established for the 20 groundfish stocks included in the Northeast Multispecies FMP. Consider the adequacy of technical analyses supporting estimates of F_{MSY}, B_{MSY} or their proxies, as provided in the *Report of the Working Group on Re-Estimation of Biological Reference Points for New England Groundfish Stocks* (the "Report"). Comment on

issues related to the simultaneous achievement of B_{MSY} values for the groundfish complex.

In responding, reviewers should consider the following. Of particular note, the NEFMC's Science and Statistical Committee recommended that additional work was needed "…specifically to explore the implications of the uncertainty in the stock recruitment relationship." For this reason, more specific questions are included in order to add clarity to the issues to be addressed by the reviewers.

A. Comment on the technical basis for the estimation of F_{MSY} and B_{MSY}, and choices regarding the use of parametric (Beverton-Holt, Ricker, other candidate models, etc.) and non-parametric stock-recruitment relationships applied to yield per recruit estimates, surplus production models, or proxies for biomass and fishing mortality rate targets and thresholds.

- Are the Working Group assumptions (growth, maturity ogive, natural mortality, partial recruitment) appropriate for estimating a B_{MSY} proxy, which establishes a minimum biomass threshold and a rebuilding target?

- Comment with reference to specific species on whether the use of Beverton-Holt type stock-recruitment curves, as opposed to the use of dome-shaped (Ricker type) curves, represent reasonable scientific judgment employing sound methodology and appropriate data sources. Is there a theoretical or practical basis to detect overcompensation (Ricker curve) from the stock-recruitment curve for each groundfish species based on the magnitude of the intrinsic rate of population increase (r) and the carrying capacity (K) parameter estimates from ASPIC production models?

- Could alternative non-equilibrium production models for groundfish species be examined for estimating F_{MSY} and B_{MSY} thresholds?

B. Comment on the justification for changing the overfishing threshold to $F_{40\%}$ (the proposed proxy for most groundfish stocks) from $F_{20\%}$ that generally defined overfishing before Amendment 9, or from the F_{MSY} estimates in Amendment 9? Are the proposed proxies for F_{MSY} (e.g. $F_{40\%\ MSP}$ for Georges Bank haddock, $F_{50\%\ MSP}$ for Acadian redfish) more appropriate to achieve MSY, given the groundfish stock dynamics? Are the proposed proxy reference points overly conservative or too liberal for a fishing mortality threshold that complies with the Magnuson-Stevens Act?

- Reconstruction of the theoretical S-R curve can be done indirectly for each groundfish species by merging results (YPR, SSB/R) from the Thompson-Bell yield-per-recruit model and expected equilibrium yield (mt) from various stock production models. Are the resulting F_{MSY} values similar to the $F_{40\%}$ values (e.g. for haddock) from the Y/R curve? Is $F_{40\%}$ a suitable proxy for F_{MSY} under these conditions?

C. Evaluate evidence for density-dependent regulation of population size (e.g., simultaneous occurrence of various stocks at higher population sizes, predator-prey, and growth rate information) for the groundfish complex. Are potential non-stationary stock dynamic processes (i.e. environmental variations in

recruitment survival) and/or trophic limitations adequately accounted for in estimates of B_{MSY}? Is there evidence that B_{MSY} values estimated for the 20 groundfish stocks cannot be simultaneously achieved?

3. Stock rebuilding and related projections

The Sustainable Fisheries Act requires that various resources be rebuilt to B_{MSY} in no more that 10 years, unless life history attributes of individual stocks dictate a longer rebuilding period (e.g. Georges Bank cod, Acadian redfish). Considering the uncertainty in stock dynamics and the ability to achieve target rebuilding fishing mortality rates for all stocks in the complex simultaneously, comment on stock projection methodology used to advise on management strategies intended to achieve stock rebuilding goals.

In responding, reviewers should consider the following:

A. Evaluate the adequacy of projection methods used to guide the attainment of B_{MSY}, specifically focusing on estimates of uncertainty in starting stock sizes, recruitment, and implementation uncertainty in the attainment of target fishing mortality rates. Comment on potential biases and precision of stock projection methodologies.

B. Are stock projection methodologies sufficient to distinguish the relative merits of various management scenarios?

C. The Magnuson-Stevens Act requires that overfished stocks be rebuilt to a biomass level consistent with producing the maximum sustainable yield from the fishery. Is there a scientific basis for arguing that an intermediate biomass target meets that requirement?

Schedule

The independent peer review is to be completed by March 1, 2003. In order to meet that deadline, the following review format and timeline is proposed.

3-5 February: Public workshop **(including participation of independent reviewers**) on the GARM Report and report of biological reference points during this week.
6-8 February: Independent reviewers meet in executive session to discuss results from the two workshops and supporting documentation.
10-14 February: Independent reviewers prepare their individual reports and submit them to the summarizer.
17-21 February: Summarizer prepares his/her report summarizing findings of individual reports prepared by panel members, which will be made available to the public.

The February 3–5, 2003 public workshop will begin with an introduction followed by a series of presentations summarizing the various documents presented to the panel. Open comment periods will allow for additional scientific input from various members of the public regarding additional analyses and comments. Peer reviewers will interact with agency and independent scientists and members of the public to ask appropriate questions and discuss results.

Specific

The consultant shall be provided with all background material required to prepare for the review, and the consultant shall chair the February 3–5, 2003 workshop, the February 6–8, 2003 executive session, receive all members' reports and summarize all findings in a summary report by February 21, 2003 for public release.

The consultant's duties shall not exceed a maximum total of 17 days: Several days prior to the workshop for document review; the three-day workshop; the three-day executive session; and several days following the meeting to prepare a report summarizing the findings of the individual reports prepared by panel members, which will be made available to the public.

The consultant's duties include:

1. Reading all background material provided;

2. Chairing the February 3–5, 2003 workshop on the Groundfish Assessment and Review Meeting (GARM) Report and report of biological reference points;

3. Chairing the February 6–8, 2003 executive session to discuss results from the two workshops and supporting documentation;

4. No later than February 21, 2003, submitting a report summarizing findings of individual reports prepared by panel members, which will be made available to the public. The report should be submitted to the workshop summarizer and to the CIE[1]; the CIE report should be addressed to the "University of Miami Independent System for Peer Review," and sent to Dr David Sampson, via email to david.sampson@oregonstate.edu, and to Mr Manoj Shivlani, via email to mshivlani@rsmas.miami.edu.

[1] The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

APPENDIX 2

TERMS OF REFERENCE
GROUNDFISH SCIENCE PEER REVIEW, FEBRUARY 3-8, 2003

Introduction

This document presents terms of reference for peer review of the stock assessment and population dynamics science supporting the New England Fishery Management Council's (NEFMC's) Northeast Multispecies Fishery Management Plan (FMP). Specifically, the review will focus on three major terms of reference:

- Effects on the accuracy and present usefulness of trawl survey data due to uneven trawl warps and other recently discovered gear-related trawl survey problems. These evaluations will be based on gear-testing cruises and related workshops conducted during autumn 2002 as well as any other information available to the reviewers.
- Estimates of stock biomass and fishing mortality targets and thresholds for the complex of stocks constituting the groundfish resource, and,
- The adequacy of projections of stock rebuilding to achieve the biomass targets, consistent with time frames mandated under the Sustainable Fisheries Act.

The staff of the NEFMC originally proposed these three focus areas. Specific comments appropriate to the three terms of reference provide guidance to the review committee, recognizing that the independent reviewers are likely to be unfamiliar with the specifics of the Northeast Multispecies FMP and the provisions of the Sustainable Fisheries Act. Overall, the terms of reference generally concentrate on the adequacy of the science currently available to support FMP development.

For each subject area, a brief objective statement is provided to give an overall context for the terms of reference to the reviewers. Within these subject areas, specific questions are provided with the intent of providing a minimum set of questions to consider in formulating the group's responses. It is envisioned that the reviewers' responses will take the form of detailed reviews of the information and conclusions reached in the various supporting documents and verbal presentations made to the group, along with their own summaries and opinions regarding the adequacy of existing science in supporting fishery management decisions. The reviewers are encouraged to pay particular attention to alternative methods presented by the independent experts, if any, in concluding whether the conclusions of the Report, or other approaches, represent the best science available.

TERMS OF REFERENCE

1. TRAWL SURVEY ISSUES AND INFLUENCE ON MANAGEMENT ADVICE

Considering the results of the Groundfish Assessment Review Meeting (GARM), subsequent results from experimental trawl comparisons, and other appropriate information, provide an evaluation of the significance of potential differences in trawl survey catchability resulting from recently discovered survey gear problems on management advice for groundfish stocks managed under the Northeast Multispecies Fishery Management Plan.

In responding, reviewers should consider the following:

A. Are conclusions regarding the use of 2000-2002 trawl survey data adequately supported by analyses reported by the GARM? Were those analyses sufficient to detect differences in survey catches arising from unequal warps and other survey problems? Did the sensitivity analyses presented in the GARM report adequately bound the range of potential effects inferred from analyses of historical and comparative data? Did the GARM adequately characterize the uncertainties in estimated stock sizes and rebuilding mortality rates potentially arising from unequal warp offsets?

B. Was the design and analysis of data from experimental trawl comparisons adequate to estimate the magnitude of differences resulting from the use of unequal trawl warps and other experimental treatments? Were estimates of the power of these experiments to detect statistical differences in fish catches between treatment and control survey configurations adequately described?

C. Advise on the significance of differences in species composition and relative catch rates resulting from side-by-side tows performed by commercial and government vessels in the recent trawl experiment with respect to model- and index-based estimates of stock size and fishing mortality rates.

D. Comment on the precision of model-based calculations of stock size and fishing mortality rates in relation to variability in trawl survey catches and other sources of information included in assessments. Are the methods used for incorporating uncertainty into management advice sufficient? How should other sources of uncertainty (e.g. model selection, estimates of total removals) be incorporated?

2. BIOLOGICAL REFERENCE POINTS

Review the fishing mortality and biomass targets and thresholds established for the 20 groundfish stocks included in the Northeast Multispecies FMP. Consider the adequacy of technical analyses supporting estimates of F_{MSY}, B_{MSY} or their proxies, as provided in the *Report of the Working Group on Re-Estimation of Biological Reference Points for New England Groundfish Stocks* (the "Report"). Comment on issues related to the simultaneous achievement of B_{MSY} values for the groundfish complex.

In responding, reviewers should consider the following. Of particular note, the NEFMC's Science and Statistical Committee recommended that additional work was needed "…specifically to explore the implications of the uncertainty in the stock recruitment relationship." For this reason, more specific questions are included in order to add clarity to the issues to be addressed by the reviewers.

A. Comment on the technical basis for the estimation of F_{MSY} and B_{MSY}, and choices regarding the use of parametric (Beverton-Holt, Ricker, other candidate models, etc.) and non-parametric stock-recruitment relationships applied to yield per recruit estimates, surplus production models, or proxies for biomass and fishing mortality rate targets and thresholds.

- Are the Working Group assumptions (growth, maturity ogive, natural mortality, partial recruitment) appropriate for estimating a B_{MSY} proxy that establishes a minimum biomass threshold and a rebuilding target?

- Comment with reference to specific species on whether the use of Beverton-Holt type stock-recruitment curves, as opposed to the use of dome-shaped (Ricker type) curves, represent reasonable scientific judgment employing sound methodology and appropriate data sources. Is there a theoretical or practical basis to detect overcompensation (Ricker curve) from the stock-recruitment curve for each groundfish species based on the magnitude of the intrinsic rate of population increase (r) and the carrying capacity (K) parameter estimates from ASPIC production models?

- Could alternative non-equilibrium production models for groundfish species be examined for estimating F_{MSY} and B_{MSY} thresholds?

B. Comment on the justification for changing the overfishing threshold to $F_{40\%}$ (the proposed proxy for most groundfish stocks) from the $F_{20\%}$ that generally defined overfishing before Amendment 9, or from the F_{MSY} estimates in Amendment 9? Are the proposed proxies for F_{MSY} (e.g. $F_{40\%\ MSP}$ for Georges Bank haddock, $F_{50\%\ MSP}$ for Acadian redfish) more appropriate to achieve MSY, given the groundfish stock dynamics? Are the proposed proxy reference points overly conservative or too liberal for a fishing mortality threshold that complies with the Magnuson-Stevens Act?

- Reconstruction of the theoretical S-R curve can be done indirectly for each groundfish species by merging results (YPR, SSB/R) from the Thompson-Bell yield-per-recruit model and

expected equilibrium yield (mt) from various stock production models. Are the resulting F_{MSY} values similar to the $F_{40\%}$ values (e.g. for haddock) from the Y/R curve? Is $F_{40\%}$ a suitable proxy for F_{MSY} under these conditions?

C. Evaluate evidence for density-dependent regulation of population size (e.g. simultaneous occurrence of various stocks at higher population sizes, predator-prey, and growth rate information) for the groundfish complex. Are potential non-stationary stock dynamic processes (i.e. environmental variations in recruitment survival) and/or trophic limitations adequately accounted for in estimates of B_{MSY}? Is there evidence that B_{MSY} values estimated for the 20 groundfish stocks cannot be simultaneously achieved?

3. STOCK REBUILDING AND RELATED PROJECTIONS

The Sustainable Fisheries Act requires that various resources be rebuilt to B_{MSY} in no more than 10 years, unless life history attributes of individual stocks dictate a longer rebuilding period (e.g. Georges Bank cod, Acadian redfish). Considering the uncertainty in stock dynamics and the ability to achieve target rebuilding fishing mortality rates for all stocks in the complex simultaneously, comment on the stock projection methodology used to advise on management strategies intended to achieve stock rebuilding goals.

In responding, reviewers should consider the following:

A. Evaluate the adequacy of projection methods used to guide the attainment of B_{MSY}, specifically focusing on estimates of uncertainty in starting stock sizes, recruitment, and implementation uncertainty in the attainment of target fishing mortality rates. Comment on potential biases and precision of stock projection methodologies.

B. Are stock projection methodologies sufficient to distinguish the relative merits of various management scenarios?

C. The Magnuson-Stevens Act requires that overfished stocks be rebuilt to a biomass level consistent with producing the maximum sustainable yield from the fishery. Is there a scientific basis for arguing that an intermediate biomass target meets that requirement?

APPENDIX 3

REVIEWERS AND AGENDA
GROUNDFISH SCIENCE PEER REVIEW (GPR)

Public Meeting – 3-5 February 2003, New England Center, University of New Hampshire, Durham, New Hampshire
http://www.necc.unh.edu/

Independent Peer Reviewers (contracted through the Center for Independent Experts (CIE: University of Miami)

Dr Ewen Bell, Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, England

Dr Robin Cook, FRS Marine Laboratory, Aberdeen, Scotland

Dr Murdoch McAllister, Imperial College, London, England

Dr Robert Mohn, Department of Fisheries and Oceans, Halifax, NS, Canada

Dr Andrew Payne (Chair), Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, England

Public Session Moderator

Mr Don Perkins, Gulf of Maine Aquarium

AGENDA - modified during meeting to accommodate participants' availability

Monday, 3 February

0900-1700 Public Session – Topic: **Trawl Survey Issues**

Background Documents:

-Report of the Workshop on Trawl Warp Effects on Gear Performance
http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0215/

- Report of the Groundfish Assessment Review Meeting
http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0216/

- Report of the Trawl Survey Experiment Workshop
Available online

- Other contributed documents

Order of the Day:

Introduction of peer reviewers, presentation of terms of reference, and discussion of ground rules (Moderator)

Formal Presentations:

- *An overview of trawl survey issues* – **Russell Brown** (30 minutes)
- *Intervention analyses to detect trawl warp offset problems, sensitivity analyses, scale of potential offset factors*- **Paul Rago** (1 hour)
- *Trawl warp and related experiments*- **Michael Fogarty** (1 hour)
- *An evaluation of Paul Starr's analysis of the fishing gear experiment*- **Doug Butterworth** (20 minutes)
- *Comparison of length composition data from trawl experiments*- **Tom Nies** (30 minutes)

Facilitated discussion regarding presented materials in relation to terms of reference (all)

Tuesday, 4 February

0800-1700 Public Session – Topic: **Biological Reference Points**

Background Documents:

- Report of the Working Group on Re-Evaluation of Biological Reference Points for New England Groundfish

http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0204/

- Report of the Overfishing Definition Review Panel:
http://www.nefmc.org/documents/overfishing/

- Report of SAW 36

- Report of the Groundfish Assessment Review Meeting:
http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0216/

- NEFMC Council Meeting Report for July 2002, summarizing Scientific and Statistical Committee review of re-estimated reference points

http://www.nefmc.org
(Go to "News and Motions," then click on "Council Reports")

- Other contributed documents

Order of the Day:

Formal Presentations:

- *Re-Evaluation of biological reference points: goals and objectives*- **Steven Murawski** (1 hour)
- *A Strategy to evaluate alternative stock-recruitment models*- **Christopher Legault** (30 minutes)
- *Evidence for density-dependence in species and ecosystem responses*- **Ralph Mayo** (30 minutes**)**
- *An age-structured production model based assessment and reference point evaluation for the Gulf of Maine cod stock*- **Doug Butterworth** (1 hour)
- *Decision analyses using biological reference points in evaluating groundfish stock status*- **Yong Chen** (30 minutes)
- *Overfishing thresholds (F_{MSY}, B_{MSY}) for New England groundfish from empirically based stock recruitment models*- **Victor Crecco** (30 minutes)
- *A general biological reference point working group model*- **Andy Applegate** (20 minutes)

Facilitated discussion regarding presented materials in relation to terms of reference (all)

Wednesday, 5 February

0800-1700 Public Session – Topic: **Projections of Stock Rebuilding**

Background Documents:

-National Standard Guidelines for Overfishing Definitions: Final Rule
http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=1998_register&docid=fr01my98-23.pdf

- AgePro Users manual:

- GARM Report Revised Projections
http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0216/

- Other Contributed Documents

Order of the Day:

Formal Presentations:

- *NMFS National Standard Guidelines and Stock Rebuilding*- **Pamela Mace** (40 minutes)
- *Projection Methodologies used to evaluate medium-term impacts*- **Jon Brodziak** (40 minutes)
- *A phased rebuidling strategy, using the cod stocks from Gulf of Maine and Georges Bank as examples*- **Doug Butterworth** (30 minutes)
- *Rebuilding strategies for three key stocks*- **Dave Lincoln** (30 minutes)

Facilitated discussion regarding presented materials in relation to terms of reference (all)

Thursday, 5 February – Saturday, 8 February

Executive Session – Invited Peer Reviewers and support staff person (Karena Jolles, New Hampshire Fish and Game Department)

Discuss issues raised at public workshop and in supporting documents. Develop strategy for completing individual reports and how summarizer will convert them to a final document.

Commence the report drafting process individually and through debate.

Consult other participants for clarity purposes.

APPENDIX 4

BIBLIOGRAPHY

1. Formal Documentation (received before or at the meeting)

Almeida, F. and L. Jacobson. Working Paper: Species Compositions from the NMFS/Industry Survey Trawl Study Conducted by the R/V *Albatross IV* and F/V *Sea Breeze* 28 October-6 November, 2002. 24 pp.

Almeida, F., and L. Jacobson. Species Size Compositions from the NMFS/Industry Survey Trawl Study Conducted by the R/V *Albatross IV* and F/V *Sea Breeze*, 28 October - 6 November 2002.

Almeida, Frank. Working Paper: Comparison of R/V *Albatross IV* and F/V *Sea Breeze* Catch during the NMFS/Industry Survey Trawl Study. Presence vs. Absence by Species. 9 pp.

Almeida, Frank. Working Paper: Composition of the R/V *Albatross IV* 'Other Catch' Component during the NMFS/Industry Survey Trawl Study Conducted 28 October-6 November, 2002. 5 pp.

Almeida, Frank. Working Paper: Cruise Report of the NMFS/Industry Survey Trawl Study Conducted by the R/V *Albatross IV* and F/V *Sea Breeze*, 28 October-6 November, 2002. 6 pp.

Brodziak, J. K. T. and P. J. Rago. AGEPRO Version 2.02 User's Guide. July 23, 2002. 107 pp.

Brodziak, Jon. Comparison of Average Catch Rates of 20 Species for Optimal and Worst-Case Scenario Net Configurations by Area. January 14-15, 2003.

Butterworth, D. S., Rademeyer, R. A. and E. E. Planganyi. An Age-Structured Production Model Based Assessment and Reference Point Evaluation for the Gulf of Maine Cod Stock. 41 pp. (3 pp. Addendum added)

Butterworth, D. S., Rademeyer, R. A. and É. E. Plaganyi. Results for Georges Bank Cod of Age-Structured Production Model Based Assessments Similar to those Conducted for the Gulf of Maine Cod Stock. 22 pp.

Crecco, Victor. Overfishing Thresholds (F_{MSY}, B_{MSY}) for New England Groundfish from Empirically-Based Stock-Recruitment Models. January 26, 2003. 21 pp.

Fogarty, Michael J. Analysis of R/V Albatross IV - F/V *Sea Breeze* Trawl Configuration Experiment. 9 pp.

Lovgren, Jim. Observations from the *Albatross IV* correctional cruise. February 5 2003. 4 pp.

National Oceanic and Atmospheric Administration. 50 CFR Part 600 Magnuson Stevens Act Provisions; National Standard Guidelines; Final Rule. May 1, 1998. Federal Register 63(84): 24212-24237.

New England Fishery Management Council. Council Report. July 2002. 6 pp.

New England Fishery Management Council. Correspondence received by Council regarding the trawl gear survey information.

New England Fishery Management Council. Report of the Groundfish Overfishing Definition Committee. November 27, 2000. 12 pp.

Nies, Tom. Working Paper: Analysis of Catch-at-Length Data from the NMFS Industry Survey trawl Study Conducted by the R/V *Albatross IV* and F/V *Sea Breeze*. October 28 - November 6, 2002. 18 pp.

Northeast Fisheries Science Center Reference (NEFSC) Document 02-15. Report of the Workshop on Trawl Warp Effects on Fishing Gear Performance. October 2-3, 2002. 80 pp.

Northeast Fisheries Science Center (NEFSC) Document 02-16. Assessment of 20 Northeast Groundfish Stocks through 2001. A Report of the Groundfish Assessment Review Meeting (GARM) October 8-11, 2002. 511 pp.

Northeast Fisheries Science Center (NEFSC)/Industry Cooperative Survey Gear Study 28 October-6 November, 2002. Source Document: Specifications for Construction of NEFSC Standard #36 Bottom Trawl.

Northeast Fisheries Science Center (NEFSC), National Marine Fisheries Service. Final report of the Working Group on Re-Evaluation of Biological Reference Points for New England Groundfish. March 19, 2002. 232 pp. + 163 pp. of Appendix 7.0.

Northeast Regional Stock Assessment Review Committee (36th SARC). Draft Advisory Report on Stock Status. January 2003. 50 pp.

Overfishing Definition Review Panel. Final report: Evaluation of Existing Overfishing Definitions and Recommendations for New Overfishing Definitions to Comply with the Sustainable Fisheries Act. June 17, 1998. 179 pp.

Restrepo, V. R. et al. Technical guidance on the Use of Precautionary Approaches to implementing National Standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act. 1998. NOAA Technical Memorandum NMFS-F/SPO-31.

Starr, Paul. Memorandum: Analysis of NMFS Trawl Survey Data: R/V *Albatross IV* and F/V *Sea Breeze*. January 10, 2003. 16 pp.

Stauffer, Gary. NOAA Protocols for Groundfish Bottom Trawl Surveys of the Nation's Fishery Resources. December 16, 2002. 81 pp.

2. Presentation or illustrative material (received at the meeting)

Applegate, Andy. Handout: General Biological Reference Point Working Group Model.

Brodziak, Jon. Presentation: (Age-Structured) Projection Methodologies Used to Evaluate Medium-Term Impacts. February 5, 2003.

Brown, Russell W. Presentation: Issues with NOAA Fisheries Bottom Trawl Surveys Conducted.

Butterworth, Doug. Summary of Paul Starr's Analysis Presented to the trawl Experiment Workshop, January 14, 2003.

Butterworth, D. Rademeyer, R. and E. Plaganyi. Updated Projections covering phased rebuilding.

Chen, Yong. Presentation: Decision analyses using biological reference points in evaluating groundfish stock status. February 2, 2003.

Correspondence Received by Council Regarding the Trawl Gear Survey Information

Fogarty, Mike. Presentation: Effects of Trawl Warp Offsets and Gear Configuration on Survey Catches.

Goudey, Clifford A. Letter to Paul Howard (NEFMC). Comments on the significance of the warp offset issue and on the utility of the recent R/V *Albatross IV* and F/V *Sea*

Breeze comparison cruise in determining the possible sampling errors in recent trawl surveys. January 28, 2003.
Industry Stakeholder Concerns raised by those who participated in the September 25-27 experimental cruise, including a list of questions from fishermen. Handout.
Legault, Christopher M. Presentation: A Strategy to Evaluate Alternative Stock-Recruitment Models.
Lincoln, Dave. Presentation: Rebuilding Strategies vs. Catch.
Mace, Pamela M. Presentation: The implementation of National Standard 1 since the SFA. February 2003.
Mayo, Ralph. Presentation: Ecosystem Implications of Revised Biomass Targets.
Murawski, Steve. Presentation: Age-Specific Catchabilities Estimated for Four Stocks w/ ADAPT.
Murawski, Steve. Presentation: Reference Point Re-Estimation.
O'Malley, James. From Science to Illusion: Mathematics in Fishery Management. In *Pacem in Maribus* XXVI, Halifax, November 29-December 3, 1998.
O'Malley, James D. Letter to Mr Ricks Savage. East Coast Fisheries Federation, Inc. May 16, 2002.
Rago, Paul. Presentation: Intervention Analyses to Detect trawl Warp Offset Problems for NMFS R/V Survey Indices from 2000-2002. February 3, 2003.
Stevenson, Barbara. Handout: Trawl Data for R/V *Albatross IV* and F/V *Sea Breeze*.

3. Extra references consulted by members of the panel for the Report

Clark, W. 1991. Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries and Aquatic Sciences, 48: 734-750.
Clark, W. 1993. The effect of recruitment variability on the choice of a target level of spawning biomass per recruit. University of Alaska Sea Grant College Program, Report 93-02: 233-246.
Getz, W. M. and G. L. Swartzman. 1981. A probability transition model for yield estimation in fisheries with highly variable recruitment. Canadian Journal of Fisheries and aquatic Sciences, 38: 647-855
Kass, R. E. and A. E. Raftery 1995. Bayes factors. Journal of the American Statistical Association, 90: 773–795.
McAllister, M. K., Pikitch, E. K. and E. A. Babcock 2001. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1871-1890.
Mace, P. M. 1994. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Canadian Journal of Fisheries and Aquatic Sciences, 51: 110-122.
Mace, P. M. and M. P. Sissenwine. 1993. How much spawning per recruit is enough? *In* Risk Evaluation and Biological Reference Points for fisheries management. Ed. by S. J. Smith, J. J. Hunt and D. Rivard. Canadian Special Publication in Fisheries and Aquatic Sciences, 120: 101-118.
Meyer, R. and Millar, R. B. 1999. BUGS in Bayesian stock assessments. Canadian Journal of Fisheries and Aquatic Sciences, 56: 1078-1086.

Mohn, R. K. 1999. Retrospective patterns in… ICES Journal of Marine Science, 56: 473-488.

Myers, R. A., Bowen, K. G. and N. J. Barrowman 1999. Maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences, 56: 2404-2419.

Myers, R. A, Mertz, G. and P. S. Fowlow 1997. Maximum population growth rats and recovery times for Atlantic cod *Gadus morhua* Fishery Bulletin, 95: 762-772.

Parma, A. M. 2002 In search of robust harvest rules for Pacific halibut in the face of uncertain assessments and decadal changes in productivity. Bulletin of Marine Science 70: 423-453.

Shepherd, J. G. 1981. Cautious management of marine resources. Math. Biosci., 55: 179-187.

APPENDIX 5

CLOSING QUESTIONS POSED AND COMMENTS MADE TO THE PANEL BY PARTICIPANTS

Doug Butterworth

1. Have ADAPT assessments explored a sufficient set of sensitivities, for example in respect of alternative values of M, and what are the implications for reference point estimates?
2. Comment on the appropriateness of MSY-based management targets given the imprecision of the estimates and difficulties associated in particular with changes over time resulting from new data and changed methodologies.
3. Given the ASPM-based reference points for two cod stocks, albeit based on initial analysis, are appreciably different from those based on ADAPT methodology, how important is it that further ASPM-based results be developed to be taken into account in the next set of management decisions for these stocks?
4. Given that assessment method, with current data, appear unable to estimate parameters such as stock-recruitment steepness (and hence B_{MSY}) with great precision, what is the potential role of adaptive management towards improving such precision? (Note the relevance of this question in respect of scientific aspects of the TOR 3C.)
5. Is it important for the *Albatross* to survey efficiently (as well as comparably over time) to be able to use associated swept-area estimates of absolute abundance to "ground-truth" estimates provided by population model assessment methods?

Geoffrey Smith

1. Given the fact that B_{MSY} values are generally set at one half of the carrying capacity of the stock, is it unreasonable to assume that all 19 stocks of groundfish can be rebuilt to B_{MSY} simultaneously?
2. Do rebuilding strategies that allow continued over fishing in the near term pose a greater biological risk than those that reduce fishing mortality rates to levels at or below F_{MSY}?
3. Is the question of the National Standard Guidelines requirement to rebuild overfished stocks to B_{MSY} in 10 years or less a scientific question or a legal and/or public policy question?

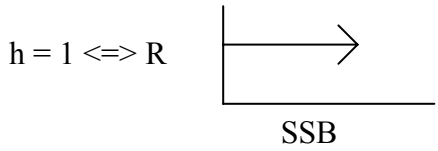
Priscilla Brooks

1. Fisheries management in the United States is governed by the Sustainable Fisheries Act and National Standards that dictate narrowly the parameters with in which management plans are developed. Pamela Mace's presentation gave you a fairly thorough overview of the law and guidelines. I believe that you must keep in mind the legal reality in which we work and in which your report will be incorporated B_{MSY} and F_{MSY} must be estimated and stocks must be rebuilt within 10 years, except

in circumstances in which the natural history of the stock dictates more time. Given these realities, is the NMFS science related to the biological reference points, that is the GARM report, sound?

Jon Brodziak

1. Are the steepness parameter (h) values implied/estimated in the Butterworth production models for GOM cod credible, in the context of Myers *et al*. (1999. Maximum reproductive rate...CJFAS)?



Ron Smolowitz

1. What is the sensitivity of the trawl survey to towing speed changes over time?

Eric Smith

1. Perhaps a useful follow-on question is to ask "Is there justification, given scientific uncertainty in biological reference points and projection methodology, for setting a lower intermediate 10-year rebuilding target that can be adjusted upwards as the stock builds and our estimate of that value becomes more certain?" This better captures the essence of the Council's question/concern from a management standpoint. TOR 3C
2. Is a Ricker-type S-R curve more (or equally) justified relative to a B-H type curve for cod and haddock? TOR 2A, bullet #2

Phil Ruhle

Please look over NMFS protocol for groundfish surveys, recently developed.

1. The speed issue is of great concern but the gear used is also a problem. In all other surveys gear is well addressed but NEFSC survey net design and age is 40 years. The design has not been used by industry in 20 years.
2. Bottom contact on this gear is very lax as is all aspects of handling of this gear; this is shown in NEFSC protocol as compared to other science centers.

Pamela Mace

Note about the Precautionary Approach:

1. See page 11 of Technical Guidance for a statement about how the precautionary approach is appropriate to management decisions, but not to scientific estimation of assessment-related parameters and variables.

Andy Applegate

1. Which other analytical methods can be used to validate the reference point estimates and rebuilding projections given the heavy reliance on less robust and variable recruitment estimates? How do managers use the scientific advice while this effort is made?
2. Are there better methods within the context of the current National Standards to evaluate the performance of the plan and monitor rebuilding of a set of multispecies fisheries?

Tom Nies

Question on Trawl Experiment:

1. Was the design of the experiment adequate to determine if errors in the trawl warp cable affected recent survey results?
   Points to consider:
   -The control net differed from the design of the survey net used for the past two years (ignoring the issue of warp length). Some differences: different doors, use of swivels on doors, different backstraps, different ground cable rigging.
   -Experimental tows were all conducted either into or with the current. Survey tows are towed in the direction of the next station, without regard to current. The experiment never towed cross-current, and a poorly rigged net may tow differently in a cross-current.
   -If, as suggested by Paul Rago based on Pennington's work, the effective sample size for frequency distributions is closely related to number of tows (as opposed to fish caught), were there enough tows to draw conclusions on catch at length/age?
   -Is the assumption of a covariate relationship between the commercial vessel and Albatross catches justified by the analysis? (I have not seen the final paper by Dr Fogarty).
   -For the covariate analysis, how does the insertion of values for missing *Sea Breeze* catches affect the results?

Jim O'Malley

1. Is there evidence of any application of the precautionary principle in the assessments or rebuilding targets?
2. Is such an application legitimate in science?

David Frulla

1. If one manages towards B_{MSY} for every species in a mixed stock fishery at the same time, can this result in under utilisation of certain species? TOR 3
2. Can differing F reduction strategies accommodate considerations relating to a mixed stock fishery, economic consideration, and uncertainties related to significantly

higher new reference points, while achieving the appropriate biomass target over the relevant rebuilding period? TOR 3