New England Fishery Management Council

Scientific and Statistical Committee Sub-Panel

Peer Review Report for the Groundfish Plan Development Team Analyses of Groundfish Monitoring

Conducted April 24 and 25, 2019 in Providence, Rhode Island

Prepared by the Peer Review Panel: Dr. Dan Holland[1], Dr. Lisa Kerr[2], Dr. Jason McNamee[3] (chair), and Dr. Hiro Uchida[4]

[1] NOAA, Northwest Fisheries Science Center

[2] Gulf of Maine Research Institute

[3] Rhode Island DEM, Division of Marine Fisheries

[4] University of Rhode Island

Executive Summary

A sub-panel of the Scientific and Statistical Committee ("the peer review panel") was convened on April 24 and 25, 2019 to review four analyses along with a conclusions statement conducted by the New England Fishery Management Council's (NEFMC) Groundfish Plan Development Team (PDT). The analyses were conducted to look at the potential effects on harvest and discards in the Northeast Multispecies (here after "groundfish") fishery when an at-sea observer is present on a fishing trip relative to when no observer is present. This is important because sector program accountability and estimation of discards is largely derived from observed trips, so if there are differences in fishing behavior on those trips, this could impact the effectiveness of management of this fishery.

The first analysis modeled incentives to discard in the groundfish fishery. The analysis is done at the trip level, but results are presented on an annual species/stock level. The analysis shows that, while on average there are no positive incentives to discard for any species across years, there are positive discard incentives for a proportion of trips for some species/stocks, notably for cod and yellowtail flounder stocks in certain years. There are reasons to believe that the estimated discard incentives are conservative such that even when the estimated incentive is not above zero, there may still be incentives to discard that are not captured by the model. The analysis was not able to estimate the frequency of trips or the magnitude of catch that may be subject to positive discard incentives and thus cannot quantify the magnitude of the problem. Rather it provides an indicator of where and when discarding may have been incentivized, and therefore indicates that discarding in this fishery is economically incentivized in some instances. This could help managers focus efforts on to the areas of the fishery where the main problems likely exist.

The second analysis examined whether fishing vessels in the groundfish fishery alter their behavior in response to human observers. The measures examined in this analysis cover a broad range of impacts that are relevant for observer-related fisheries management policy. The analysis found statistically significant differences in many measures (but not all) between unobserved and observed fishing trips of the same vessels, strongly suggesting that fishers do alter their fishing behavior when a human observer is onboard. However, since a key difference is shorter duration of unobserved trips, this may explain at least part of the differences in other variables such as kept catch.

The third analysis used observed trips to model cod (and pollock) catch while accounting for typical effort attributes in addition to spatial and temporal covariance in catch. The approach creates a predictive model, which was used to predict total cod catch (kept + discarded) on observed trips, and then was also used to predict catch for unobserved trips, which were then compared to catch reports from the National Marine Fisheries Service (NMFS). The method indicated differences between the predictions of unobserved catch and the catch as reported by NMFS. This finding suggests a potential for unreported catch on the unobserved trips if it is assumed that observed trips can adequately represent unobserved trips. This method has an ability to predict unobserved discards controlling for differences in spatial distribution of unobserved trips relative to observed trips and showed promise for informing the Council

quantitatively in their deliberations on Amendment 23 with some additional refinement and testing.

The final analysis compared species/stock landings to effort and total catch ratios on observed and unobserved trips in the multispecies groundfish fishery to determine whether there is evidence of an observer effect. The analysis assumes that differences are due to the observer effect and are not due to the deployment effect, so this is an important consideration when interpreting results from this method. The reviewers appreciated the parsimony of this approach, but felt it needed a little more refinement before it could be used by managers. The method did indicate the potential for differences between observed and unobserved trips, and therefore it corroborated the other results from the other analyses.

Generally, the reviewers appreciated all of the work done by the PDT and felt that the analyses, taken comprehensively, create a weight of evidence that disproves the null hypothesis, namely that there is no effect from the presence of an observer on a fishing trip. In other words, the work taken collectively show that there is an observer effect, and therefore managers need to account for this when basing management off information derived from observed trips. The analyses suggest that estimates of discards on unobserved trips derived from discards rates on observe trips may not be accurate, and likely to be an underestimated reflection of actual discards. In their current form the analyses do not offer a specific quantification of the problem, but the methods show promise for being able to focus efforts on to the most problematic species in the fishery, and some of them also show promise for being able to quantify the magnitude of the issue with additional work. The following report details these findings.

Table of Contents

Report

I. Methods to explore discard incentives of groundfish stocks

a. Reviewer Summary Comments:

This paper models' incentives to discard catch of groundfish stocks based on the estimated economic incentives to retain or discard the catch. The analysis is done at the trip level and, for each species, subtracts the benefits of retaining catch (mainly ex-vessel value) from the cost of retaining catch (mainly quota leasing costs) to estimate the incentive to discard. The discard incentive measure is standardized by the ex-vessel value of catch so might be thought of as the monetary incentive to discard per dollar value of retained catch. Results are presented on an annual species/stock level. The analysis shows that, while on average there are not positive incentives to discard for any stock or year, there are positive discard incentives for a proportion of trips for some species/stocks, notably for cod and yellowtail flounder stocks in certain years. There are reasons to believe that the estimated discard incentives are conservative such that even when the estimated incentive is not above zero, there may still be incentives to discard. The presentation of methods and results is clear, and the methodology and assumptions made are generally sound as are the conclusions drawn from the analysis. The analysis is applied only to unobserved trips but could and should be applied to observed trips. The analysis is not able to estimate the frequency of trips or magnitude of catch that may be subject to positive discard incentives and thus cannot quantify the magnitude of the problem. Rather it provides an indicator of where discarding may have been incentivized.

b. Terms of Reference:
    1. Are the methods adequately described and based on sound analytic techniques and statistical principles?

The methods for this paper are generally well described and based on sound theoretical and analytic techniques. As with most economic models that represent human decision making as a purely economic and rational process, there are several assumptions made in the model that may not strictly hold. In addition, the model does not represent the potential heterogeneity of decision makers that may alter what costs and benefits of landing fish are included in decision making. However, the assumptions made are generally reasonable and clearly stated, and their rationale is described.

The paper should be clearer about the type of decision maker that is represented. The decision of a vessel owner might be different from a hired captain based on differences in costs and benefits considered. For example, a hired captain or crew member on a boat that does not deduct the cost of quota when determining crew share might not consider the quota cost and thus would have a lower incentive to discard than is estimated here. Alternatively, a captain who is also the owner of the vessel and the quota might consider the net value of the fish after both crew share and quota costs are deducted and thus have a higher incentive to discard than is estimated. It appears the model best represents the viewpoint of a hired captain or crew member that expects quota costs to be deducted off the top and their compensation to reflect a share of the difference between ex-vessel value and quota cost. To the extent it is known, it would be useful to discuss

how prevalent this type of decision-making is, e.g., how prevalent is it to deduct the value of quota from ex-vessel value before calculating crew share for hired captains and crew?

Other background that is not made clear in the paper is that observed trips were excluded from the analysis. The analyst stated in post presentation discussions that only data from unobserved trips was used, and this should be clarified in the paper. It is not clear why observed trips should be excluded as these trips seem likely to be a substantial source of data on trips where discard incentives existed for legal size fish but where those fish were retained due to the presence of the observer.

2. Are important uncertainties in the data and the analyses (possibly including the effects of year to year variations in fishing practices) identified, and are the impacts of these uncertainties on the analyses adequately described?

There are many uncertainties in the data and analysis, but they are generally identified in the paper and the impacts of these uncertainties are discussed. The most important uncertainty is the accuracy of quota prices that are a key factor in the decision to land or discard fish. The analysts noted in discussions after the presentation that it is the magnitude of the margin between ex-vessel price and quota price that is the primary determinant of the estimated discard incentive. Quota prices are based on reported data on quota leases (i.e., annual catch entitlement (ACE) sales) within and between sectors and provide the best representation available of quota value; however, they undoubtedly depart somewhat from the perceived opportunity cost of quota in the mind of the decision maker deciding whether to discard or retain fish. Quota prices are mostly based on quarterly estimates but there may be variations in market values within quarters.

More fundamentally, there may be a variety of reasons for decision makers to use something other than the market price of quota in making their determination of whether to discard or land fish. A basic assumption of the model is that there are known market prices for quota, and that the decision maker uses these prices to estimate the cost of landing fish. The analysis assumes away any impact that the quota holdings of the vessel at the time of landing, or the expected need for future quota, might have on the discard decision. It basically assumes the quota can be bought for that price upon landing the fish and that the same quota price will continue to prevail. This is a reasonable assumption if the quota market is efficient, information on quota prices is readily available, transactions costs are low, and decision makers are rational. However, the assumed value of quota to the individual may in fact be affected by quota endowments and expectations of whether vessel owners/decision makers will or will not need to seek quota on the market at some point and what prices may be at that time. For decision makers that expect to have to lease quota to cover catch of otherwise discarded fish, there may be a risk premium applied to account for the risk of not being able to acquire quota in a timely fashion. This would increase the discard incentive. There might be pressure from vessel/quota owners/holders to limit the landings of stocks for which quota is limiting. On the other hand, a vessel owner with extensive quota holdings might be less concerned about quota prices and might assume a lower quota value in the discard decision.

As noted in the paper, a positive discard incentive does not necessarily mean discarding will occur. There are many reasons including moral values and social norms that might motivate landing of fish when discard incentives are positive. Furthermore, sectors may have implemented programs that change incentives, either by increasing the likelihood of detection and penalty for illegal discarding, or by changing incentives. An example of the latter would be if the sector maintains a pool of quota available to cover catch that can be purchased with certainty at a price lower than ex-vessel price or the prevailing market price of quota.

Overall the uncertainties and caveats of the analysis are well described, and the reader can draw their own conclusions. It seems likely the analysis is conservative in estimating the discard incentive. A key factor that may make the estimates of the discard incentives conservative is risk. If a captain (or decision maker) is uncertain they will be able to acquire quota to cover the fish they would otherwise discard and may be forced to stop fishing, they may add a risk premium to the cost of quota considered in Equation 2 from the document (see page 3), which will tend to increase the discard incentive. Because the estimates of discard incentives may have a conservative bias, it may be that discard incentives did exist in some cases where the estimated discard incentive is less than zero.

3. Are the analyses conducted at the appropriate temporal and spatial scale such that the existence of regional or seasonal differences in monitoring performance can be identified?

To the extent possible, the analysis is done at an appropriate temporal and spatial scale. The primary limiting factor is the data used to estimate quota values. These are generally quarterly estimates but sometimes annual estimates. However, the prices or expected prices could vary within quarters, by sector or by individual. This creates uncertainties in the analysis, but the decisions of the analysts appear appropriate.

It might be of interest to disaggregate the analysis and presentation of results to the sector level or some other grouping of vessels for which we might expect heterogeneity in incentives or behavior in relation to incentives. For example, different ports may have different ex-vessel prices. Some sectors may have different internal quota prices or a pool of quota that can be accessed at a known price. It might also be useful to present results on a quarterly basis. This might better identify times and stocks where discard incentives were positive.

4. What are the strengths and weaknesses of the methods? Are there constraints that would hinder the use of the catch monitoring analyses?

As the PDT notes in their own summary conclusions: "The model when it is applied to unobserved trips can only estimate a discard incentive when landings or trips comply with the discarding prohibition, even when it may not be economically rational to do so." The method is useful in showing that incentives to discarding did likely exist in certain circumstances. It is, however, unable to quantify the frequency or magnitude of actual discards. The analysis almost certainly underestimates the percentage of trips where discard incentives existed to the extent that there are trips where catch was discarded and thus was not included in the analysis. We

know, as the paper notes, that fish were landed in compliance with regulations despite a positive incentive to discard it. What we don't know is how often catch was discarded.

A simpler analysis might also be informative. Simply comparing ex-vessel prices (by grade and perhaps port) to quota lease prices would probably be almost as good an indicator of incentives to discard. It would not indicate the percentage of trips or catch for which incentives exist, but the current method underestimates this anyway since it does not apply the estimate to discarded catch. It would be possible, using this suggested method, or the method used in the analysis reviewed, to estimate the percentage of catch from observed trips where a discard incentive existed. This would be a useful exercise though it may also underestimate the percentage of catch subject to positive discard incentives to the extent that behavior on observed trips is different and leads to lower catch of fish with positive discard incentives.

5. Are the conclusions of the Plan Development Team supported by the analyses (see 1e)?

The conclusions of the PDT related to this analysis are reasonable and capture most of the key conclusions and caveats associated with the analysis. None of the conclusions of the PDT seem inappropriate, and the limitations of the analysis are acknowledged. The PDT notes that a positive discard incentive may indicate a bias in landings data, but they do not state that the analysis definitively proves the existence of substantial discards. They note that the analysis cannot quantify the proportion of trips where discards did occur or the amount of discards. As the PDT notes, more precise estimates of quota prices would enhance the ability of the model to identify where discard incentives existed.

6. Are there recommendations for improving the analyses, or for additional research or data collection that can help address improving groundfish monitoring?

A key to this analysis and a limiting factor is good approximations of the opportunity cost of quota that was used in making the decision to land or discard the fish. As noted by the PDT, better or more complete quota lease price data would help this. However, even with better estimates of market value of quota, the value assumed by the decision maker may depart from market value for reasons described above. This is probably an unresolvable problem.

The primary determinant of the discard incentive is the margin (or lack of margin) between ex-vessel price and quota price. It would be useful to include a chart or table comparing quarterly average ex-vessel price for various grades or the lowest grade and quarterly average quota lease price for stocks where discard incentives were relatively higher (though perhaps still negative on average). It appears that in most cases quota prices do not exceed average ex-vessel value, but the question remains as to whether this is the case for the lowest market grade. The table or chart suggested would provide an indication of where incentives for high grading exist, namely where higher grades of the species might be profitable to retain while lower grades of fish are not.

The analysis did not include data on observed trips. It would be useful to analyze these trips and compare the results with those from the unobserved trips already analyzed.

It might be useful to more carefully analyze trips where incentives to discard were positive or closer to zero to determine what drove these higher incentives in relation to the average incentive for that period and stock. For example, we might find that those trips tended to land a substantial amount of small or lower grade fish, or perhaps unmarketable fish. This would indicate a high grading issue rather than an incentive to discard all catch of that species. It might also identify heterogeneity across groups (e.g. ports with lower ex-vessel prices that have higher discard incentives).

Editorial comments for authors:

Throughout this paper and some of the other documents reviewed, authors refer to quota and quota price instead of ACE and ACE prices. If this is a standard convention in discussions in the Council arena, this may be appropriate. However, referring to ACE and ACE sales or trades would be more accurate than referring to quota and quota lease.

Pg. 1 paragraph 2: It is not clear what the authors mean by "we lack the terminology to communicate precisely what we are estimating." Please clarify this.

Pg. 2. 2nd paragraph, 2nd sentence. There is not necessarily an incentive to discard in all quota-based fisheries. It depends whether quota is limiting and how limiting it is and the cost of avoidance.

Pg. 2 last sentence of paragraph 3: Those less able or willing to discard may gain from a lower quota price. Quota owners lose from lower quota value to the extent that they lease rather than use quota and to the extent that discarding undermines the long term value of the quota.

Pg. 8, 1st full sentence. The value of forgone future catches of jointly caught stocks should be captured in the quota price unless quota price is being kept down by discarding or other factors (e.g. unwillingness to "gouge" in lease transactions as found by Holland (2013)).

II. Methods to evaluate observer effects in the groundfish fishery

a. Reviewer Summary Comments:

This analysis demonstrates that fishing vessels in the groundfish fishery alter their behavior in response to human observers. The analysis looked at eight measures: namely (1) trip duration, (2) kept catch, (3) kept groundfish, (4) kept non-groundfish, (5) total revenue, (6) groundfish average price, (7) opportunity cost of quota, and (8) number of groundfish market categories included in kept catch. These measures cover a broad range of impacts that are relevant for observer-related fisheries management policy. The analyses were conducted separately for four stanzas (one pre-sector stanza and three post-sector stanzas) and also by fishing gear (gillnet and trawl). Additionally, the approach was unique in how it chose its sequence of trips for analysis, which the reviewers appreciated. The analyses found statistically significant differences in many measures (but not all) between non-observed and observed fishing trips of the same vessels, strongly suggesting that fishers do alter their fishing behavior when a human observer is onboard.

b. Terms of Reference:

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

Yes. The core method used to construct the comparison pairs of non-observer and observer trips followed the procedure of Benoit and Allard (2009) and was explained well.

The only suggestion the reviewers have is to clearly state that the comparisons were made also with the same vessels. This is clear in Equation 3 from the document (see page 3), however only if one pays close attention to the use of subscripts. The term "exact matching" is used in the matching method literature and does not necessarily point to same-entity matching. Use of the same vessel for constructing the sequence of UUU and UOU trips as well as conducting the comparisons is the key feature of this study and what makes the analysis so clean, and thus deserves to be emphasized.

2. Are important uncertainties in the data and the analyses (possibly including the effects of year to year variations in fishing practices) identified, and are the impacts of these uncertainties on the analyses adequately described?

Yes. The use of four stanzas and constructing the UUU/UOU sequences that are within 45 days apart minimizes the temporal unobservable impacts on the analyses, such as seasonal and other environmental effects. Depending on the time of year, the 45-day decision is an important uncertainty in the analysis, as this could be a biologically meaningful period of time at the change of a season (for instance). However, the analyst made this decision as a trade-off between minimizing the amount of time while keeping enough data to analyze, which is justified. The use of the same vessels for the trip sequence and comparison controls for unobservable fixed effects unique to a vessel.

3. Are the analyses conducted at the appropriate temporal and spatial scale such that the existence of regional or seasonal differences in monitoring performance can be identified?

Yes, as described above.

4. What are the strengths and weaknesses of the methods? Are there constraints that would hinder the use of the catch monitoring analyses?

The strengths of the method used are as described above. Not as weaknesses but more as suggestions to make the results more robust, the reviewers have following suggestions.

For each of the eight measures, state the expected direction(s) of change or impacts of the human observer being present. The paper should explain not just that some behavioral changes were detected, but also whether those changes make sense. Some measures may be ambiguous, i.e., the impacts can go both ways, which is fine and makes a stronger case for having the "two-sided problem" section in the document. Also, along this line of thought, it would be helpful to link the five reasons for behavioral change outlined on p.2 to the eight measures.

Related to the comment above, it is strongly recommended that some of the measures be standardized by the trip duration. For example, a "kept catch" measure might be expected to go up, ceteris paribus, with an observer because there will be no illegal discarding but could also go down if vessels avoid areas with higher target catch but also higher catch of unwanted species. The results show, for trawls and gillnets, that kept catch was lower on observed trips. This may be due, at least in part, to the decline in trip duration. Revenue on observed trips also was lower than unobserved trips, which is in line with a priori expectation, but it also could be an artifact of shorter trip duration. It was discussed during the presentation that if all detected differences are stemming from shortened trip duration then there is no behavioral change that regulators need to be concerned about. Economic intuition tells us that is highly unlikely (i.e., there are likely to be behavioral changes in discarding and fishing location choice), but to make that point clearer and stronger these measures need to be standardized in a meaningful way. This may be partially accomplished by looking at changes in kept catch per day, but if catch increases nonlinearly with duration (e.g., because steam and search time are a lower proportion of trip length) we might expect kept catch to decline more than proportionately with duration. Alternatively, vessel monitoring system (VMS) data may be utilized to distinguish the tow/fishing time from steaming or searching time and it could be used to standardize the measures.

5. Are the conclusions of the Plan Development Team supported by the analyses (see 1e)?

Yes, but they will be strengthened by standardizing the measures as described above.

6. Are there recommendations for improving the analyses, or for additional research or data collection that can help address improving groundfish monitoring?

See item #4 above about standardizing the measures. Additionally, for clarification purposes, it will be helpful to define "opportunity cost of quota" (or simply linking the reason #5 described on p.2 of the document), along with how this was calculated.

Conjectures (or anecdotes) on the differences between the trawl and gillnet in the context of pre-harvest behavioral manipulation should be explicitly explained for the two following reasons. One is it motivates why the analysis should be conducted separately between these two fishing gear types. Another is it may be the core factor of why some results, i.e., behavioral responses to observed trips, are different.

The reviewers were told that more than the presented eight measures were analyzed but some were excluded from the report primarily because no statistically significant differences were detected. Some of them, however, would merit inclusion. Herfindahl-Hirschman Index (HHI) is one example; HHI is a measure of concentration and in this context can be used to measure whether the variety of fish species landed is more "concentrated" in handful of species. Intuitively, we think fishers will attempt to alter the composition of fish caught (e.g., less bycatch) through a strategic choice of fishing location in response to an observer on board. HHI is a measure that can detect such behavioral change.

Lastly, during the presentation the authors explained why they chose the max days between the fishing trips within a sequence (UUU or UOU) to be 45 days, and how they tested for both

shorter and longer durations. The reviewers feel this discussion merits inclusion as it would more strongly justify the choice of 45-days.

III. Methods to predict groundfish catch in the presence of observer bias

a. Reviewer Summary Comments:

This method used observed trips in the Gulf of Maine (GOM) stock area to model expected cod catch while accounting for typical effort attributes (e.g., total kept catch, vessel size, trip length) in addition to spatial and temporal covariance in catch. The approach creates a predictive model, which was used to predict total cod catch (kept + discarded) on observed trips, to test the performance of the model. The predictive model was then used to predict catch for unobserved trips. Both predictions were compared to the summed predictions across a fishing season to the catch estimates for sectors reported by NMFS.

The method did a fair job of predicting catch for the observed trips, which is the dataset that the model was developed on, showing that the modeling approach has value and predictive power for the data used. The method also indicated discrepancies for the prediction of unobserved catch relative to the catch as reported by NMFS. This finding suggests a potential for unreported catch on the unobserved trips if it is assumed that observed trips can adequately represent unobserved trips with regard to "pre-catch" behavior, meaning a fisherman will operate in the same manner prior to catching fish on a fishing trip whether an observer is present or not. Pre-catch behavior was also modeled using several attributes of effort (e.g., kept all, location) that were expected and shown to influence catch outcomes.

Due to its current configuration, the method is likely conservative in its predictions (i.e. will produce lower expected catches than is true), because it is not proven that "pre-catch" behavior is similar between trips that are observed and unobserved. In particular for the case of otter trawl, even with short notice of a trip being observed, a fisherman could alter behavior more easily than a gillnet operation, which likely already has gear set before leaving the dock. There are still modifications that a gillnet operation could make, like shortening trips and only hauling certain strings of gear that they believe will be less likely to have unwanted species, but there are less options for modification for gillnetters relative to otter trawl.

The review panel offered suggestions on ways to improve the model. Generally, these had to do with the addition of new explanatory variables (potentially some of the information generated for the other methods examined during this review), additional data sources (vessels with cameras, VMS data, Industry-Based Survey (IBS) cod survey), and potential ways to include trips that did not land cod.

This method has an ability to predict unobserved discards controlling for differences in spatial distribution of unobserved trips relative to observed trips. The approach was novel and interesting, and showed some promise for informing the Council quantitatively in their

deliberations on Amendment 23 with some additional refinement and testing. For these reasons the review panel believes additional effort should be invested in this approach.

b. Terms of Reference:
    1. Are the methods adequately described and based on sound analytic techniques and statistical principles?

Yes. The approach uses well defined statistical modeling principles, the documentation of the equations and underlying assumptions was thorough, and the model choices (e.g. error distribution) and other underlying theoretical aspects of the model were well justified for the hypothesis being tested.

Spatial and temporal covariance are two of the more interesting aspects of this approach. This was an area that was highlighted as worthy of additional investigation as far as how these aspects were parameterized, however even in their current state, these aspects of the model provide valuable insight into the issue of potential things that could influence observer bias and are based on sound principles and techniques.

There is one potentially important problem with the model specification that could cause bias and inconsistency in model predictions. The explanatory variables include a variable for all kept catch (namely, "kept_all"). Part of this catch is cod, which is the dependent variable for the model. This is likely to lead to correlation between the residuals in the model and the variable kept_all. For observations where cod is a large proportion of the total catch, the residual is likely to be increasingly negative (the model will likely be underpredicting the cod catch). This may account for the fact that the model, when applied to observed trips, tended to underpredict catch.

    2. Are important uncertainties in the data and the analyses (possibly including the effects of year to year variations in fishing practices) identified, and are the impacts of these uncertainties on the analyses adequately described?

Yes. As highlighted by the analyst, other statistical distributions for expected catch on a trip (e.g., quasi-Poisson, negative binomial) could have been used. These alternate distributions may provide a better fit to the catch data; however, the analyst made a good case as to why he used the selected distributional assumption, namely that the random error term was useful at capturing over-dispersion in the predictions. This could be an area of further exploration.

As noted, the most unique aspect of the analysis was the temporal and spatial aspects of the model. The scale chosen for the spatial and temporal knots were not explored beyond those presented. Other choices should be investigated with a focus on finding the best trade-off between more refinement in the number of "knots" but not decreasing sample size to the point of adding an unreasonable amount of uncertainty in to this part of the analysis. This portion of the analysis also highlighted the limitations of using VTR data only as the data source in particular for the spatial information. VMS data might offer a refinement of this aspect of the data.

The reduction in effort and observer coverage across time was also an area highlighted as an uncertainty. The reviewers suggested adding more years of information to the model and the analyst noted that this would leverage parameter pooling across years and might yield more

accurate parameter estimation, the current configuration only used a small subset of years. A final note was that this model makes inferences from observed trips on to unobserved trips. This is important with respect to testing the hypothesis that there are likely differences between these two treatments, however this is an important uncertainty as you might never be able to know the full extent of the issue through this process. The reviewers offered additional analyses and modifications that might help address some of these uncertainties and are presented in this section (see #6 below).

3. Are the analyses conducted at the appropriate temporal and spatial scale such that the existence of regional or seasonal differences in monitoring performance can be identified?

Yes. In particular, the temporal scale of two weeks is likely adequate to capture things like seasonality in species abundance. The spatial resolution should be investigated to see if there could be more refinement, however there is a balance between refinement of the scale and loss of information in the data, so this should be investigated further.

4. What are the strengths and weaknesses of the methods? Are there constraints that would hinder the use of the catch monitoring analyses?

Strengths of the analysis include: it is predictive; it quantifies the potential difference between observed and unobserved trips; it is based on sound principles and statistical techniques; it accounts for heterogeneity in catch due to differences in space and time; it is able to account for biases of underlying data sources; and, as currently configured, it does not need additional or unique data sources.

Weaknesses of the analysis include: it was developed mainly for cod at this point; it is limited in some aspects by available data; it is computationally intensive; there is variability in the predictions for otter trawl that need further investigation; its use as a predictive tool will be bolstered by more data informing potential changes in "pre-catch" behavior and more refinement in spatial information; and more work on justifying the co-occurrence species is needed.

The model may be conservative in predicting cod catch on unobserved trips if fishermen were generally attempting to avoid cod catch on observed trips used to parameterize the model. This is partially controlled for with the spatial correlation terms, but fishermen could have employed other methods or information to avoid cod catch on observed trips.

5. Are the conclusions of the Plan Development Team supported by the analyses (see 1e)?

Yes, all of the comments about this method made by the PDT are noted in the documentation of the analysis or in this review of the method. The one caveat to this statement is with regard to the last statement made on the analysis for pollock by the PDT. This was done to show robustness of the method; however, the reviewers don't believe that as much time was invested in this analysis, nor should there be an expectation that the same covariates that are useful for cod are useful for pollock. With a more species-specific model, the power of this method for use on other species should improve.

6. Are there recommendations for improving the analyses, or for additional research or data collection that can help address improving groundfish monitoring?

The reviewers made several recommendations for improvement to the model. These include eliminating cod from the right-hand side of the model either by replacing kept_all with a new parameter comprised of kept_all minus cod catch or by eliminating kept_all and adding additional species catch variables but leaving cod out. Investigating new and/or additional covariates to help better inform the model could also improve the model performance. The reviewers suggested trying covariates that help in the prediction based on changes in population size such as using acceptable biological catch (ABC), or spawning stock biomass in the model as available or appropriate. Additionally, the reviewers thought that adding in some of the information generated by the other methods reviewed, such as using the discard incentive work might also be helpful. The reviewers suggested running the analysis across multiple years and to better refine the species selected as appropriate covariates in the model coupled with potentially finding ways to use these co-occurring species to bring in tows with no cod catch and develop an index such as employed in the development of a Jaccard Index (Jaccard 1901). The reviewers suggested using other data sources beyond observer data such as the Massachusetts IBS for cod and/or using data from vessels with cameras. Some other factors noted by the reviewers were that model validation was needed. The method was validated with respect to its predictive capacity, but should also be tested against an independent dataset, such as a fishery independent survey dataset). The reviewers also suggested running the model without discarded catch, and finally that the description of the results should try to better characterize the most likely outcome and uncertainty around that most likely outcome as many will likely be unfamiliar with the Bayesian approach to describing solutions and results.

IV. Methods to evaluate groundfish catch ratios

a. Reviewer Summary Comments:

The objective of the study was to compare ratios of stock-specific landings to effort and total catch on observed and unobserved trips in the multispecies groundfish fishery to determine whether there is evidence of an observer effect. The hypothesis of the study was that if constraining stocks lead to illegal discards, this should be evident in differences in the stock-specific ratios of landings to effort and total catch between observed and unobserved trips. The study assumes that differences are due to the observer effect (i.e., observed trips do not represent unobserved trips) and not due to the deployment effect (i.e., observers are not randomly distributed among fishing trips). Landings ratios were characterized at an aggregate level by gear type and broad stock area over an annual time step for both observed and unobserved trips. The reviewers appreciated the parsimony of this approach, but felt it needed a little more refinement as described below before it could be used by managers.

b. Terms of Reference:
    1. Are the methods adequately described and based on sound analytic techniques and statistical principles?

Yes, the methods for this study are relatively straightforward and clearly described, although some aspects of the study would benefit from additional description.

The reviewers suggest that the report more clearly explain that the calculation of ratios of landing to total catch or effort ratios was conducted at the aggregate gear-area level rather than individual trip level. For example, the equation should include subscripts for species, area, and gear type, and a subscript on the sum function indicating summing across trips. Additionally, the description of the data should be clarified, and the "AA table" should be specifically defined.

While there is no statistical analysis of the differences in landings ratios, color coding is used to identify patterns that are consistent across landings ratios. Further description of criteria for what constitutes the identification of a pattern should be included as this plays an important role in interpretation of the results. The analyst should describe the rationale for assigning results using the yellow and gray color assignment and make clear that these do not indicate statistical significance.

2. Are important uncertainties in the data and the analyses (possibly including the effects of year to year variations in fishing practices) identified, and are the impacts of these uncertainties on the analyses adequately described?

The uncertainties in the data and analyses are not discussed extensively in the report, although it was noted that Sector IX trips were excluded from the analysis because they are known to be subject to misreporting. A key uncertainty not addressed in the study was the large differences in the sample size of observed to unobserved datasets, which introduces potential issues for the validity of the comparison of landings ratios. A more detailed description of the breakdown of data by finer spatial resolution and time step (season or month) would allow a more thorough evaluation of whether the observed samples were representative and comparable to unobserved samples. However, a more statistically robust treatment of the data is recommended (further detail is provided below).

3. Are the analyses conducted at the appropriate temporal and spatial scale such that the existence of regional or seasonal differences in monitoring performance can be identified?

The analyses appear to be conducted at the appropriate temporal and spatial scale. The analyses are conducted at the broad stock area (Gulf of Maine, Georges Bank, mixed stock statistical area 521) over the years 2011-2017. This allows for identification of regional and annual differences in observer effect but does not allow for seasonal differences to be identified. A more detailed description of the spatial and temporal distribution of the data would allow a comprehensive evaluation of whether the analysis could be conducted at a finer spatial or temporal scale. The limited number of observed trips included in the analysis suggests this may not be possible.

4. What are the strengths and weaknesses of the methods? Are there constraints that would hinder the use of the catch monitoring analyses?

Strengths of the analysis include**:** The study provides an evaluation of the observer effect in the groundfish fishery by gear type and area across years. The analysis was conducted at the

species/stock level which allowed for evaluation of the relative importance of the observer effect for certain stocks (e.g. Gulf of Maine cod). This study enabled identification of the direction of the observer impact (i.e. landings ratios on observed trips are greater/less than unobserved trips) and the relative magnitude of the effect through the ratio of observed to unobserved landings ratios (ratio of ratios). However, it stops short of characterizing the magnitude of the observer effect in quantities of pounds of fish landed (or discarded).

Weaknesses of the analysis include**:** The characterization of the results of this study seems to be narrow in scope and there is an opportunity to expand on this. An expanded description of the findings in the text and presentation of results in figures, rather than table format, would be helpful. For example, plots of landings ratios across years, may be easier for the reader to resolve patterns than the table format. The conclusions of this study could be strengthened through approaches that would allow for characterization of the variance between observed and unobserved landings ratios and explicit statistical testing for an observer effect. There is unequal sample size between observed and unobserved trips, with unobserved trips being an order of magnitude greater in number than observed trips. This unequal sample size could result in unequal variance that would challenge the ability to draw robust conclusions about an observer effect. Bootstrap resampling of trips could resolve this issue and could provide a good alternative to derive confidence intervals around the ratio estimates. This would allow for evaluation of whether observed landings ratios fall outside the bootstrap estimates of unobserved landings ratios. Alternatively, statistical testing of an observer effect could be addressed by conducting the analysis at the trip level and employing a generalized linear model or generalized mixed model (which may accommodate unbalanced design) to test for the effect of year, area, observer, and vessel factor (e.g. size/tonnage). Care should be taken in selection of the statistical distribution for modeling trip-level ratios however, since the distributions of ratios are likely to be highly skewed.

5. Are the conclusions of the Plan Development Team supported by the analyses (see 1e)?

In general, the conclusions of the PDT are consistent with the results of the analysis. The PDT concluded that: 1) there are discrepancies between the observed and unobserved trips, 2) more cod landings were seen on observed vs. unobserved trips in the Gulf of Maine stock area, and 3) less haddock landings were seen on observed vs. unobserved trips. The PDT cautioned against interpretation of the magnitude of the differences and indicated that the results were not likely an accurate estimation of the true extent of the potential missing removals. The study identifies absolute and relative differences in landings to effort ratios between observed and unobserved trips and may be useful in informing identification of patterns across years, species, and area. However, without characterization of the variance of observed and unobserved landings to effort ratios and statistical testing of differences it is challenging to draw robust conclusions. This study is helpful in identifying the problem, but it is challenging to use this in characterizing the magnitude of the problem. Also, the paper is helpful in that is supports findings of other papers.

However, the review panel would recommend further research to strengthen the robustness of this analysis.

6. Are there recommendations for improving the analyses, or for additional research or data collection that can help address improving groundfish monitoring?

The conclusions of this study could be strengthened through additional analysis that would allow for characterization of the variance between observed and unobserved landings ratios and statistical testing of the observer effect. To characterize variance, the reviewers suggest conducting bootstrap resampling of trips. This would allow for evaluation of whether observed landings ratios fall outside the bootstrap estimates of unobserved landings ratios.

Furthermore, the reviewers suggest calculating landings ratios at both the individual trip and aggregated gear level. This would allow for different types of statistical analysis to be applied to the dataset. For example, utilizing data at the trip level would allow for application of a generalized linear model or generalized mixed model to test for observer effect as well as the influence of other factors, such as year, area, and vessel factors.

The analyst could take a next step toward characterizing the magnitude of the illegal discard problem by calculating what the landings would have been based on the ratio of observed to unobserved trips on a species basis. This could provide perspective on the magnitude of the discard problem; however, the reviewers suggest that the analyst is the best judge as to whether calculation of magnitude is appropriate based on knowledge of the data and analysis.

V. Term of Reference 8

a. Reviewer Summary Comments:

This was a term of reference for all of the methods and was titled "Are the data, methods, and analytic tools sufficient for the Council to identify and analyze monitoring alternatives for the Northeast Multispecies Fishery Management Plan Amendment 23 management action?" The reviewers felt that this term of reference was cross cutting for all of the methods, therefore it was given its own section and answered in a way that is comprehensive across all of the methods examined.

b. Reviewer Comments:

Each of the methods has strengths and weakness, but together the set of studies provide substantial support to conclude that there are differences both in discarding behavior and in fishing behavior between observed and unobserved trips. The analyses suggest that discard estimates from observed trips should not be used to estimate discards from unobserved trips, or at minimum not without some adjustments. In addition, this suggests it is not appropriate to determine a level of observer coverage that should be deployed by considering the coefficient of variation of discard estimates from observer coverage since observed trips are not representative of unobserved trips. Furthermore, these studies suggest that the direction of the impact of the observer effect on landings (positive or negative) appears to vary by species/stock.

The analyses do not quantify the magnitude of the problem of unaccounted discards. Both PDT analyses reviewed in section III (*methods to predict groundfish catch in the presence of observer bias*) and section IV (*methods to evaluate groundfish catch ratios*) of this report both could be used to provide estimates of the total quantity of unreported discards relative to annual catch limits (ACL) or ABCs with some additional refinement.

By way of some potential pathways for the Council to use this information, the reviewers offer the following comments. First, if the percentage of the ACL that is discarded on unobserved trips is not large (e.g. less than 10%) then it might be feasible to use the section III (*methods to predict groundfish catch in the presence of observer bias*) approach to estimate discards on unobserved trips and use this to determine an appropriate buffer between the ABC and ACL to account for management uncertainty.

Second, if discards are a large proportion of the ACL, then the above approach is unlikely to be successful and may be counterproductive. What we mean by this is that the increased buffer would have to be large and would aggravate the illegal discard problem, which could make estimating discards for unobserved trips more difficult and uncertain. In this situation, rather than attempting to estimate the discards, the analysis reviewed in section I (*methods to explore discard incentives of groundfish stocks*) suggests that there may be a need for increased monitoring and enforcement or increased penalties to deter illegal discarding. It would be useful to apply the discard incentive analysis to observed trips to see if a higher percentage of landed catch has positive discard incentives which would indicate more clearly the likelihood that discarding is occurring on unobserved trips. However, this would still suffer from the problem that observed trips' catch composition may differ from unobserved trips due to differences in fishing behavior.

In conclusion, the reviewers note that unaccounted mortality from the fishery is one of several contributors to issues in our understanding of groundfish populations. Resolving to better understand this potential bias will be a step forward in improving our understanding of groundfish populations and will contribute to improved accounting of fishery mortality in our management process.

References

Holland, D.S. 2013. Making cents out of barter data from the British Columbia groundfish ITQ market. *Marine Resource Economics*, *28*(4), pp.311-330.

Benoît H. P., Allard J. 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? Canadian Journal of Aquatic Sciences 2009; 66:2025-2039.

Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37: 547–579.

Appendices

Appendix 1 – Terms of Reference

New England Fishery Management Council
50 WATER STREET | NEWBURYPORT, MASSACHUSETTS 01950 | PHONE 978 465 0492 | FAX 978 465 3116
John F. Quinn, J.D., Ph.D., *Chairman* | Thomas A. Nies, *Executive Director*

MEMORANDUM

DATE: April 12, 2019

TO: Scientific and Statistical Committee Subpanel for Groundfish Monitoring

FROM: Tom Nies, Executive Director

SUBJECT: **Terms of Reference – Review of Groundfish PDT Information and Analyses of Groundfish Monitoring Issues**

Background

The New England Fishery Management Council has organized this review to ensure that any new and novel analyses of Amendment 23 issues and management alternatives get sufficient independent review.

Terms of Reference

1. For each of the Plan Development Team's four analytic methods listed below (1a – 1d), please address the questions in numbered sections 2 through 8 below:
    a. Methods to explore discard incentives and estimate prohibited discards of groundfish stocks.
    b. Methods to evaluate observer effects in the groundfish fishery.
    c. Methods to predict groundfish catch in the presence of observer bias.
    d. Methods to evaluate groundfish catch ratios.
    e. Groundfish PDT conclusions based on the analyses conducted

2. Are the methods adequately described and based on sound analytic techniques and statistical principles?
3. Are important uncertainties in the data and the analyses (possibly including the effects of year to year variations in fishing practices) identified, and are the impacts of these uncertainties on the analyses adequately described?
4. Are the analyses conducted at the appropriate temporal and spatial scale such that the existence of regional or seasonal differences in monitoring performance can be identified?
5. What are the strengths and weaknesses of the methods? Are there constraints that would hinder the use of the catch monitoring analyses?
6. Are the conclusions of the Plan Development Team supported by the analyses (see 1e)?
7. Are there recommendations for improving the analyses, or for additional research or data collection that can help address improving groundfish monitoring?
8. Are the data, methods, and analytic tools sufficient for the Council to identify and analyze monitoring alternatives for the Northeast Multispecies Fishery Management Plan Amendment 23 management action?

1. **Groundfish Plan Development Team (PDT) Analyses to be Reviewed**
   a. Methods to explore discard incentives and estimate prohibited discards of groundfish stocks
   b. Methods to evaluate observer effects in the groundfish fishery
   c. Methods to predict groundfish catch in the presence of observer bias
   d. Methods to evaluate groundfish catch ratios
   e. Groundfish PDT conclusions based on the analyses conducted
   f. Presentations

2. **Additional background materials**
   a. Draft Alternatives for Amendment 23/Groundfish Monitoring
   b. Gulf of Maine cod hotspot analysis by the Groundfish PDT

Appendix 2 – Review Agenda

New England Fishery Management Council
50 WATER STREET | NEWBURYPORT, MASSACHUSETTS 01950 | PHONE 978 465 0492 | FAX 978 465 3116
John F. Quinn, J.D., Ph.D., *Chairman* | Thomas A. Nies, *Executive Director*

Meeting Agenda

New England Fishery Management Council

Peer Review of
Groundfish Plan Development Team Analyses of Groundfish Monitoring

Hotel Providence
139 Mathewson Street, Providence, RI 02903
Telephone: (401) 861-8000

Day 1- Wednesday, April 24, 2019

10:00	Welcome, introductions and agenda review (Chair)
10:15	Review of Terms of Reference (Chair)
10:30	Groundfish Plan Development Team (PDT) Presentations and Panel Questions • Overview (Cournane) • Methods to explore discard incentives and estimate prohibited discards of groundfish stocks (Errend) • Methods to evaluate observer effects in the groundfish fishery (Demarest) • Methods to predict groundfish catch in the presence of observer bias (Linden) • Methods to evaluate groundfish catch ratios (Nitschke) • Groundfish PDT conclusions based on the analyses conducted (Cournane)
12:30	LUNCH
1:15	Presentations and Panel Questions (*continued*)
3:30	Opportunity for public comments
3:45	Review Panel discussion *****
5:30	Adjourn

- ***Note: The Chair may take public comments throughout the Review Panel discussion as appropriate.***

Day 2- Thursday, April 25, 2019

8:30	Review Panel comments and development of Review Panel report *****
1:30	Adjourn

- ***Note: The Chair may take public comments throughout the Review Panel discussion as appropriate.***

This meeting is physically accessible to people with disabilities. This schedule is subject to change. If you have questions, please call the Council office for final confirmation of meeting times, dates and locations.