

A. GULF OF MAINE ATLANTIC COD (*GADUS MORHUA*) STOCK ASSESSMENT FOR 2011, UPDATED THROUGH 2010

The Northern Demersal Working Group (NDWG) prepared the assessment. The working group (Appendix 1) held three different meetings over a three month period. The meeting dates and locations are listed below. Working group participation differed by meeting. A complete summary of working group participants by meeting and day are presented in Appendix 1.

- NDWG Gulf of Maine Cod Industry Meeting (NDIM)
 - August 16, 2011
 - Massachusetts Department of Marine Fisheries (MADMF) Annisquam Field Station, Gloucester, MA

- NDWG Gulf of Maine Cod Data Working Group (NDDWG) Meeting
 - September 7-9, 2011
 - Northeast Fisheries Science Center (NEFSC), Woods Hole, MA

- NDWG Gulf of Maine Cod Models and Biological Reference Points Working Group (NDMBRPWG) Meeting
 - October 17-21, 2011
 - Falmouth Technology Park, Falmouth, MA

SAW 53 Terms of Reference

A. Gulf of Maine Atlantic cod (*Gadus morhua*)

1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch.

2. Present the survey data being used in the assessment (e.g., indices of abundance, recruitment, state surveys, age length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

3. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, catch recruitment and fishing mortality.

4. Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (TOR-3).

5. If time permits, consider the small-scale distribution of cod (e.g., spawning sites, resource distribution, fishing effort) in the Gulf of Maine and advise on its management implications.

6. State the existing stock status definitions for “overfished” and “overfishing”. Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY} , $B_{THRESHOLD}$, F_{MSY} , and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the “new” (i.e., updated, redefined, or alternative) BRPs.
7. Evaluate stock status with respect to the existing model (from the most recent accepted peer reviewed assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
 - a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
 - b. Then use the newly proposed model and evaluate stock status with respect to “new” BRPs (from Cod TOR-6).
8. Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
 - a. Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F , and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
 - b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
 - c. Describe this stock’s vulnerability (see “Appendix to the SAW TORs”) to becoming overfished, and how this could affect the choice of ABC.
9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments and review panel reports. Identify new research recommendations.

Executive Summary

1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch.

Since 1964, catch of Gulf of Maine Atlantic cod has ranged from 3,242 mt to 22,272 mt. Recent catches over the past five years have ranged from approximately 5,000 mt to 11,000 mt. Catch estimates prior to 1981 do not include commercial discards or estimates of recreational removals. Since 1982, commercial landings have been the largest source of fishery removals, comprising 40-90% of the total catch. Commercial discards constituted a large proportion of the catch between 1998 and 2003 when trip limits ranged from 30-500 lb/day (13.6 – 226.8 kg/day). Since 2006 commercial discards have accounted for <10% of the catch. Major uncertainties in the commercial catch include mis-allocation of commercial landings stemming from industry mis-reporting of statistical area and uncertainty in the discard estimation method. The uncertainty with respect to mis-reporting is likely to be small (5%). In recent years precision of the estimated discards has been high with coefficients of variation (CV) <20%. The updated assessment has included hindcasted commercial discard back to 1982 and the uncertainty on these estimates is unknown.

There is a large recreational fishery in the Gulf of Maine that, over the last decade, has accounted for approximately 30-50% of the total catch. Recreational discards have become an increasingly important component of fishery removal and as of 2010, constitute 20% of total catch. Uncertainty in the recreational catch is on the order of 10-25% in terms of percent standard error (PSE). An additional source of uncertainty is the age composition of recreational discards prior to 2005. The updated assessment has attempted to hindcast recreational discard length frequency distributions back to 1981 so that this fraction of the catch could be incorporated into an age-based assessment. Previous Gulf of Maine cod assessments have not accounted for recreational discards.

The Northern Demersal Data Working Group (NDDWG) reviewed findings from the scientific literature about the discard survival of Atlantic cod and other similar species. It must be emphasized that the working group found this TOR very difficult to address. The working group discussed all gears for which discards were estimated in the updated SAW 53 assessment, with each gear being evaluated separately based on the gear-specific information available from the literature. While each study provided an estimate of survival, no single study could address every factor implicated in mortality. Important factors in determining discard survival from the available scientific literature include: water and air temperature, sunlight exposure, depth of capture, time of handling, type of handling, length of time on deck, short term and long term survival (one study estimated that only about 50% of mortality occurred in first few days—the length of most observation periods), impacts on growth due to reduced feeding ability, whether predator avoidance was compromised or predator exposure was increased at release time (birds, mammals, other fish predators), whether fish were held on deck in tanks or in an aquarium or held in a cage at depth. Each gear was evaluated with respect to available studies with survival estimates, what factors had been accounted for, what factors had not been accounted for, and whether it was possible to determine what conditions were likely to have existed for unobserved trips. Because it is not possible to characterize the

temperature/depth/season for all unobserved trips, a single, annual discard mortality rate is required. The working group was consistent in how it approached the evaluation of each gear, first by reviewing the available studies, discussing what factors were and were not controlled for, and whether the estimates in the literature were likely to be biased high or low. In the end, the working group did agree that the published studies probably overestimated survival, although it was difficult to characterize the extent of the bias. The discard mortality rates to be used in SARC53 for Gulf of Maine cod are 100% for all gears.

2. Present the survey data being used in the assessment (e.g., indices of abundance, recruitment, state surveys, age length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

The Northeast Fisheries Science Center (NEFSC) spring and fall bottom trawl surveys began in 1968 and 1963 respectively, providing a long time series of fishery independent indices. Age-specific indices for Gulf of Maine cod began in 1970. All previous Gulf of Maine cod assessments have used only the offshore survey strata. The impacts of including the inshore survey strata in the NEFSC survey indices was examined by the NDDWG and resulted in increased indices of age 0 through 2 fish. The overall trend in the age specific indices of older fish was not markedly different with the inclusion of the inshore strata, and there were several strata/year combinations with poor sampling. For this reason, and because the inshore areas are largely covered by the Massachusetts Department of Marine Fisheries (MADMF) bottom trawl survey, the NDDWG decided to maintain the status quo and exclude the inshore strata from NEFSC indices. The NEFSC survey vessel was replaced in spring 2009 resulting changes to the survey protocol. Calibration experiments to estimate differences in catchability between the two survey series were conducted and peer-reviewed. Length based calibration models were used to express the 2009-2011 NEFSC indices in units equivalent to the longer time series.

The MADMF bottom trawl survey began in 1978, with two surveys (spring and fall) conducted annually. Age-specific indices are available beginning in 1981 for the fall and 1982 for the spring. In previous assessments the MADMF fall survey has been used primarily as a recruitment index. In the updated assessment, the utility of this survey was evaluated and was not included in the final base model.

Previous Gulf of Maine cod assessments have included a landings per unit effort (LPUE) index that extended from 1982 to 1993. The time series has not been extended beyond 1994 due to uncertainties in VTR reported fishing effort since 1994, the impact of reductions in days at sea, rolling closures and trip limits. All of these issues would affect the comparability of LPUEs estimated from 1994 onward with the earlier time series. Additionally, these same issues would make standardization of a contemporary catch per unit effort (CPUE) index difficult. The continued inclusion of the existing LPUE index was evaluated by the Northern Demersal Models and Biological Reference Point Working Group (NDBRPWG). Model results were found to be insensitive to this index, and the decision was made to exclude this index from the final base model.

Several other surveys were evaluated including the Maine – New Hampshire inshore trawl survey and the MADMF Cod Industry Based Survey. For several reasons including lack of age-specific information and short time series, these surveys were not included in the assessment models. The surveys were however used to inform several decisions made by the NDDWG and NDMBRPWG with respect to assumptions about spawning time and gear selectivity.

3. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, catch recruitment and fishing mortality.

The VPA assessment model used for the most recent assessment of Gulf of Maine Atlantic cod (GARM III, 2008) was updated to account for the major changes to the data inputs as well as three additional years of catch and survey data. The major changes to the input data include:

- Updated length-weight equations.
- Updated maturity ogive.
- Re-estimated commercial landings-at-age.
- Re-estimated discards-at-age including extension of discards back to the beginning of the assessment time series.
- Re-estimated recreational landings-at-age.
- Estimation of recreational discards-at-age.
- Updated catch and stock weights-at-age.
- Re-estimated survey indices.

*The updated VPA estimates SSB_{2010} at 12,270 mt and $F_{5-7(2010)}$ at 1.48. The GARM III VPA assessment estimated SSB_{2007} at 33,877 mt and $F_{5-7(2007)}$ at 0.46. Comparatively, the updated VPA now estimates SSB_{2007} at 10,714 mt and $F_{5-7(2007)}$ at 0.68. The general conclusions from the updated VPA are that the weights-at-age used in GARM III likely overestimated the true stock weights-at-age. In addition, the GARM III results overestimated the size of the 2003 and 2005 year classes. The size of these year classes was derived almost exclusively from survey information. As of 2007 these year classes were only partially recruited to the fishery, so there was little information to counter the signals coming from the surveys. Relative to the 2010 update of the VPA assessment, the 2008 VPA assessment over estimated spawning stock biomass, the strength of incoming year classes and underestimated fishing mortality. **The updated VPA is not the base model for this assessment.***

In this updated assessment a statistical catch-at-age model (ASAP) represents the new base model. The reasons for selecting the ASAP model include: the ability to explore alternative model formulations to counter/lend support to VPA results, additional flexibility to explore starting condition assumptions (e.g., extending the time series beyond 1982), ability to estimate a stock-recruit relationship internal to the model, and the ability to explicitly handle data uncertainty, particularly given the lessons learned from the update of the VPA model.

The ASAP base model configuration (BASE) reflects the best model with which to evaluate stock

status and provide catch advice. The assessment indicates that total SSB has ranged from 7,270 mt to 23,675 mt during the assessment time period, with current SSB in 2010 estimated at 11,868 mt (90% CI = 9,479 – 16,301 mt). The base model estimates SSB in 2007 at 12,561, 37% of the 33,877 mt estimated at GARM III. Currently, total biomass is estimated at 20,589 mt (90% CI = 17,638 – 25,996 mt). Current F 's are near historic highs with fully recruited $F_{full} = 1.14$ (0.79 – 1.54) and $F_{5-7} = 1.10$ (90% CI = 0.74 – 1.46).

A retrospective analysis for the 2003-2010 terminal years indicates retrospective error in both F and SSB with the tendency for the model to underestimate F and overestimate SSB. The F retrospective error ranged from -0.10 in 2009 to -0.52 in 2003. SSB retrospective error ranged from 0.09 in 2009 to 0.90 in 2003. Over the last 5 years, retrospective bias has resulted in a 22% overestimation of SSB and 22% underestimation of fishing mortality. Retrospective error in age 1 recruitment varied from -0.07 in 2005 to 4.32 in 2003. It is worth noting the decreased retrospective pattern in age-1 recruitment in the ASAP BASE run, relative to the updated VPA model. The ASAP model does not exhibit the severe retrospective pattern in the recent period, particularly in the 2008 assessment peel (coinciding with the timing of the GARM III assessment). Consequently, had an ASAP model been used at GARM III, it is likely that the 2005 year class would have been estimated to have been much lower and the perception of the stock would have been far less optimistic than the GARM III results suggested.

4. Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (TOR-3).

Historically, the recreational fishery has been split between Georges Bank and the Gulf of Maine. Since 1999, recreational landings of Atlantic cod have been predominately in the Gulf of Maine region (NEFSC 2008). The potential for misallocation of recreational landings is unknown; however, given the behavior of the recreational fleet operating in the Gulf of Maine, the magnitude of impacts is likely to be small. The issue is misallocation of commercial landings is likely to be larger and have a greater impact on model performance. With respect to Gulf of Maine Atlantic cod, the allocation procedure itself does not likely contribute additional uncertainty as indicated by the low CVs on the allocated landings. A more likely source of allocation uncertainty arises from the misreporting of statistical area on the VTRs. The work of Palmer and Wigley (2007, 2008, and 2010) suggests that these impacts are likely to be small (<5%), but consistently unidirectional (under-reporting of total Gulf Maine cod catch).

Sensitivity runs were conducted to bound the potential impacts of mis-allocation. Two sensitivity runs were conducted, one which inflated landings by 5% and another which decreased landings by 5%. Spawning stock biomass changed +/- 5% with no change in F . The 2010 estimates of SSB were within the 90% confidence intervals achieved from the MCMC estimate of uncertainty (9,479 – 16,301 mt).

5. If time permits, consider the small-scale distribution of cod (e.g., spawning sites, resource distribution, fishing effort) in the Gulf of Maine and advise on its management implications.

Discussion related to resource distributions occurred throughout the NDDWG meeting as both surveys (NEFSC, MADMF, ME-NH, and the Industry Based Survey) and fleet activity were reviewed. Given the full agenda, and extent of reanalysis of data, there was not an abundance of time available to delve into this TOR. Nevertheless, some time was set aside and the working group attempted to review as much as possible during that time block. One presentation summarizing tagging in the western Gulf of Maine was presented, however further discussion of this TOR was reserved until after the discard mortality TOR had been completed. The work examined confirmed that most of the fish on the spawning aggregations show site fidelity; that the timing and extent of the closures is appropriate; and that when fishing resumes at the end of the closure, it can be very disruptive to the cod (interrupts any residual spawning because the fish rapidly disperse from the spawning grounds). Moreover, the industry based survey confirms generalized patterns observed in both MADMF and NEFSC surveys, with cod moving offshore in the fall and inshore in the spring. Additionally, information from a preliminary longline survey in Downeast Maine identified the scarcity of cod in that region.

6. State the existing stock status definitions for “overfished” and “overfishing”. Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY} , $B_{THRESHOLD}$, F_{MSY} , and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the “new” (i.e., updated, redefined, or alternative) BRPs.

The existing MSY reference points are based on a spawning potential ratio (SPR) of 40%. The overfishing definition is $F_{MSY} = F_{40\%} = 0.237$. A stock is considered to be overfished if spawning biomass is less than half of SSB_{MSY} . The existing overfished definition is $0.5 \times SSB_{MSY} = 0.5 \times SSB_{40\%} = 0.5 \times 58,248 \text{ mt} = 29,124 \text{ mt}$. The existing MSY reference points were derived from a VPA model with a plus group-at-age 11. There are a number of reasons why new reference points are needed for the proposed base model for the current assessment. The number of age classes modeled is 9 instead of 11 (this changes the weight and selectivity in the plus group), commercial and recreational discards are included (this changes the weights and selectivities at all ages), the parameters of the $L-W$ equation were re-estimated (this also affects weights at all ages), and the time elapsed before spawning was increased from 0.1667 (March 1) to 0.25 (April 1) which will affect biomass discounting in the YPR calculations.

The current reference points were derived at GARM III, and are based on $F_{40\%}$. The decision to use $F_{40\%}$ as a proxy for F_{MSY} was endorsed by the independent reviewers at the GARM III meeting, who commented that $F_{40\%}$ is supported by published studies on sustainability. It was pointed out that the published studies focused on F_{MSY} proxies that emphasized sustainability while minimizing yield loss rather than the implications for rebuilding. There were different views within the NDMBRPWG as to the relative priorities of focusing on sustainability and minimization of yield loss, versus implications for biomass targets and rebuilding. Several F_{MSY} proxies were debated: $F_{22\%}$ (F_{MAX}), $F_{35\%}$, and $F_{40\%}$ (status quo). The SARC Panel determined that $F_{40\%}$ was an appropriate reference point for the analyses considered.

To arrive at estimates for $SSB_{40\%}$ and corresponding MSY , long term projections were run,

sampling from the empirical distribution of recruitment estimates from the preferred ASAP model (recruitment estimates from 1982-2008, final two years excluded). The resulting reference points and their 90% confidence intervals corresponding to $F_{MSYproxy}=F_{40\%}=0.20$ are $SSB_{MSY} = 61,218$ mt (46,905 – 81,089 mt), $MSY = 10,392$ mt (7,825 – 14,146 mt).

7. Evaluate stock status with respect to the existing model (from the most recent accepted peer reviewed assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.

a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

The existing peer reviewed assessment model is a VPA. A meticulous bridge was built from the existing VPA model structure to the updated VPA model structure. The updated VPA model, which includes changes to the catch (inclusion of discards), weights-at-age, etc., estimates that in 2010 SSB is 12,270 mt. This is less than the existing overfished threshold of 29,124 mt; therefore, the stock is overfished. The updated VPA estimate of average fishing mortality on ages 5-7 in 2010, F_{5-7} is 1.48, while the fully recruited F from the VPA is $F_{full}=2.46$. These are both greater than the overfishing limit, and therefore, overfishing is occurring.

b. Then use the newly proposed model and evaluate stock status with respect to “new” BRPs (from Cod TOR-6).

The revised reference points are $F_{MSYproxy}=F_{40\%}=0.20$ and $SSB_{MSY} = 61,218$ mt ($0.5 \times SSB_{MSY} = 30,609$ mt). The proposed ASAP base model estimate of 2010 SSB is 11,868 mt. This is less than the overfished threshold of 30,609 mt; therefore, the stock is overfished. The 2010 estimate of average fishing mortality on ages 5-7 from ASAP is $F_{5-7}=1.10$, while the fully recruited F_{full} is 1.14. This is greater than the overfishing limit of 0.20, and therefore, overfishing is occurring.

Accounting for the retrospective bias does not result in a change of stock status and the revised stock status lies within the confidence intervals of the unadjusted point. The NDMBRPWG reached consensus that the stock status determination from the ASAP base model without accounting for retrospective bias was preferred. The precedence established at GARM III was to only make retrospective adjustments when the adjusted point fell outside the confidence intervals of the unadjusted point. This approach was supported by the SARC Panel.

For both the existing VPA model with respect to existing reference points and the new proposed ASAP base model with respect to updated reference points, the stock is overfished and overfishing is occurring. Consequently, for both models and reference point sets, the stock is not rebuilt.

8. Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

a. Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F , and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

Short term projections of future stock status were conducted based on the current assessment results without accounting for retrospective bias. The NDMBRPWG did not support the use of hindcasted recruitment for the same reasons they rejected the historical ASAP sensitivity runs; recruitment estimates based solely on survey information have proven unreliable to use as the basis for stock determination. Projections were run under three different F assumptions: $F_0=0.00$, $F_{MSY(F40\%)} = 0.20$, and $F_{75\%FMSY} = 0.15$.

Projection results indicate that even the most optimistic scenario in terms of rebuilding (F_0), the stock cannot rebuild to SSB_{MSY} by the current rebuilding date of 2014.

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

Given the noted retrospective patterns, there should be additional uncertainty in catch advice based on these projections. Moreover, the projections will be sensitive to realized recruitment. Recent recruitment has been weak with no strong recruitment observed in the last twenty years. Continued weak recruitment will impede the ability for this stock to rebuild. Given the poor performance of past projections beyond a time period of two to three years, the longer term projections presented in this report should be considered highly uncertain.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

Uncertainties that were not accounted for by assessment and reference point models were evaluated using model diagnostics. Standard model diagnostics (e.g., residual analyses, retrospective analyses) were used for model validation. Vulnerabilities that were not accounted for by assessment and reference point models were evaluated using exploratory modeling, habitat observations and preliminary results from studies conducted in the spawning closure areas. Those studies indicate strong site fidelity to the spawning grounds, and the almost immediate disruption of spawning activity when those areas are opened. This would suggest that area closures to protect spawning grounds is beneficial and could reduce vulnerability. Additional considerations of vulnerability and productivity are the implications of shifts in distribution, recruitment dynamics and increased natural mortality. Consumption of Atlantic cod by other fishes and mammals may be increasing as predator populations increase, however empirical evidence is lacking to support this hypothesis directly. A considerable source of additional vulnerability is the continued weak recruitment and low reproductive rate (e.g., recruits per spawner) of Gulf of Maine cod. If weak recruitment and low reproductive rate

continues, productivity and rebuilding of the stock will be less than projected.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments and review panel reports. Identify new research recommendations.

Five of the six previous research recommendations have either been addressed or shown to be no longer relevant. The one research recommendation that has not been addressed (Maine-New Hampshire Inshore Trawl Survey) has been carried forward as a new research recommendation for SARC 53. There were a total of four new research recommendations to come out of the NDDWG.

Introduction

Stock structure

Atlantic cod (*Gadus morhua*) is a demersal gadoid species whose range in United States (US) waters extends from Cape Hatteras north to the Canadian border. Globally, Atlantic cod occur on both sides of the North Atlantic Ocean, extending southward in the eastern Atlantic to the Bay of Biscay. Within the United States Exclusive Economic Zone (EEZ) there are two recognized stocks of cod: Gulf of Maine and Georges Bank. Recent reviews of historical and contemporary tagging studies (O'Brien et al. 2005, Tallack 2007, Loehrke and Cadrin 2007) suggest, that while there is movement of fish between the Gulf of Maine and Georges Bank stocks, the degree of mixing is less than 20% (Tallack 2009, T. Miller pers. comm.). Additionally, within the Gulf of Maine there are likely localized metapopulations (Ames 2004), between which, the degree of mixing is unknown. The Gulf of Maine of Maine stock complex extends from the northern tip of Cape Cod east to the US/Canadian border and north to the coast of Maine (Fig. A.1).

Assessment history

The initial analytical assessment of the Gulf of Maine stock was conducted using a virtual population analysis (VPA) model by Serchuk and Wigley (1986) and presented at the 7th Northeast Fisheries Science Center (NEFSC) Stock Assessment Workshop (SAW) in 1988 (NEFSC 1989). Subsequently, the stock was reviewed again at SAW 12, 15, 19, and 24 (NEFSC 1991, 1993, 1995, 1997, 1998; Mayo 1995, 1998, Mayo et al. 1993, 1998). Additionally, interim assessments were reviewed outside of the SAW framework by the Northern Demersal Working Group in July 1999 (NEFSC 2000) and again in August 2000 (NEFSC 2001a).

Amendment 4 (1991) to the Multispecies Fisheries Management Plan implemented $F_{20\%}$ as an overfishing mortality threshold for Gulf of Maine cod. Estimates of $F_{20\%}$ and F_{max} are shown below (*note $F_{20\%}$ was not reported in the SAW 7 documents):

Stock assessment workshop	Year	$F_{20\%}$	F_{max}	Model type	Notes
SAW 7	1988		0.27	VPA	Commercial landings only
SAW 12	1991	0.40	0.27	VPA	Commercial landings only
SAW 15	1993	0.36	0.25	VPA	Commercial landings only
SAW 19	1995	0.35	0.27	VPA	Commercial landings only
SAW 24	1997	0.37	0.29	VPA	Commercial landings only
SAW 27	1998	0.39	0.29	VPA	Commercial landings only

The 1996 re-authorization of Magnuson-Stevens Conservation and Management Act required the redefining of overfishing and overfished with respect to the rate of fishing mortality associated with producing maximum sustainable yield. SAW 27 provided estimates of F_{MSY} and B_{MSY} based on the ASPIC surplus production model with survey catchability coefficients conditioned on biomass estimates from the SAW 27 VPA. These estimates were mean age 1⁺ biomass $_{MSY}$ = 33,000 mt and age 1⁺ biomass weighted F_{MSY} = 0.31. This method was used in the Report of the Overfishing Definition Review Panel (Applegate et al. 1998) and the corresponding reference points were adopted in Amendment 9 to the multispecies FMP. The biomass threshold was set at $\frac{1}{4} B_{MSY}$ (8,300 t).

In the last decade, the Gulf of Maine cod stock has undergone four peer-reviewed assessments: SAW 33 (NEFSC 2001), the Groundfish Assessment Review Meeting (GARM, NEFSC 2002), GARM II (NEFSC 2005) and GARM III (NEFSC 2008). Summaries of these assessments and the resulting stock status are provided in Table A.1 and A.2. All of these assessments were conducted using the ADAPT VPA model with a starting year of 1982. The data inputs from SAW 33 through GARM II were nearly identical, with

GARM I and II representing updates to the SAW 33 model inputs. Commercial discards were accounted for by increasing the total landings by 500 mt increments; the size of the increase was determined based on the estimated discards. This method assumes that the discarded fraction of the catch is of the same size composition as the landed catch. In the existence of trip limits, this assumption may be appropriate, but when discarding is occurring primarily as a result of minimum retention sizes, such a method may incorrectly characterize the age composition of the catch. Recreational landings were included in these assessments, but recreational discards were not. Additionally, catch and stock weights-at-age were estimated solely from the landed fraction of the catch. When discards due to minimum sizes restrictions contribute a sizeable fraction of overall removals, this method has the potential to overestimate stock biomass.

SAW 33 included catch through 2000 and survey indices through 2001 (spring only). SAW 33 re-evaluated reference points using an age based production model with a Beverton-Holt stock recruit relationship (NEFSC 2001b). Reference points were estimated as total stock age 1⁺ total biomass $B_{MSY}=90,300$ mt, $SSB_{MSY}=78,000$ mt, and $F_{MSY}=0.23$. The SAW 33 assessment concluded that Gulf of Maine cod were not over fished, but overfishing was occurring. It is noteworthy that the stock status determination applied at SAW 33 was different than the current basis. For SAW 33 the overfished definition was based on $\frac{1}{4} B_{MSY}$ criteria (Applegate et al. 1998) unlike the $\frac{1}{2} SSB_{MSY}$ that was later adopted by the Working Group on Re-estimation of Biological Reference Points for New England Groundfish (NEFSC 2002b). The 2001 total stock biomass was estimated at 24,000 mt (18,000 mt SSB); just over 25% of B_{MSY} . Fishing mortality (F) was estimated at 0.73 which was over three times higher than F_{MSY} .

The Working Group on Re-evaluation of Biological Reference points for New England Groundfish (NEFSC 2002a) further revised Gulf of Maine cod reference points; SSB_{MSY} was revised to 82,800 mt based on change in the period used to derive mean stock weights. F remained unchanged. Amendment 13 (2004) to the Multispecies FMP adopted the Working Group's revised reference points ($SSB_{MSY}=82,800$ mt, $F_{MSY}=0.23$). The biomass threshold was revised to $\frac{1}{2} SSB_{MSY}$ (41,400 t). GARM I updated the data inputs by one year (through 2001) using the same VPA formulation as SAW 33. Spawning stock biomass in 2001 was estimated at 22,040 mt, approximately 25% of SSB_{MSY} . F was estimated at 0.47, two times greater than F_{MSY} . As of 2002 Gulf of Maine cod were overfished and overfishing was occurring. GARM II was a three year update (through 2004) to the GARM I assessment. Biological reference points remained unchanged from GARM I. Spawning stock biomass had declined to 18,800 mt in 2004 and F had increased to 0.63. The stock complex was still overfished and overfishing was occurring. The GARM II assessment exhibited a retrospective pattern in both F and SSB , with a tendency for F to be underestimated and SSB to be overestimated in the most recent three years.

The 2008 GARM III assessment represented a benchmark assessment update. Major changes from the previous assessments include a more thorough consideration of commercial discards and updates to the biological reference points. Unlike previous assessments where landings-at-age were increased in fixed amounts, the GARM III method applied an estimated discard ratio to the landings-at-age. While this method better characterizes the true trends in discards, it still makes the assumption that the age composition of the discards is identical to the landed fraction. It should be noted that the ratio increase in landings-at-age was only applied from 1999 to 2007. Prior to 1999, commercial discards were not accounted for. As in previous assessments, catch and stock weights-at-age were estimated solely from the landed fraction of the catch and recreational discards were not included in the catch estimates. Biological reference points were based on the non-parametric yield and SSB per recruit analysis with $F_{40\%}$ used as a proxy for F_{MSY} . The reference points were estimated as follows: $F_{MSY} = 0.237$ and $SSB_{MSY} = 58,248$ mt. Terminal year estimates of F were 0.46 and SSB was estimated to have increased to 33,877 mt. The stock was perceived to no longer be overfished, but overfishing was still occurring. The large increase in SSB was contingent on the relative strength of the 2003 and to a greater degree, the 2005 year classes. The

2005 year class was estimated at 23.9 million fish (age 1) which represented the second largest observed year class in the assessment time period. Given that these fish were only age 2 in 2007, they had yet to enter the fishery. The 2007 estimates of partial recruitment indicated that the vulnerability of this year class to the fishery was at less than 1%. The entire strength of the 2005 year class was primarily derived from the NEFSC spring and MADMF fall survey indices.

Fisheries management

Gulf of Maine Atlantic cod have been managed under two different management authorities in recent history. Prior to 1977 the 5Y component (statistical areas 511-515) of the stock was managed under an international treaty through the International Commission for the Northwest Atlantic Fisheries (ICNAF). Fisheries management was primarily controlled through annual total allowable catches (TACs) and minimum mesh sizes (Serchuk et al. 1994). The TACs remained constant at 10,000 mt between 1973 and 1975 followed by reductions to 8,000 mt in 1976 and then to 5,000 mt in 1977. The Magnuson Fishery Conservation and Management Act (MFMCA) was passed in 1977 and subsequently the management authority of the Gulf of Maine cod stock, as well as all other New England groundfish stocks, shifted to the New England Fishery Management Council (NEFMC).

The use of TACs continued under the NEFMC authority through 1982, with TACs dispersed among quarters and vessel tonnage classes. The early quota period was accompanied by poor catch monitoring and reported black markets for quota managed species and may have contributed to increased uncertainty over catches. The system adopted in the mid-80's had numerous exceptions and special programs to mesh and minimum size requirements that make it difficult to draw conclusions about how regulations influenced fishery selectivity. In 1982, the "Interim" Groundfish Fisheries Management Plan (FMP) was implemented which replaced the quota system (TAC) with input controls such as mesh sizes and minimum retention sizes (Table 3). The "Interim" FMP was replaced by the initial Groundfish FMP in 1985 which largely carried forward the existing measures from the interim FMP. Amendment 4 to the FMP required the use of a Nordmore grate in the northern shrimp fishery as well as placing a prohibition on the retention of groundfish bycatch. Beginning with Amendment 5 (1994), there was a concerted attempt to reduce fishing effort through a days-at-sea (DAS) reduction schedule. Additionally, Amendment 5 brought about mandatory vessel reporting in the way of the Vessel Trip Reports (VTRs). Effort controls were increased under Amendment 7 through further acceleration of the DAS reduction schedule, and the addition of seasonal and year round closures in the Gulf of Maine. Between 1997 and 1999 trips limits on Gulf of Maine cod were reduced from 1000 lbs/day to 30 lbs/day. Amendment 13, implemented in May 2004, placed additional restrictions on DAS usage while allowing for the use of regular B DAS to target healthy stocks. Additionally, Amendment 13 implemented mandatory electronic reporting for all primary federally permitted seafood dealers. In 2006, Framework 42 established reference point thresholds for the 18 groundfish stocks reviewed at GARM II as well as formalized rebuilding plans for all overfished stocks ($< \frac{1}{2} SSB_{MSY}$), such as Gulf of Maine cod. Through 2010 a series of additional framework actions and interim rules placed additional restrictions on DAS usage and seasonal closures on the recreational fishery.

The effort controls first adopted in 1994 were frequently changed making it difficult to isolate the effects of individual regulations. The use of often-changing trip limits led to increased discard rates and may have contributed to high-grading. Seasonal (rolling) and year-round closures may have limited fishery access to larger spawning fish, and strict DAS limits focused effort on easily caught nearshore cod and led to the increased use of sink gillnet gear.

In 2010 the groundfish fishery experienced a major management change with the passage of Amendment 16. Amendment 16, with the introduction of annual catch limits (ACLs), represented a return to the use of

hard TACs. Additionally, 17 new groundfish sectors were approved and those vessels not members of a groundfish sector were subject to additional cuts in DAS and restrictive trip limits. Vessels fishing under the sector management were exempt from DAS restrictions and instead, each sector was given a share of the total commercial groundfish sub-ACL. How the catch was divided up amongst sector vessels or how catch was allocated throughout the year was left to the sole discretion of the sector. One of the requirements of Amendment 16 was an increase in the overall level of observer coverage. This was accomplished using observers trained through the existing Northeast Fisheries Observer Program (NEFOP) as well as a new class of observers termed At-Sea Monitors (ASMs). The data collection protocols for ASMs were restricted to catch estimation and the collection of limited biological information (e.g., lengths). The recent shift to a catch share system in 2010 appears to have dramatically reduced discards but it is too soon to fully understand the overall impacts of the sector management system.

Length-weight relationship

Previous assessments of the Gulf of Maine cod stock have used an annual NEFSC research vessel survey length-weight (LW) equation as the basis for converting catch weights to numbers-at-age (Equation 1). The origin of the equation and nature of the data used to estimate it (survey or commercial landings) are uncertain. The equation differs from updated NEFSC survey-based LW equations estimated by Wigley et al. (2003). Because the source of the original equation could not be documented and because continued use of it would not account for seasonal differences in the LW relationship, a decision was made to re-evaluate the existing LW relationship with respect to re-estimated length-weight equations.

$$(1) \quad W = 0.000008104L^{3.0521} \quad (\text{GARM III and prior})$$

There are two schools of thought as to whether it is more appropriate to use a landings-based length-weight equation versus a survey-based length-weight equation. Advocates for a landings-based derivation argue that since the fishery may catch larger (heavier) fish at length, there is the possibility that a survey-based length weight equation may be biased low, particularly at greater lengths. A survey-based approach may be preferred when a large portion of the catch comes from discards (or some other fraction not sampled such as recreational landings) or when the catch weights-at-age are also used to estimate stock weights due to sparse sampling of older ages in the surveys (missing or highly variable estimates of weights-at-age). In the case of Gulf of Maine Atlantic cod, the arguments for a survey-based LW relationship are valid (large fraction of catches not from commercial landings and use of catch weights to estimate stock weights). Currently in the Northeast Region, fishery surveys are the only source of individual length-weight sampling.

Since 1992 the NEFSC bottom trawl surveys have used digital scales to record individual fish lengths. Using these data, updated survey-based length weight equations were compared to the existing length weight equation. Both seasonal (spring/fall) and annual updates were evaluated. First, to address concerns that Gulf of Maine cod condition have changed over time, the 1992-2010 time series was divided into roughly five year blocks and the relationships from each of the blocks examined (Fig. A.2). The relationships were nearly identical for both spring and fall seasons for all but one block (1996-2000). The 1996-2000 periods suffered from low sampling in both seasons and it was believed that these differences were more an artifact of sampling variability rather than a biological difference. Overall, the results suggested temporal stability of the seasonal LW relationships and indicated that cod condition has been constant, at least within the 1992 to 2010 period examined. Given this stability, the 1992-2010 data were aggregated to estimate updated spring, fall and annual relationships (Equations 2-4). These were then compared to the existing LW relationship (Fig. A.3). The updated relationships were statistically significant from one another as evidenced by the non-overlap of the 95% confidence intervals. All three

updated LW relationships tended to estimate heavier fish at length than the existing length weight equation.

$$\begin{aligned} (2) \quad & W = 0.000004714L^{3.1741} \quad (\text{Spring}) \\ (3) \quad & W = 0.000006178L^{3.1322} \quad (\text{Fall}) \\ (4) \quad & W = 0.000005132L^{3.1625} \quad (\text{Annual}) \end{aligned}$$

Based on these results a decision was made to use the revised LW relationships in the SAW 53 assessment update. Application of these LW equations back to the start of the assessment time period in 1982 requires an assumption that the stationarity observed in cod condition between 1992 and 2010 persisted back in time.

Growth and maturity

Atlantic cod in the Gulf of Maine and Georges Bank reach a maximum size around 130 cm (\approx 25 kg). Cod in the Gulf of Maine tend to grow slower than on Georges Bank (Fig. A.4). For the SAW 53 assessment update, von Bertalanffy growth parameters were re-estimated using NEFSC survey data from 1970 to 2011 (Equations 5 and 6). A summary of the number of ages included in the analysis are presented in Table A.4. Given the sparseness of the sampling of older ages, the L_∞ may be poorly estimated. Generally, the differences in growth parameters lend support to the treatment of Gulf of Maine and Georges Bank as separate stocks. These results are consistent with that of previous research on the topic (Penttila and Gifford 1976, Begg et al. 1999).

$$\begin{aligned} (5) \quad & L = 142.6 \cdot (1 - e^{-0.1261(t-0.1303)}) \quad (\text{Spring}) \\ (6) \quad & L = 162.4 \cdot (1 - e^{-0.1034(t-0.8103)}) \quad (\text{Fall}) \end{aligned}$$

Examination of monthly trends in the mean length of Gulf of Maine cod landed in the commercial fishery suggests that the majority of somatic growth occurs between March and December, with little growth occurring January through February (Fig. A.5). Examination of mean catch weights-at-age suggests that fish size-at-age may have declined in recent years, particularly at older ages (ages 5⁺; Fig. A.6). The declines are less evident in survey data (Fig. A.7), with many of the ages showing increases in the most recent two to three years. Generally, both current catch and survey weights-at-age are below those observed in the early 2000-period.

A logistic regression method (O'Brien et al. 1993) was used to fit maturity-at-age from the NEFSC spring survey data. In an attempt to smooth the noise in the data and increase sample sizes for those years with low sampling (Table A.5) a 3-year centered moving average was applied (Fig. A.8). The use of a 3-year moving average as opposed to some other time interval was based in part on the precedence of the GARM III assessment and also due to the fact that the 3-year average was sufficient to increase the sample size so that ogives could be estimated for years with few observations. The Northern Demersal Data Working Group (NDDWG) examined the 3-year moving average, and determined that the estimated A_{50} (age at which 50% of fish are mature) varied about the time series average A_{50} , but without any persistent trends.

The number of distinct stations from which fish were sampled for maturity was compared among years, to determine if differences in sampling protocol could explain the two high A_{50} estimates at the beginning of the time series. The age sampling design from 1970 to 1990 was based on achieving a sampling target number per watch; since 1991, the design has been to sample a target number per tow. The number of distinct stations was variable through time, but nothing indicated that sampling was more clustered earlier in the data compared to recent years. In fact, the number of stations sampled in the 1970s was higher than the middle of the time series, probably because abundance was so low in the 1990s that sample sizes

suffered in general. As the length of a survey watch has been 6 hours for most of the time series, it is likely that the protocol to target sample sizes by watch still managed to spread out the stations sampled. An alternative analysis was suggested to fit models that tested for year effects in the slope, the intercept, or both. These analyses encountered the same problem with small sample size in some years, leading to infeasible solutions in certain years. Because no persistent trends were detected, and sampling protocol did not appear to have produced non-representative measurements, the NDDWG decided to use a single time series average maturity ogive estimated from data in years 1970-2011 (Fig. A.9). The time series $A_{50\%}$ for male cod was 2.86 and 2.67 for females.

Natural mortality (M)

Previous assessments of Gulf of Maine cod have assumed a constant, age-invariant rate of instantaneous natural mortality (M) or 0.2 (NEFSC 2008, Mayo et al. 2009). The NDDWG evaluated the sufficiency of this assumption through life history analyses of natural mortality. Hoenig (1983) demonstrated that natural mortality can be estimated as a function of the maximum observed age (t_{max}) in a population (*ibid*; Equation 7). Depending on whether the maximum age observed from the surveys ($t_{max}=17$) or the maximum age observed in the fishery ($t_{max}=15$) is used, this approach yields estimates of $M = 0.246$ or 0.279 . This approach was further refined by Hewitt and Hoenig (2005; Equation 8), though the revised approach yields similar results of $M = 0.248$ or 0.281 . Because the Gulf of Maine cod stock has been heavily exploited for most of its recent history, and age samples are only available from the 1970s, M values in the range of 0.246 to 0.281 estimated from maximum age likely overestimate the true M .

An alternative approach relies on the gonadosomatic index (GSI) which is the ratio of gonad weight to somatic weight (Gunderson 1997). The general premise is that M is positively correlated with reproductive effort (*ibid*; Equation 9), more specifically, female reproductive effort. Estimates of GSI were not readily available for Gulf of Maine cod; however using a GSI value of 0.117 reported for the adjacent Georges Bank cod (McIntyre and Hutchings 2003) yields an M estimate of 0.209. Pauly (1980) first showed that M is proportional to the von Bertalanffy growth parameter, K . Using a variant of the relationship (Jensen 1996; Equation 10) and an estimate of $g=1.598$ (Gunderson et al. 2003) provides estimates of $M = 0.165$ or 0.201 depending on whether the K value is taken from the growth parameters estimated from the fall or spring surveys respectively.

$$(7) \quad \ln(Z) = a + b \cdot \ln(t_{max})$$

$$(8) \quad M = 4.22/t_{max}$$

$$(9) \quad M = 1.79 \cdot GSI$$

$$(10) \quad M = gK$$

where:

- Z is total mortality,
- $a = 1.46$,
- $b = -1.01$,
- t_{max} is the maximum observed age in a population,
- M is natural mortality,
- GSI is the gonadosomatic index,
- $g = 1.598$ (after Gunderson et al. 2003),
- K is the von Bertalanffy growth parameter

From this the meta-analysis of life history-based estimates the working group decided that the evidence

available suggested that 0.2 was reasonable. As in all previous assessments for this stock, natural mortality will be assumed to be 0.2 for this assessment for all years. The lack of observed change in condition, as evidenced by a constant LW equation, does not support a hypothesis for a shift in life history parameters.

The NDDWG did discuss the possible impacts of seal predation on assumptions of natural mortality. There is a general presumption that seal populations have been increasing in the region over the past twenty years, though no definitive estimates exist to evaluate the trends or relative scale of a population increase. It is possible that increases in the seal population could lead to increased cod predation which could suggest that M should be temporally increasing in the more recent time period. While these concerns were noted, there is no empirical basis to evaluate the current size of the seal populations and trends over the last thirty years, nor are there estimates of cod consumption of cod and how rates may have varied over time. Additionally, while seals are known to prey on cod, they are generalist feeders and the importance of cod in the diet of Gulf of Maine grey seals is unknown. There is limited information that suggests that cod represent only a minor component of harbor seal diet along the Maine coast (Wood 2001).

TOR A.1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data. Evaluate available information on discard mortality and, if appropriate, update mortality rates applied to discard components of the catch.

Overview

In the recent period (1982 to present) total catch has ranged from 22.3 thousand metric tons (mt) to 3.8 thousand mt (Table A.6, Fig. A10). Prior to 1999, commercial landings constituted 70-80% of the total catch, but since 1999 they have constituted only about 40-60% of the total catch (Fig. A.11). There were three primary reasons for this shift: (1) significant restrictions on commercial landings leading to (2) an increase in commercial discards, and (3) increased contribution from the recreational fishery.

Beginning in 1999, commercial discards became a significant component of the catch, accounting for greater than 30% of the overall catch (Fig. A.11). Notable increases in commercial discards were primarily the result of restrictive trip limits between 1998 and 2000 (Table A.3). Trip limits were gradually relaxed from 2000 through 2004 resulting in an overall decrease in the contribution of commercial discards to the overall catch.

Recreational landings peaked in 1987, but generally, recreational landings prior to 1999 constituted approximately 15% of the overall catch, whereas they accounted for, on average, about 20% from 1999 through 2010. Recreational discards became an increasingly important component of the overall Gulf of Maine cod catch as the minimum retention size of cod was progressively increased from 15 in. in 1982 to the current size limit of 24 in., which has been in effect since 2006.

Commercial landings

In 1982, the United Nations Convention on the Law of the Sea (UNCLOS) defined a countries exclusive economic zone (EEZ) as a zone extending up to 200 nautical miles from a nation's coast. The EEZ defines the region where each country has sovereign rights to marine resources including fisheries. The geographic proximity of the US and Canada in the Gulf of Maine and Georges Bank Regions results in an overlap of each nation's EEZ. Given the importance of these areas with respect to resource extraction (among other reasons), the US and Canada both submitted cases to the International Court of Justice at

The Hague, Netherlands seeking clarification. The Court issued a final ruling on October 12, 1984 formally delineating the US and Canadian EEZ. Hereafter, this demarcation line became informally known as the “Hague Line”.

Within the Gulf of Maine the US EEZ splits statistical areas 464, 465 and 467 (Fig. A.1). Prior to Hague line implementation, landings of cod in US ports from these statistical areas could have been either from the Gulf of Maine or Scotian Shelf stocks. Current management of Gulf of Maine cod includes catch from these areas against the fisheries ACLs. Previous assessments have not included these catches. While landings from these statistical areas have been low since 1985, accounting for less than two percent of the total Gulf of Maine landings (Fig. A.12), the NDDWG concluded it was important to include these landings in the current assessment to maintain consistency with current ACL monitoring. No attempt was made to adjust landings prior to 1985.

Since 1964 when modern catch statistics began, commercial landings of Gulf of Maine cod have ranged from 1.4 thousand mt to nearly 18 thousand mt (Table A.6). Landings statistics for area 5 (Gulf of Maine and part of Georges Bank stocks) exist back to 1893 (e.g., Mayo et al. 2009). The methods used to apportion landings to individual stock complex are not well documented and generally, these stock landings are considered less certain. It is worth noting that the estimates of historical Gulf of Maine cod landings reported in past assessment documents are of similar magnitude as landings between 1964 and 2010. Total species landings are derived from the weighout reports of commercial seafood dealers and these data are generally considered a census of total landings. A secondary source is required to apportion out the species landings to statistical area (stock) and assign basic information on fishing effort (e.g., gear and mesh). Prior to 1994, the partitioning of stocks from total cod landings was accomplished, in part, through a port-interview process conducted by port agents working for the National Marine Fisheries Service (NMFS).

In 1994, with the requirement of vessel-reported VTRs, the port interview process stopped and the area and effort information had to be inferred directly from the VTRs. Currently, a standardized procedure is used to assign area and effort from VTRs to dealer-reported landings from 1994 onward (Wigley et al. 2008). The product from this process is stored in the NEFSC allocation (AA) database tables. Landings are matched to VTRs in a hierarchical manner, with landings matched at the top tier (level A, direct matching) having a higher confidence than those matched at the lower tiers. The matching rates have improved over time with approximately 80% of Gulf of Maine cod landings being matched at the highest level since 2004 (Fig. A.13). Interestingly, there is a seasonal component to the matching success, with generally poor matching success around the month of May (Fig. A.14). This phenomenon has not been fully explained, but does coincide with the start of the groundfish fishing year and annual renewal of vessel permits. The overall precision associated with this process, in terms of a CV is estimated at less than 0.1 (Table A.7).

An additional area of uncertainty with stock landings stems from the mis-reporting and/or under reporting of statistical areas on VTRs. Federal regulations require that a separate VTR logbook sheet be filled out for each statistical area or gear/mesh fished. Vessels fishing in multiple statistical areas frequently under-report the number of statistical areas fished (Palmer and Wigley 2007, 2009 and 2011). The impacts of this misreporting on Gulf of Maine landings estimates are thought to be small. Between 2004 and 2008, the errors are estimated to have only resulted in small (<5%) underestimates of total stock landings, with the impacts decreasing over time (<1% in 2007 and 2008; Palmer and Wigley 2011).

For some species, there may be a component of the catch that does not get reported by seafood dealers. In the case of Gulf of Maine cod, fish retained by the crew for home consumption represent the largest likely fraction of landings that would not be reported by seafood dealers. Estimates of home consumption can be derived from VTRs, but these estimates probably represent underestimates of total home consumption

landings due to incomplete reporting. From 1994 to 2010, home consumption landings are estimated at < 0.3% of total commercial landings (Table A.8). Even if these represent underestimates, it is unlikely that home consumption landings represent a significant source of fishery removals. Given this, home consumption estimates were not included in total estimates of commercial landings.

The commercial fishery is primarily conducted by vessels fishing trawl and gillnet gear with gillnet gear having become progressively more important over time (Fig. A.15). Current landings by trawl and gillnet gear are about equal and account for nearly 95% of the total landings. Landings by longline and handline (jig) are minor. There is a seasonal component to fleet activity in the Gulf of Maine whereby gillnet landings drop in the spring months (March through June) when parts of the western Gulf of Maine are inaccessible due to rolling closures. Larger trawl vessels which have the capacity to fish further off shore, to the east of the rolling closures, dominate the landings during the spring months (Fig. A.16).

The ports of Gloucester and Portland have historically been the primary offload ports of Gulf of Maine cod (Fig. A.17). Portland landings have declined over the last twenty years and Gloucester now accounts for over 60% of total commercial landings. The impacts of the rolling closures in the western Gulf of Maine impacts port landing patterns in a manner similar to their impact on the gear trends. Landings in Gloucester drop off during the months of April and May when the nearshore waters are closed to groundfishing (Fig. A.18). During these months cod are primarily landed in ports along the Maine coast. The rolling closures cycle clockwise around the western Gulf of Maine, and by June, when the rolling closures are off the coast of Maine, Gloucester again becomes the dominant port for Gulf of Maine cod landings.

The patterns for landings by statistical area are nearly identical to the port trends. Over the last twenty years landings have become increasingly concentrated in statistical area 514 which is the statistical area in closest proximity to Gloucester (Fig. A.19). Landings from statistical areas to the north and east have declined. Currently, statistical area 514 accounts for >70% of total stock landings. The rolling closures have similar impacts on the statistical area landing patterns (Fig. A.21). The spatial aggregation of the fishery in the western Gulf of Maine over the past twenty years is also evident in observer data (Fig. A.20). It is not fully understood whether the aggregation of the fishery in the western Gulf of Maine has been driven by regulations, stock availability/distribution, or some combination of the two.

Commercial landings of Gulf of Maine cod are classified by four primary market categories: scrod, market, large and unclassified. Other market categories exist such as snapper, whale and steaker, but these are considered variants of the scrod (snapper) and large (whale and steaker) market categories. Market sized fish typically dominate annual landings with scrod sized fish having become less common over time, possibly in response to increasing minimum retention sizes (Fig. A.22). Over the past five years, market cod have accounted for approximately 70% of the total landings (Fig. A.23).

The temporal landing patterns of Gulf of Maine cod has changed slightly over the past five years, likely in response to the major changes brought about by Amendment 16. From 2006 through 2009 the fishery was most active from May through March, with very little landings occurring during the months of March and April (Fig. A.24). Presumably, the low landings during these months were as result of a combination of limited availability of DAS and rolling closures. In 2010 landings were more constant over the course of the year. It is not exactly clear how the transition to a sector management scheme altered the landings in March and April 2010, but it is possible that vessels that were entering sectors in May 2010 sought to fully utilize any remaining DAS as its currency would be useless under a sector-based system.

Commercial landings: biosampling

Biological sampling (length and age) of Gulf of Maine cod prior to 1982 was poor (Table A.9). The sufficiency of biological sampling has always limited age-based assessments of Gulf of Maine cod to the period from 1982 onward. Prior to 1982 it was not uncommon for sampling to be absent across entire market categories, or even for an entire year. From 1982 to 1995 sampling was relatively constant at around approximately 30 to 60 samples per year. When sampling dropped off, it was typically sampling of the smaller (scrod) and larger (large) market categories that suffered. Beginning in 1996 there was a notable increase in overall sampling. The years 1998 to 2000 were exceptions to this trend and were marked by years of low landings, including the lowest level of commercial landings (i.e., 1999, 1407 mt).

Length sampling of the commercial landings has varied from 28.1 to 517.9 mt per 100 lengths (Table A.10). A sampling intensities less than 200 mt per 100 lengths has traditionally been considered an unofficial NAFO/ICNAF standard. Sampling intensity has generally increased over time and has exceeded the standard since 1996. Prior to 1982 length sampling was poor with sampling intensities exceeding 1000 mt per 100 lengths sampled. The sampling density (number of lengths per sample) has ranged from 3 to 345 lengths per sample with an average of 79 lengths per sample. In the earlier periods, while sampling intensity was lower than the current period, the density was generally higher. Part of the trend in declining sampling densities has come about from a relaxation of the requirement to collect the full number of desired lengths per sample. In the past, samplers would frequently not sample unless they could collect a full sample (typically 100 lengths, but has varied by market category over time). Given that age sampling is conducted at the same time as length sampling (but lower density), it is not surprising that the sampling of age structures (otoliths) has followed similar trends as lengths. From 1995 onward the metric tons per 100 ages have been less than 1000 mt with sampling in the last five years on the order of 100 mt per 100 ages (Table A.11).

Previous Gulf of Maine cod assessments have estimated numbers-at-age by aggregating lengths into 3 cm bins. The current assessment performed a complete update of the catch-at-age. In doing so, an attempt was made to use 1 cm intervals. This requires a greater degree of age imputation to manually fill in gaps in the age length key (ALK). An examination of the amount of imputation that would be required suggested that the level of imputation was not unacceptable (Table A.12). The majority of market/time blocks required no imputation and for those that did, generally the percentage of landings requiring imputation was less than 5%. ALK imputation was primarily restricted to the older ages; given the small numbers of the population in these ages combined with the plus group handling of older ages, the impacts of this imputation are likely negligible.

When estimating the number of fish landed-at-age, every attempt was made to maintain the market category/quarter sampling design. However, when the availability of lengths for a particular market/quarter block was low, either a semiannual or annual time block was used. A criteria of 100 lengths per block was applied to the commercial landings for use as an objective basis to decide when it was appropriate to bin across quarters. In situations where an annual time block was required, the annual LW relationship (Equation 4) was used to convert landings to numbers-at-age. Otherwise, the appropriate seasonal LW equation was applied (Equations 2 and 3). A summary of the amount of binning that was required is presented in Table A.12. Total numbers-at-age are presented in Table A.13. The bootstrapped generated CVs on the landings-at-age estimates are shown in Table A.14. CVs are generally less than 30% for those ages that make up the majority of the landings (Ages 3-6). Prior to 1984, the calculation of bootstrap CVs were not possible due to the inability to identify individual sampling events. There is considerable uncertainty in the estimates of landings-at-age among some of the older ages, particularly beyond age 9 where the average CV begins to exceed 40%. Overall, younger ages have become less prevalent in the commercial landings with increases in the minimum retention size (Fig. A.25). Older fish were less common in the landings back in the late 1990's, likely due to a truncated population age structure.

Changes in the methods used to estimate landings-at-age relative to GARM III included: revised LW equations, 1 cm length bins compared to 3 cm length bins and complete re-estimation of the landings-at-age time series. Given these changes the revised estimates were compared to the GARM III estimates. Overall the differences were small (<10%), with the revised landings-at-age tending to be lower than the GARM III estimates (Table A.15). This was expected given that the revised LW relationships estimated heavier fish at length. Large differences were observed at older ages, but these represent large changes of a small number of fish (see Table A.13). Estimates of weights-at-age from landings in the commercial fishery are presented in Table A.16.

Commercial discards

Gulf of Maine Atlantic cod are primarily discarded in the commercial fishery for three reasons: (1) fish are below the minimum retention size (too small), (2) fish are of poor quality, and (3) high grading of smaller or poor quality fish in situations where a limited amount of fish can be landed (e.g., under trip limits). Discarding of smaller/poor quality fish became increasingly important from 1999 onward when the trip limits became more restrictive. However, the primary reported reason for fish discards has been because the fish were too small (Fig. A.26). With increases to the commercial minimum retention sizes in 2002, discarding due to undersized fish accounts for approximately 70% of total fish discards. This finding is in contrast to the conclusions of the GARM III assessment that “...*presumed that cod of all sizes and ages are discarded without prejudice.*” The GARM III conclusion was based on an examination of the years 1998 to 2000 when trip limits were most restrictive; however, this conclusion does not hold for other periods. This distinction is important to consider when determining how best to estimate the discards-at-age. Given that the majority of discards are of fish that are below minimum retention size, the method used in GARM III to account for discards in the catch-at-age was inappropriate and lead to an underestimation in the fishing mortality on younger fish and an overestimation in older fish.

Direct sampling of the commercial fishery for discards has been conducted by fisheries observers since 1989. Of the Gulf of Maine cod observed discarded by fishery observers, the following gear types account for greater than 99% of the total observed discards: benthic longline, small mesh (<5.5”) otter trawl, large mesh (≥ 5.5 ”) otter trawl, shrimp trawl, and large mesh (5.5”-7.99”) and extra large mesh (≥ 8.0 ”) sink gillnet gear (Table A.17). GARM III discard estimates included otter trawl, shrimp trawl and sink gillnet, but no distinction was made for the different mesh sizes.

The total number of trips observed of these gear types ranged from a low of 62 in 1997 to a current high of 2250 trips (Table A.18). The large increase in the number of observed trips in 2010 was due to the additional contribution of ASMs that were required for the groundfish fishery under Amendment 16. ASM coverage averaged approximately 25% of total groundfish trips whereas regular observer coverage (NEFOP) averaged about 7% (M. Palmer, NEFSC, unpublished data). A comparison of the estimated discard rates between ASM and NEFOP observers was undertaken in SARC 52 (Wigley et al. 2011) and showed no statistical difference for the majority of gears and quarters examined. Generally, the Gulf of Maine cod ASM discard rates were statistically indistinguishable from the NEFOP discard rates as evidenced by the fact that the 95% confidence intervals of the difference between estimates include zero (Fig. A.27).

While handline gear does not constitute a large fraction of observed discards, this is partly because this gear type is not frequently observed owing to the small size of these vessels and regulatory exemptions from observer coverage for some handline permit categories. Regardless, it is known that discarding by this gear does occur and it is accounted for in the in-season groundfish monitoring programs. Attempts were made to estimate discards for this gear type, but the NDDWG concluded that the proportion of

observed trips for handline was too low to give confidence in the derived estimates of discard amount (maximum number of observed trips in any year was 9).

The previous GARM III assessment used a variant of the Standardized Bycatch Reporting Method (SBRM; Wigley et al. 2007). The ratio method applied for Gulf of Maine cod was similar to that used for other groundfish stocks except that rather than using the amount of all catch retained in the ratio denominator, the amount of retained cod was used. This decision was made on the basis that it was thought that the discard estimates provided by the $d_{\text{cod}}/k_{\text{cod}}$ ratio better represented the high discarding that likely occurred under the severe trip limits that existed in 1999 (30-200 lb/day). It is unknown whether this is true, but the methodology used in GARM III is inconsistent with that used for other groundfish stocks as well as the current in-season groundfish monitoring programs, which also utilizes a $d_{\text{cod}}/k_{\text{all}}$ ratio. To resolve this discrepancy, the SAW 53 Gulf of Maine cod assessment utilizes the $d_{\text{cod}}/k_{\text{all}}$ ratio estimator.

Prior to arriving at the final estimates of commercial discards, several different temporal stratification schemes were evaluated with respect to their impact on total discards and relative precision. Quarterly, semi-annual and annual stratifications were explored. All achieved nearly identical results with respect to total discards, with the annual stratification having slightly lower CVs, though generally all CVs were below the informal target of 30% (Fig. A.28). Given the lack of sensitivity to choice of temporal stratification, a decision was made to use a semi-annual stratification owing to its ease of use from an operational perspective when estimating discards-at-age.

Final estimates of discards ranged from under 100 mt in 1998 to a high of 2,198.2 mt in 1990 (Table A.19). While there are exceptions, large-mesh otter trawl is the major source of cod discards. Shrimp trawl discards were an important component of cod discards in the early years, but the required use of a Nordmore grate for the Gulf of Maine shrimp fishery beginning in 1992 was highly effective at reducing cod discards. The resulting CVs on the discard estimates are variable on a gear-specific basis. At the aggregate level, CVs of total discards are typically less than 30% and below 20% over the last four years (Table A.20). Comparison of the updated discard estimates to those of GARM III shows close agreement between the two, with both showing similar trends and scales and having overlapping 95% confidence intervals in all of the years (Fig. A.29). The largest difference between the two estimates occurs in 1999.

As a means of evaluating the accuracy of the discard estimation procedure, a check was conducted to attempt to estimate total landings using the same methodology used to estimate discards. Instead of estimating a $d_{\text{cod}}/k_{\text{all}}$ ratio, a $k_{\text{cod}}/k_{\text{all}}$ ratio is estimated. When compared to the total cod landings, the results show close agreement with respect to scale and trends lending support to the accuracy of the discard estimation procedure (Fig. A.30).

Commercial discards: biosampling

Observers collect length and age information from the discarded fraction of the catch (as well as on the retained catch); however, only length samples are currently available. ALKs were created using both commercial landings and NEFSC survey ALK corresponding to the appropriate season (spring/fall). Length sampling extends back to 1989 and has generally been quite good with sampling intensities for most years less than 100 mt of discards per 100 lengths (Table A.21). The length distributions by gear are shown in Figure A.31 on an aggregate basis and by year in Figures A.32 through A.38. Increases in the minimum fish size as well as the impacts of trip limits leading to the discarding of larger sized fish are evident in the time series plots. Generally, shrimp trawl captures the smallest fish with the sink gillnet gear having a much broader distribution of lengths including a large proportion of lengths in excess of the minimum size. The reasoning for the expanded length distribution in the gillnet fishery is largely due to

the prevalence of poor quality discards in this fishery (e.g., damage due to seals, dogfish or sand fleas that is occurring during the gear soak).

When estimating discards at length, attempts were made to maintain the separate semi-annual estimates so that the most appropriate seasonal LW equation could be applied. For some years and gear types this was not possible owing to limited sampling. A criterion of 50 lengths per block was applied to the commercial landings to provide an objective basis to decide when it was appropriate to bin across semesters and or gear types. Binning across gear types was only done between the two gillnet gears owing to the similarities of their length frequency distributions.

Commercial discard hindcasting: pre-1989

Direct observations of discards by fishery observers only exist from 1989 to present. The model formulations used in past assessments have started in 1982 owing to the availability of information on the age structure of the commercial landing. Previous assessments have made no attempts to hindcast discards back to 1982. In this assessment update a survey filter method described in Palmer et al. (2008) and previously applied to groundfish stocks in the Northeast Region (e.g., Mayo et al. 1992, O'Brien and Esteves 2001) has been used to extend discard estimates back to 1982. Discards were only hindcasted for the three primary discard gear types during this period: large mesh otter trawl, shrimp trawl and large mesh sink gillnet.

The survey filter method requires information on survey numbers at length (N_i), estimates of gear selectivity at length (m_i), a scaling factor (q) and an estimate of total fishery effort (f). Assuming these are available, discard-at-length can be estimated using the following equations:

If:

$$(11.a) \quad C_i/f = q \cdot (N_i \cdot m_i), \text{ then}$$

$$(11.b) \quad C_i = (q \cdot f) \cdot (N_i \cdot m_i) \text{ as above.}$$

If:

$$(11.c) \quad K_i = C_i \cdot s_i, \text{ and}$$

$$(11.d) \quad D_i = C_i \cdot (1-s_i), \text{ then}$$

$$(11.e) \quad D_i = (q \cdot f) \cdot (N_i \cdot m_i) \cdot (1-s_i), \text{ and}$$

$$(11.f) \quad D_i/f = q \cdot [N_i \cdot m_i \cdot (1-s_i)]$$

where:

- C_i is the catch retained by a given commercial mesh at length i ,
- N_i is the abundance of fish in the survey at length i ,
- m_i is the proportion of the available population retained by a given mesh at length i ,
- s_i is the proportion of the retained catch kept at length i ,
- K_i is the kept portion of the catch at length i , and
- D_i is the discarded portion of the catch at length i .
- f is some estimate of total fishing effort.

If it is assumed that the fish discarded pre-1989 were all less than the minimum size, the above equation can be simplified by setting s_i to 0. This assumption is likely valid for large mesh otter trawl and shrimp trawl, but may not hold for large mesh sink gillnet gear (Fig. A.39). The impacts of this assumption on the estimation of proportion at age is evaluated later. Using a set of years when management was similar to the hindcast years, gear selectivity at length (m_i), and the appropriate scaling factor (q) can be estimated and the accuracy of the overall method can be evaluated. The years 1989 to 1993 were used for method

development and evaluation of trawl and gillnet gear and the years 1989 to 1991 for shrimp trawl due to the major changes in the shrimp trawl discard patterns that occurred in 1992 (i.e., Nordmore grate).

Using Pope’s (1966) ‘alternate tow’ approach, the ratios of observed proportion-at-length discarded from the fishery to the proportion-at-length present in the survey are generated (e.g., Fig. A.40). Equation 12 (Wileman et al. 1996) is then fit to the aggregate ratios (across all years) to generate selectivity ogives (Fig. A.41). The fits to the shrimp trawl were poor, and given the small size distribution of cod discarded in the shrimp trawl fishery, an assumption was made that the selectivity of the shrimp trawl was identical to that of the NEFSC bottom trawl survey. The mesh sizes of the shrimp fishery during this period (1.75”/4.45 cm) were not all together dissimilar from those of the survey gear (11.5 cm codend with a 1.27 cm liner). Comparison of the proportions at length between the survey-filter method and the direct observations recorded by observers shows reasonably close agreement in the length distributions across years for large mesh otter trawl and shrimp trawl gears (Figs. A.42 and A.43). There was less agreement among the length frequency distributions for sink gillnet gear, with only two of the five years showing close agreement (Fig. A.44). Conversion of the number-at-length to numbers-at-age using a combined spring and fall NEFSC survey ALK showed even closer agreement between the survey-filter approach and the direct estimates (Fig. A.45 – A.47). This suggests that while the assumptions of the survey filter method may not accurately reflect the length distribution of gillnet discards, the overall impacts on the age distribution are mitigated.

$$(12) \quad r(l) = \left[\frac{\exp(a+bl)}{1+\exp(a+bl)} \right]$$

By regressing the ratio of observed discards-at-length to the total fishing effort (K_{all} was used similar to the contemporary discard estimates) on the ratio of selectivity-adjusted survey numbers-at-length, the gear-specific scaling factor (q) can be estimated as the slope of the regression line (Equation 11.f, Fig. A.48). In performing these regressions it was noted that the relationship of the two ratios was different in 1990 relative to other years. It’s possible that this reflects some effects of the 1987 year class moving into the fishery. Based on the GARM III assessment, the 1987 year class was the largest year class observed during the assessment time series (Mayo et al. 2009).

Total discards estimated using the survey filter approach reflected the relative trends and scales from the direct estimates (Table A.22). The large mesh gillnet estimates were underestimated relative to the direct estimates, possibly due to the assumption of smaller fish in the survey filter method. In 1990 the survey filter underestimated across all gear types, possibly due to poor fit of q in that year as described above.

The NDDWG considered an alternative metric to the survey-filter hindcast: use an average of the d_{cod}/k_{all} ratio from years 1989-1993 and raise it by the annual K_{all} in years 1982-1988. The NDDWG discussed whether the average d_{cod}/k_{all} ratio could be biased from including the 1990 value in the estimate, which may have been much higher owing to the anomalously large 1987 year class. As an intermediate approach, the NDDWG suggested a third calculation of hindcasted discards using the average d_{cod}/k_{all} ratio for years 1989 to 1993, *excluding 1990* (Fig. A.49). The NDDWG discussed the appropriateness of hindcasting, and whether assuming that discards are zero is better than making assumptions to derive estimated amounts. Ultimately, the NDDWG concluded that the true discards are likely between zero and the d_{cod}/k_{all} ratio estimates that included the 1990 value (which provides a likely upper bound). The final approach applied the average d_{cod}/k_{all} ratio for years 1989 to 1993, *excluding 1990* as the basis for the amount of hindcasted annual discards with the proportion at age determined using the survey filter method. Commercial discards-at-age and weights-at-age are presented in Tables A.23 and A.24 respectively. Bubble plots of discards-at-age over time are shown in Fig. A.50.

Recreational landings

Estimates of the recreational Gulf of Maine cod catch were obtained from the Marine Recreational Fishery Statistics Survey (MRFSS). This survey has been conducted annually since 1979. MRFSS breaks the total catch into three components: directly observed landings (A), unobserved landings (B1), and unobserved discards (B2). Similar to the treatment of MRFSS data in GARM III, recreational catches were partitioned into Gulf of Maine and Georges Bank stocks using updated MRFSS data and site register lists. Recreational catches attributed to site register lists in Maine and New Hampshire as well as Massachusetts landings from Essex, Suffolk, and Plymouth counties are allocated to the Gulf of Maine stock. Landings from Barnstable County (Massachusetts) are split such that intercept sites bordering Cape Cod Bay are allocated to the Gulf of Maine stock and those on the east and south side of Cape Cod are allocated to the Georges Bank stock.

While MRFSS is the source for official recreational catch estimates, VTRs provide a useful source for understanding some of the finer spatial and temporal trends that cannot be easily determined from the MRFSS data. They also help inform the validity of the MRFSS sampling scheme and treatment of data. VTR data are only available for the federally permitted party (head boats) and charter modes. VTR data do not cover the private recreational fleet or party/charter vessels operating only within state waters. Federally permitted recreational vessels only represent from 14 to 69% of the total recreational catch in a given year (Table A.25), thus VTR-based estimates will underestimate the total recreational landings (Fig. A.51). The MRFSS program does not sample the New England region in Wave 1 (January/February); however, VTR data suggest that historically, very low recreational activity occurs in these months (Table A.26). Since May 1, 2006 the recreational fishery has been prohibited from possessing cod in the Gulf of Maine between November 1st and March 31st. This prohibition was extended to April 15th in 2009. MRFSS-based estimates of total catch by sampling wave show highly variable temporal patterns, but are generally consistent with VTR data, with waves 2-5 having the highest proportion of total annual catch (Table A.27). It may be important to note that an anomalously high proportion of the 2010 MRFSS catch was estimated in wave 2. Since wave 2 was only open to the recreational fishery beyond state waters for two weeks in 2010 it seems unlikely that wave 2 could be responsible for 50% of the total recreational catch. The majority of VTR-reported recreational landings come almost exclusively from statistical areas 513 – 515 (Table A.28). Based on the VTRs, there are virtually no landings of Gulf of Maine cod in ports south of Massachusetts (Table A.29). This finding supports the existing allocation scheme based on the site register lists that is used to partition MRFSS recreational catch into Gulf of Maine and Georges Bank stocks.

The MRFSS survey is a numbers based survey and conversion of MRFSS estimates to removals in terms of total weight can be accomplished in several ways. Total weight estimates typically provided by the MRFSS program convert numbers to weight using the average sampling weights by state and semester. In the earlier time periods, sampling was poor such that average MRFSS weights did not exist for all cells. This can lead to an underestimation of removals in terms of average weight (Method 1). Imputing the missing cells using the averages from other cells within the same year addresses the issue of missing cells (Method 2). The quality of the MRFSS weight sampling is unknown, though it is generally perceived that the quality of the length information is more reliable. Length sampling of recreational landings has improved over time, though the sampling intensity is not as good as that of the commercial fishery (Table A.30). An alternative method is to use the annual length frequency distributions (Fig. A.52) to generate numbers at length and then apply the annual LW equation to estimate total removals in terms of weight (Method 3). Because the majority of recreational catch occurs mid-way between the spring and fall NEFSC surveys, it was not appropriate to partition out catch into spring and fall components. Methods 2 and 3 achieve similar results in terms of total landings, Method 1 tends to underestimate total removals

early in the time series when sampling was sparse (Fig. A.53).

The SAW 53 assessment update will use Method 3 for all final estimates of catch removals. Total landings estimated in terms of weight track closely with the numbers-based estimates of landings (Fig. A.54). Since 1997, there has been a proportional increase in the weight-based estimates relative to numbers due to incremental increases in the recreational minimum retention size. The numbers-based estimates of recreational landings were converted to numbers-at-age using ALKs borrowed from the NEFSC survey which include age information collected from the inshore strata. The inclusion of the inshore strata provided a better spatial overlap with the recreational fishery compared to the use of just the offshore strata (Fig. A.55). Recreational landings-at-age show similar trends with respect to the impacts of increasing minimum retention sizes (Fig. A.56). Like the commercial landings, older ages are absent from the recreational landings throughout much of the 1990s.

Recreational discards

In previous Gulf of Maine cod assessments, recreational discards have been reported, but they have not been included in the catch-at-age. The primary reason was that there has historically been no length sampling of discarded component of the recreational fishery, and thus no information to convert the total recreational discard estimates (B2 catch) to estimates of discards-at-age. The largest fraction of discards is attributed to the party/charter mode in areas that are greater than 3 miles from shore and the private/rental mode, which has seen an increasing trend in the fraction taken more than 3 miles from shore (Table A.31). Beginning in 2005 direct sampling of cod discards from party boats began in the Gulf of Maine (i9 sampling; Table A.32). Sampling intensities have averaged approximately 200 mt of discards per 100 lengths sampled which is slightly higher relative to the length sampling of recreational landings during the same period.

With increases in the minimum recreational retention sizes, the contribution of recreational discards to total recreational catch has been increasing over time (Fig. A.57). Currently, recreational discards are approximately double the recreational landings in terms of numbers. Because of the increasing importance of recreational discards over time the NDDWG concluded it was worthwhile to attempt a hindcast of recreational discards using the available length frequency information and a variant of the survey filter method was used to hindcast commercial discards. Unlike commercial discards, estimates on the magnitude of recreational discards in terms of total numbers were already available from the MRFSS survey. The survey filter method was needed only to construct the length frequency distribution of the recreational discard catch back in time. Similar to commercial discards, the assumption was made that all discarding was done due to minimum retention sizes. This assumption appears to be valid for the recreational fishery, with almost no discarding of legal-sized fish occurring in the 2005 – 2010 period (Fig. A.58). Using the alternate-tow approach used for commercial discards, a gear selectivity ogive was constructed (Fig. A.59). Comparing the survey-filter length frequency distributions to the observed length frequency distributions showed close agreement (Fig. A.60). Applying the survey filter method back to 1981 (start of the length sampling of recreational landings) yielded the length distributions shown in Fig. A.61. The same NEFSC survey ALKs applied to the recreational landing was used for the recreational discards resulting in the discard-at-age patterns shown in Figure A.62.

A summary of recreational catch from 1981 to 2010 is presented in Table A.33. Recreational catch has ranged between 5.8 thousand mt and 0.6 thousand mt. The large increase in the 2010 catch should be noted for the reasons described previously. Because of the method used to apportion MRFSS cod estimates to stock areas, there are no direct estimates of precision available for recreational catches; however, the MRFSS-published estimates of percent standard error (PSE) provide some gauge as to the relative precision of the recreational catch estimates (Table A.34). Overall the general precision of these

estimates is about equal to the commercial discards. It is worth noting that despite the large Wave 2 catch in 2010, PSE values appear comparable to previous years.

Total cumulative recreational landings-at-age and landing weights-at-age are presented in Tables A.35 and A.36. Recreational discards-at-age and discard weights-at-age are presented in Table A.37 and A.38.

Discard mortality

The NDDWG reviewed a working paper (Palmer et al. 2011) which summarized findings from literature about the discard survival of Atlantic cod and other similar species. It must be emphasized that the NDDWG found this TOR very difficult to address. Discard mortality was evaluated for all gears for which discards were estimated in the updated SAW 52 assessment, with each gear being evaluated separately based on the gear-specific information available from the literature. Some members of the NDDWG argued that a presumption of discard mortalities less than 100% would ‘provide an incentive’ to influence handling the fish in such a way that mortality might actually be lowered. The majority of the working group disagreed with the rationale and considered these concerns external to an objective determination based solely on the scientific merits of each study.

While each study provided an estimate of survival, no single study could address every factor implicated in mortality. These factors include: temperature and seasonal effects, depth of capture, time of handling, type of handling, length of time on deck, short term and long term survival (one study estimated that only about 50% of mortality occurred in first few days—the length of most observation periods), impacts on growth due to reduced feeding ability, whether predator avoidance was compromised or predator exposure was increased at release time (birds, mammals, other fish predators), whether the field studies held fish on deck in tanks or in an aquarium or held in a cage at depth. It was noted that studies where fish were held in cages to evaluate survival could be biased either high or low. On the one hand, being held in a cage reduces exposure to predation, which could inflate estimates of survival. On the other hand, the cage could induce stress, damage to fish from contact with the cage, and even mortality due to cannibalism—all factors that could potentially increase mortality.

Each gear was evaluated with respect to available studies with survival estimates, what factors had been accounted for, what factors had not been accounted for, and whether it was possible to determine what conditions were likely to have existed for unobserved trips. The NDDWG concluded that it would not be possible to characterize the temperature/depth/season for all unobserved trips and therefore a single, annual discard mortality rate would be decided on. The working group was consistent in how it approached the evaluation of each gear, first by reviewing the available studies, discussing what factors were, and were not controlled for, and whether the estimates in the literature were likely to be biased high or low. In the end, the working group did agree that the published studies probably overestimated survival, although it was difficult to characterize the extent of that bias. The discard mortality rates to be used in SARC53 for Gulf of Maine cod are 100% for all gears. Sensitivity analyses at lower discard mortality rates were not explicitly explored. Building the bridge from the previous assessment to an updated VPA assessment will constitute a *de facto* evaluation of including discards with 100% mortality since many of the gears/fleets did not have discards estimated in the previous assessment (e.g., commercial longline and recreational).

Total catch-at-age and mean weight-at-age

Estimates of total catch-at-age were determined by summing the numbers-at-age across all of the catch components: commercial landings, commercial discards, recreational landings and recreational discards

(Table A.39). The age structure of fishery catch was truncated in the early 1990s relative to that observed in the 1980s. The truncation persisted through 2000 with age 9 and older fish beginning to reappear in the fishery in greater numbers beginning in 2001. These older age classes persisted through 2007 and have become less common in the fishery catches over the most recent three years.

Mean catch weights-at-age were estimated by using a numbers weighted average of the individual catch component's mean weights-at-age (Table A.40). This is a major difference relative to previous Gulf of Maine cod assessments which have estimated catch weights using only the landed fraction of the catch. The net impact is that previous assessments likely overestimated the true catch weights by not including the smaller fish-at-age in the estimation of catch weights-at-age. The relative differences between the weights used in the current assessment and those used in GARM III are presented in Table A.41. The largest differences in weights occur at the younger ages classes (i.e., those ages most likely to be in the discarded fraction of the catch). From age 5 and older, the relative differences are generally less than ten percent.

Mean weights were generally greater than average during the mid- to late-1990s, with below average mean weights being observed across many age classes during the early- to mid-2000s. Mean weights of the older age classes (\geq age 5) appear to still be below average, but an increase has been observed in the younger ages (Fig. A.6).

Sampling of older age fish in the trawl surveys has historically been low, and use of survey-based weights-at-age to estimate January 1 and spawning stock weights for use as model inputs would require extensive imputation. For this reason, catch weights-at-age were used to estimate January 1 and spawning stock weights. Prior to estimation of stock/spawning stock weights, minor imputation of the catch weights at-age were required to fill in gaps in the older age classes (primarily ages 10 and 11, Table A.40). An examination of possible approaches (e.g., moving averages or time series averages) showed that imputation using a 5-year centered moving average would be most appropriate.

January 1 and spawning stock weights were estimated from catch weights using a method described in Rivard (1980, 1982). March 1 is the assumed spawning event in the base model. Given that there is little somatic growth between January 1 and the assumed start of the major spawning period (April 1; Fig. A.5), spawning stock weights were set equal to January 1 weights-at-age. The Rivard method adjusts the catch mean weights-at-age, which are generally presumed to represent mid-year weights, back to January 1. Mean weights at the beginning of the year for a given age class are calculated as the geometric mean of the weight in the same year and of the same cohort in the previous year. No adjustments are made for the plus group calculation. Calculations for the initial and final years and ages are described in Rivard (1980, 1982). January 1/spawning stock weights are shown in Table A.42.

TOR A.2. Present the survey data being used in the assessment (e.g., indices of abundance, recruitment, state surveys, age length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

NEFSC bottom trawl survey

The NEFSC spring and fall bottom trawl surveys began in 1968 and 1963 respectively, providing a long time series of fishery independent indices. All previous Gulf of Maine cod assessments used only the offshore survey strata (Fig. A.63). During the NDDWG meeting, it was suggested that the indices be evaluated with and without the inshore strata. The current approach to generating NEFSC indices ignores the inshore strata because they are not consistently sampled. Additionally, the Massachusetts Department

of Marine Fisheries (MADMF) survey covers the inshore areas and this survey has traditionally been included in the Gulf of Maine cod assessments. The impacts of including the inshore survey strata in the NEFSC survey indices was examined by the NDDWG and resulted in increased indices of age 0 through 2 fish. The overall trend in the age-specific indices of older fish was not markedly different with the inclusion of the inshore strata and there were several strata/year combinations with poor sampling. For this reason, the NDDWG decided to maintain the status quo and exclude the inshore strata from NEFSC indices.

A frequent criticism of the NEFSC bottom trawl surveys is that they do not cover the same areas where the commercial and recreational fisheries catch cod, and thus are ‘missing’ much of the cod that exists in the Gulf of Maine. A comparison of the NEFSC spring and fall survey catches to commercial (total observed cod catches by ten minute square) and recreational activity (total number of trips catching cod by ten minute square) show close agreement between survey and industry catches (Fig. A.64).

The NEFSC bottom trawl survey has utilized three different vessels and three different door configurations throughout the time series of the survey (Table A.43). In an effort to maintain a consistent survey time series, survey indices are converted to ‘Albatross IV/Polyvalent door’ equivalents using several different conversion factors (Table A.44). The largest change in the survey time series occurred in 2009 when the FSV Albatross IV was decommissioned and replaced by the FSV Henry B. Bigelow. This resulted in changes not only to the vessel and doors, but also to the overall trawl gear as well as the survey protocols (summarized in Table A.45). Calibration experiments to estimate survey differences were conducted in the spring and fall of 2008 (Brown 2009). The results of those experiments were peer reviewed by a panel of external (non-NMFS) experts and then summarized in Miller et al. (2010). These results provide annual calibration coefficients both in terms of abundance (numbers) and biomass (weight). Further work by Brooks et al. (2010) developed length-specific abundance calibration coefficients for Atlantic cod. This method uses a segmented regression model where a constant conversion factor is applied to fish ≤ 20 cm and ≥ 54 cm, and a constantly decreasing linear regression is fit to fish between 20 and 54 cm (Fig. A.65). A comparison of the converted and unconverted spring and fall survey indices is presented in Figure A.66.

During a pre-SARC 53 meeting with the fishing industry (held August 16, 2011 in Gloucester, MA), industry expressed concern with the 24-hour operation of the survey. There was a sense that there were differences in the relative catchability of cod between daytime and nighttime hours. These observations are supported in the scientific literature (e.g., Beamish 1966), though the nature of off bottom movements is highly variable. An analysis was pursued as to whether there were appreciable differences in survey catchability between daytime and nighttime tows. The results showed that generally catchability was slightly higher in the daytime tows. However, the trends between day and night tows were similar, and in most years the day/night survey indices fell within the 80% CI of the aggregate index (Fig. A.67). Because of the similarity in the trends it is appropriate to use both day and night tows to calculate indices for the assessment. Splitting by day and night would result in reduced tows and lost strata (Table A.46), which would increase the likelihood that survey indices could be influenced by a single large tow in any year.

Aggregate survey indices over time are presented in Table A.47 along with the corresponding CVs. Generally survey indices were higher in the earlier time periods, reaching lows in the mid-1990s. There has been a slight increase in survey indices relative to the mid-1990, but survey indices have remained constant over the past decade (Fig. A.68). It is worth noting that some of the highest survey indices are associated with relatively high CVs/confidence intervals. This is an important consideration in determining how to interpret survey indices; i.e., do increases in survey indices represent true increases in the relative size of the resource, or are the indices being driven by a few influential tows that are not indicative of the resource abundance/biomass? Indices-at-age for both the spring and fall surveys are

presented in Tables A.48 and A.49 and Figures A.69 and A.70. Similar to the trends observed in the commercial and recreational fisheries, there were few older fish present in the survey catch-at-age throughout most of the 1990s.

The NDDWG examined spatial trends in the NEFSC survey catches over time to see if these could inform the understanding of small-scale distributions of cod (TOR A.5). Plots of the spring and fall survey catches (number/tow) show a general decline in the overall abundance from the 1970s through the 1990s. There is a notable increase evident in the 2000-2010 period, but the increase appears to be restricted to the western Gulf of Maine (Fig. A.71 and A.72). Moderate survey catches occurred along the coast of Maine in the 1970s, but these have not been observed in the past twenty years. To further address the aspect of spatial aggregation, a time series of Gini indices were calculated following the techniques outlined in Wigley (1996). These results support the patterns shown in distribution plots and suggest that the resource has contracted into the western Gulf of Maine over the last twenty years (Fig. A.73). These patterns are similar to the spatial aggregation that has occurred in the commercial fishery.

MADMF bottom trawl survey

The MADMF has conducted research bottom trawl surveys during the spring and fall since 1978. The survey strata included in the MADMF survey primarily includes the nearshore habitat within Massachusetts state waters in the southwestern Gulf of Maine (Fig. A.74). The MADMF survey strata closely coincide with the NEFSC inshore survey strata occurring in Massachusetts state waters (Fig. A.75). Both surveys occur around the same time of the year, though the MADMF spring survey occurs about 20 days later in the spring and 45 days earlier in the fall relative to the NEFSC survey (Table A.50). Because the MADMF surveys are conducted in relatively shallow waters and are limited in their spatial extent, they do not provide an index of the total stock resource, but may provide some information on the younger age classes inhabiting the nearshore environment (i.e., a recruitment index). Additionally, given the limited spatial extent, the MADMF survey may be more susceptible to resource availability due to timing of onshore/offshore seasonal movements (i.e., process error). A complete description of the MADMF trawl survey is provided in King et al. (2010).

In constructing MADMF survey indices-at-age, ALK information was borrowed from the NEFSC inshore survey strata shown in Figure A.75. Given the similarities in the survey extent and timing, this approach was preferred over manual imputation (Table A.51). Aggregate survey indices and the corresponding CVs are presented in Table A.52 and Figure A.76. Abundance-at-age indices for the spring and fall surveys are presented in Tables A.53 and A.54 and Figures A.77 and A.78, respectively.

Maine – New Hampshire inshore trawl survey

The Maine – New Hampshire (MENH) inshore trawl survey has not been included in previous assessments, though previous assessment reviews have encouraged a thorough examination of the information available from this survey (GARM I, NEFSC 2002b). The MENH survey began in 2000 and has been conducted in the spring and fall annually in the nearshore waters of the Gulf of Maine (Fig. A.79; Sherman et al. 2005). The ten year time series of abundance and biomass indices do not exhibit strong interannual fluctuations (Fig. A.80). The spatial distribution of catches seems consistent with the patterns observed in the NEFSC surveys with the highest catches occurring in the southwestern Gulf of Maine off the coasts of Massachusetts and New Hampshire (Fig. A.81). There were some indications of high catches along the eastern Maine coast, though annual plots examined by the NDDWG showed that these catches occurred early in the time series and have not persisted over time. A cursory examination of length frequency distributions suggests that the spring survey captures primarily age 0 through 2 fish (<35

cm) with the fall survey capturing age 0 and 1 fish as well as juvenile fish less than 60 cm (Fig. A.82). The size frequencies seem to suggest that MENH captures the same age classes observed in MADMF survey.

The biggest impediment to inclusion of this survey is the absence of age information. While otoliths have been collected, they have not been aged. It would be easier to incorporate this survey into an assessment if ages were available, and the NDDWG wanted to encourage that this be pursued. Additionally, the NDDWG encouraged that reproductive information be evaluated for the early years where Downeast Maine stations were sampled to evaluate whether any of the fish were mature and whether there was evidence to suggest the presence of a spawning aggregation. In the meantime, because the length frequencies are similar to MADMF, the working group did not feel that any important signals were being excluded from the model because there are age specific indices from MADMF in the model. The MENH survey was not included in the SAW 53 assessment update of the Gulf of Maine cod.

MADMF Atlantic cod industry based survey

The MADMF Atlantic cod industry based survey (IBS) was conducted from November 2003 through March 2007 (Hoffman et al. 2006). The survey was primarily conducted during the months cod are believed to spawn in the southwestern Gulf of Maine (November through May). Given the short time series, the survey was not considered for inclusion as an assessment tuning index. The NDDWG did however examine results from the survey as they relate to spawning times which indicate that peak spawning in the southwestern Gulf of Maine occurs in the April to May time period.

LPUE index

Trends in commercial landings per unit effort (LPUE) have been used in previous Gulf of Maine cod stock assessments. The 1982-1993 age composition of the landings corresponding to the effort of an otter trawl sub-fleet (summarized in Mayo et al. 1994) has been used to calculate LPUE-at-age indices for ages 2 through 6 (Table A.55; Mayo et al. 2009). The time series has not been extended beyond 1994 due to uncertainties in VTR reported fishing effort since 1994, the impact of reductions in days at sea, rolling closures and trip limits. All of these issues would affect the comparability of LPUEs estimated from 1994 onward with the earlier time series. Additionally, these same issues would make standardization of a contemporary catch per unit effort (CPUE) index difficult.

There is high correlation between the LPUE-at-age indices and the NEFSC abundance-at-age indices, particularly among older ages (Table A.56). While the aggregate indices do not exhibit as high a degree of correlation, they do exhibit the same basic trends (Fig. A.83). Given that the LPUE index has been used in previous assessments and it is unknown how its removal could impact assessment results, the NDDWG suggested model sensitivity runs to assess the utility of including the LPUE index. If model results were insensitive to the index, the NDDWG concluded it would be appropriate to remove the index from the SAW 53 assessment update.

TOR A.3. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, catch recruitment and fishing mortality.

Update of the GARM III VPA model

There were major changes in the treatment of the underlying data for the SAW 53 assessment update relative to the data used in the GARM III assessment. The major changes include: updated LW relationships, re-estimated landings-at-age, and inclusion of commercial discards in the catch-at-age, extension of the commercial discards-at-age back to the beginning of the model time series (1982), estimation of recreational discards-at-age back to beginning of the model time series, new estimates of weights-at-age that reflect landings and discards, and a revised maturity ogive. Additionally, there are three more years of catch and survey information that needed to be incorporated into the model. To fully understand how these data changes impact the VPA update, a bridge as constructed to transition from the GARM III assessment to a fully updated assessment.

The GARM III assessment was conducted using the Adaptive Framework Virtual Population Analysis (ADAPT-VPA) model (NOAA Fisheries Toolbox ADAPT-VPA version 2.7, 2007). This version relied on Pope's approximation to solve the catch equation and only allowed for the 'backward' calculation of the plus group. The most recent version of the ADAPT-VPA software (version 3.1.1, 2011) solves the catch equation exactly and supports both the 'backward' calculation of the plus group and the 'combined' calculation advocated by Butterworth and Rademeyer (2008a). In addition to the data changes, these model changes must also be accounted for when building the bridge from the GARM III assessment.

The model formulation used in GARM III utilized an extended age range out to age 11⁺ relative to previous assessments which had used a 7⁺ age group. Commercial and recreational landings from 1982 to 2007 as well as discards from 1999 to 2007 were accounted for in the model. Tuning indices included the NEFSC spring ages 2-8, NEFSC fall ages 1-7 lagged forward by an age and a year (e.g., 2006 age 2 fish become 2007 age 3 fish in the model), MADMF spring ages 2-4, MADMF age 1 lagged forward and commercial LPUE ages 2-6. The fully recruited F is determined as the unweighted average F on ages 5 to 7. The terminal year F on age 10 is estimated as the mean of the fishing mortality on ages 5 through 9. In years prior to the terminal year, F on age 10 was determined from weighted estimates of ages 5 through 9. The age 10 F was applied to the age 11⁺ group. Maturity-at-age was calculated from the three year moving average of maturity observations. Spawning stock biomass was calculated assuming a March 1 spawning period (0.1667 into the calendar year; **note this is inconsistent with the start of the spawning period noted elsewhere in the document and is revised in the final assessment model*).

The general approach used to build the bridge from the GARM III VPA to an updated VPA was as follows (run numbers correspond to the run summaries presented in Tables A.57 and A.58):

- **Run 1:** Recreate GARM III results using v2.7 with GARM III data set to confirm that model and data were correctly applied.
- **Run 2b:** Migrate to v3.1.1 using the GARM III data set to quantify the impact of using an 'exact' solution to the catch equation. Continue to handle plus-group using the GARM III formulation with backward calculation.
- Update the GARM III data set incrementally to understand the impacts of updated data inputs:
 - **Run 3a:** Update commercial landings and discards (exclude discards prior to 1999) and recreational landings through 2007; survey indices not updated, stock and SSB weights unchanged.
 - **Run 3b:** Update stock and SSB weights using the updated weights through 2007 that are presented in Table A.42. Everything else left untouched.

- **Run 4:** Include commercial discards back to 1982 (full time series); survey indices not updated.
- **Run 5:** Include recreational discards through 2007 (full catch update); survey indices not updated.
- **Run 6:** Update the survey indices through 2007, spring surveys through 2008. Update the maturity ogive.
- **Run 7:** Drop the commercial LPUE survey index.
- **Run 8:** Handle the plus-group using ‘combined method’.
 - *This model provides an evaluation of the sensitivity of the GARM III results to the differences in models and treatment of the data.*
- **Run 10:** Update time series through 2010; spring surveys through 2011. *This model represents an updated VPA model.*

The results from the bridge building exercise are presented in Table A.58. There were no major diagnostic problems with the GARM III model following the VPA software update (**Run 2b**). Survey residuals were largely un-patterned (Fig. A.84.a-d). NEFSC survey selectivities suggested constantly increasing selectivity up to the maximum age, with no declines in subsequent ages (i.e., flat-top selectivity) while MADMF spring selectivity decreased sharply with age (Fig. A.85). Fleet selectivity decreased slightly at older ages beyond a maximum-at-age 6, suggestive of some doming (Fig. A.86). A small retrospective pattern was evident in SSB (Fig. A.87) but there was no clear patterning in either F (Fig. A.88) or age-1 recruitment (Fig. A.89). Overall, the results were nearly identical to those of GARM III.

The largest change with respect to the GARM III results occurred from the update of the SSB/January 1 stock weights (Run 3b). In previous assessments stock weights-at-age had been derived from only the landed catch. This approach likely overestimated the true weights-at-age for ages 1 through 3. Based on the updated maturity ogive these ages range from 9.4% to 61% mature (Fig. A.9) and based on the GARM III assessment (Mayo et al. 2009) accounted for 80% of the 2007 population in terms of numbers. Overestimation of the weights-at-age for these younger fish can significantly impact estimates of SSB. The introduction of the recreational discards had minor impacts on the 2007 terminal estimates, primarily in the way of increasing F by 0.13 and decreasing SSB by 3,700 mt (approximately 15%). Minor changes resulted from the survey updates, but dropping the LPUE indices had no impact on the overall results. The net impact from all software and data changes (Run 8) relative to the GARM III results was an increase in F by 0.1 (21.7% increase), and a drop in SSB of 14,428 mt (42.6%). There was a general improvement in the overall retrospective statistics. Time series plots of the major intermediate models are presented in Figures A.90 through A.92.

Updated VPA model (through 2010)

The 2010 update of the Gulf of Maine cod VPA model (Run 10) added three additional years of data: catch and fall survey data were extended through 2010 and spring survey data through 2011. No other changes were made from the Run 8 model formulation. The updated VPA estimates 2010 SSB at 12,270 mt and $F_{5,7}$ at 1.48. The survey fits to Model 10 did not exhibit any strong residual patterns (Fig. A.93.a-c), and survey catchabilities (q) were very similar to those from the GARM III model (Fig. A.94, **note q -values are plotted in terms of area swept in this plot to compare with subsequent ASAP runs*). The fleet selectivities decreased slightly at older ages beyond a maximum between ages 5 and 7, suggestive of some doming similar to the GARM III results (Fig. A.95). Run 10 exhibited extremely high CVs on the population estimates of age 9 and 10 in the terminal year +1 (Table A.58). These high CVs are a product of imprecise estimates of very small numbers of fish (there were an estimated 1000 age 9 and 10 fish in year t_{+1}). There is evidence that there has been further truncation of the age structure since the GARM III

assessment. Continued handling of the plus group as age 11's may no longer be appropriate given this truncation. Retrospective patterning increased in Run 10 relative to Run 2b, particularly in the estimation of SSB (Fig. A.96) and age-1 recruitment (Fig. A.97). The absolute magnitude of the F retrospective statistic (*rho*) remained relatively unchanged (0.05 to -0.06), although there was a change in the overall patterning (Fig. A.98).

Relative to Run 8, Run 10 estimated higher fishing mortality (Fig. A.99) and lower SSB (Fig. A.100) in the overlapping years from 2001 onward. These large differences are driven primarily by a difference in the perception of the recruitment strength of the 2003 year class and to a greater extent, the 2005 year class (Fig. A.101). The strength of these year classes in the GARM III assessment, as well as Run 8 were derived primarily from the NEFSC spring survey (Table A.48) and MADMF fall survey (Table A.54). In the 2010 update (Run 10) not only were there three more years of survey observations with which to gauge the strength of these incoming year classes, but there were additional signals coming from the catch to balance out the high survey data points in 2007 and 2008. By 2010, the 2005 year class was almost fully recruited to the fishery. The catch-at-age (Table A.39) does not show large catches of either the 2003 or 2005 year classes, at least not to the level that would be suggestive of a strong year class. The conflict in the data between early signals of a strong 2005 year class (surveys in 2007 and 2008) and more recent signals that do not suggest a strong year class (surveys and catch for 2009-2010/2011) created tension in the model that manifested itself in the increased retrospective pattern in SSB, and the higher CVs associated with age 5 (2005 year class in 2010) between Run 10 and all earlier model runs (Table A.58). As noted above, precision was also poorer-at-ages 9 and 10, but this is likely be due to there being so few fish at those ages,

The NEFSC spring 2007 and 2008 indices have the highest CVs within the 1968 to 2011 NEFSC spring survey time series (Fig. A.102). Examination of the individual station catches for these two years shows that the high survey data points were driven by single tows in each of the years (Table A.59). The high survey abundances indicated by the NEFSC spring 2007 and 2008 indices are likely not representative of the resource. A contributing factor to uncertainty in recruitment estimates is the MADMF fall survey, which has traditionally been treated as a recruitment index in the VPA model through the inclusion of the age 1 survey index lagged forward a year and an age. Comparison of the MADMF fall age-1 index values to Run 10 age-1 recruitment estimates suggests that the MADMF fall survey is a poor index of recruitment (Fig. A.103). A sensitivity run was conducted to evaluate the performance of the Run 10 model after removal of the MADMF fall index and down weighting of the NEFSC spring survey indices in 2007 and 2008 (all ages set to weighting of 0.1) to account for the high variance of these survey indices (Run 10f). Overall, there was little change in the perception of the stock in terms of terminal estimates of F and SSB (Table A.58 and Figs. A.104 to A.106); however, there was marked improvements in the retrospective patterns, particularly with respect to age-1 recruitment (Fig. A.107) and SSB (Fig. A.108). The comparison of retrospective patterns between runs 10 and 10f suggest that had the GARM III assessment treated the survey indices similarly, the perception of the stock would have been less optimistic back in 2008. Specifically, the 2008 estimate of just under 22,000 mt of SSB would have dropped to about 16,000 mt, and the estimate of age-1 recruitment would have dropped from over 17.9 million to just under 9 million

General conclusions from the updated VPA are:

- Weights-at-age used in GARM III were estimated from only the landed fraction of the catch and likely overestimated the true stock weights-at-age.
- The 2005 year class signal that appeared in the 2007/2008 survey indices was not evident in either later surveys or in the catch.

- As of GARM III, the 2005 year class would have been unavailable to the fishery and the 2003 year class would have only been partially available to the fishery (PR patterns from GARM III suggest approx. 30%).
- The entire signal of the 2005 year class and to some extent the 2003 year class was derived primarily from the survey indices. Compared to the GARM III VPA, the updated VPA estimate of the 2005 year class decreased by 66% and the 2003 year class decreased by 22%.
- **Relative to the 2010 update of the VPA assessment, the 2008 VPA assessment over estimated spawning stock biomass, the strength of incoming year classes and underestimated fishing mortality.**

It should be noted that the VPA model reviewed at GARM III was not alone in overestimating spawning stock biomass. An alternative statistical catch-at-age model (SCAA; Butterworth and Rademeyer, 2008) also reviewed at GARM III (but not accepted as the basis for stock determination) was even more optimistic with respect to stock determination. Admittedly, as described above, there were other issues that lead to the optimistic view of the resource at GARM III, namely the handling of the stock weights, but the assumptions about the strength of the incoming year class were the greatest contributor to the optimistic view of Gulf of Maine cod at GARM III. Both models reviewed at GARM III, the VPA and the SCAA, failed to account for the uncertainty in the 2003 year class and to a larger degree the 2005 year class. Problems predicting the strength of incoming year classes has historically plagued the Gulf of Maine cod assessment:

- From GARM II (NEFSC 2005):
 - *“The estimate of the strength of the 2003 year class is very sensitive to the MA DMF 2004 autumn age 1 index, included as the 2005 age 2 index in the VPA calibration. Exclusion of this single datum results in an estimate of 15 million fish vs. 22 million fish at age 1 in 2004. This value does not substantially affect the estimate of 2004 spawning stock biomass, but does influence starting conditions for projections.”*
- From GARM III (Mayo et al. 2009):
 - *“...biomass indices began to increase substantially in 2001 and spring 2002, but the large apparent increase evident in autumn 2002 resulted from a single large haul unduly influencing the stratified mean.”*
 - *“A retrospective pattern is also evident for age 1 recruitment estimates whereby recruitment was well overestimated for the 2001 and 2003 year classes...The estimate of the size of the 2005 year class appears to not suffer the same fate, as it is supported by an additional year of data in the present assessment...”*

Sensitivity of model results to assumptions of peak spawning period

During the NDDWG’s review of the MADMF cod IBS survey data, time was spent evaluating the period of peak spawning in the Gulf of Maine. The available data suggests that peak cod spawning, particularly in the western Gulf of Maine where the stock is most heavily concentrated, seems to occur at the

beginning of April and extend into May. Previous Gulf of Maine cod assessments, including the Run 10 VPA model examined in this report, have used an assumption that the spawning period occurs at the end of February/beginning of March. The assumption of an April 1 spawning period is likely a more accurate estimate for the Gulf of Maine stock. The impacts of this change were evaluated in the context of the Run 10 VPA by performing a sensitivity run that moved the spawning period to April 1 (Run 10g). This change has virtually no impact on estimates of F (Fig. 104) or recruitment (Fig. A.105) and only minor changes in SSB (Fig. A.106). Because the revised spawning period occurs later in the year, there is an additional month of natural mortality and fishing mortality prior to the spawning period, hence a decrease in estimated SSB. This change was examined by the Northern Demersal Models and Biological Reference Point Working Group (NDMBRPWG) and it was agreed that an April 1 spawning period would be used in the base case model.

Development of an ASAP statistical catch-at-age model

The use of a statistical catch-at-age model for the Gulf of Maine cod assessment was explored. More specifically, the statistical catch-at-age model, ASAP (Age Structured Assessment Program v2.0.20, Legault and Restrepo 1998), which can be obtained from the NOAA Fisheries Toolbox (<http://nft.nefsc.noaa.gov/>). The reasons for selecting the ASAP model include: ability to explore alternative model formulations to counter/lend support to VPA results, additional flexibility to explore starting condition assumptions (e.g., extending the time series beyond 1982), ability to estimate a stock-recruit relationship internal to the model, and the ability to explicitly handle data uncertainty, particularly given the lessons learned from the update of the VPA model with respect to uncertainty in the survey data.

ASAP is an age-structured model that uses forward computations assuming separability of fishing mortality into year and age components to estimate population sizes given observed catches, catch-at-age, and indices of abundance. Discards can be treated explicitly. The separability assumption is partially relaxed by allowing for fleet-specific computations and by allowing the selectivity-at-age to change in blocks of years. Weights are input for different components of the objective function which allows for configurations ranging from relatively simple age-structured production models to fully parameterized statistical catch-at-age models. The objective function is the sum of the negative log-likelihood of the fit to various model components. Catch-at-age and survey age composition are modeled assuming a multinomial distribution, while most other model components are assumed to have lognormal error. Specifically, lognormal error is assumed for: total catch in weight by fleet, survey indices, stock recruit relationship, and annual deviations in fishing mortality. Recruitment deviations are also assumed to follow a lognormal distribution, with annual deviations estimated as a bounded vector to force them to sum to zero (this centers the predictions on the expected stock recruit relationship). For more technical details, the reader is referred to the technical manual (Legault 2008).

In developing the base ASAP model configuration over 20 preliminary models configurations were explored. These preliminary model configurations attempted to take advantage of ASAP's flexibility by handling commercial and recreational fleets separately and breaking out catch components into landings and discards. These complex model formulations suffered from strong residual patterning and/or overall model instability from being over-parameterized. Minor changes to model parameters would often lead to non-convergence. Moreover the model results from these complex models were nearly identical to some of the simpler models explored. A more in depth overview of these preliminary model configurations as well as other ASAP sensitivity runs is provided in Appendix 2. The difficulties encountered in these initial explorations led to a more parsimonious approach to the model formulation with the use of a single aggregated fleet (i.e., identical to the VPA). Sensitivity runs on these simpler model formulations examined the impacts of inclusion/exclusion of the MADMF fall and LPUE survey indices. Model

performance and stock perception were robust to the inclusion/exclusion of these data and were therefore left out of the base ASAP configuration.

ASAP base model configuration (BASE)

A decision was made to use an age 9 plus group in the ASAP base model configuration (BASE). This decision was based on the difficulties of the VPA to precisely estimate older ages due to what appears to be continued truncation in the population age structure over the most recent three years and the difficulties in precisely estimating fishery selectivities of the older ages in preliminary developmental ASAP runs. An 11⁺ ASAP sensitivity to the base configuration will be explored later.

Selectivity-at-age was freely estimated for each of the two fishery selectivity blocks, but the two NEFSC surveys were fixed at 1.0 for ages 6 and older (i.e., flat top selectivity) and the MADMF indices were fit using a double logistic functional form to capture the decreasing selectivity-at-age apparent in the VPA selectivity patterns. The choice of the flat-topped selectivity pattern for the NEFSC survey indices was informed in part by the VPA results, which suggested increasing catchability with age, and the likelihood calculated in ASAP for domed versus flat-topped scenarios. Additionally, comparison of proportion of fish age 5 and older caught in the NEFSC surveys relative to the fishery shows a higher ratio of old fish caught by the NEFSC surveys (Table A.60). This in itself does not confirm the presence of flat top survey selectivity, but does support a conclusion of higher selectivity-at-age in the survey relative to the fishery. There have been discussions during previous assessment meetings and working group meetings that adult cod may be unavailable to the NEFSC surveys due to the presence of fixed gear (primarily lobster pots) in the inshore areas. However, the MENH survey actively works with the lobster industry to have gear removed in advance of the survey and as noted before, this survey is not capturing large cod (Fig. A.82). Decreased selectivity in the fishery may be plausible, particularly if large cod are exploiting closed areas unavailable to the fishery (either permanent or seasonal). However, the NDDWG cursory examination of the Cod IBS survey length frequencies did not indicate the presence of larger cod in the rolling closure areas relative to those captured in the fishery or surveys. Additionally, an analysis of cod tagging data conducted by Hart and Miller (2008) concluded that there was no evidence that larger/older Atlantic cod are subjected to lower fishing mortality in the Gulf of Maine than smaller cod. The VPA results, however, do show some propensity for moderate doming in the fishery (Fig. A.95), but do not support the severe doming suggested by some models (e.g., Butterworth and Rademeyer 2008a). Further sensitivities to the doming assumptions will be explored later in this report. It should be noted that many of the preliminary ASAP runs allowed for domed survey selectivity and the results of these runs were generally similar to the ASAP BASE model results (Appendix 2).

Beginning with a single selectivity function for the fishery, model diagnostics were examined for trends in age composition residuals. With only one selectivity block (i.e., the same selectivity assumed for years 1982-2010), there were notable trends in age composition residuals with runs of positives and negatives. An additional selectivity block was introduced beginning in 1989 and several intermediate models were run exploring splits from 1989 to 1994. The period from 1989 to 1994 encompassed major changes in data availability, reporting sources and fisheries management. The model with a 1990/1991 split had the lowest objective function and offered improved fit to the age composition in the way of reduced residual patterning. The base model contains two fleet selectivity blocks: 1982-1990 (block 1) and 1991-2010 (block 2).

For the fishery, selectivity-at-age is freely estimated within each block for 8 out of 9 ages, with one age class fixed at full selectivity in each block. In block 1, age 5 was assumed to be fully selected, while in block 2 age 6 was assumed to be fully selected. This decision was informed on the basis of smaller mesh sizes and minimum retention sizes during the years included in block 1. Each of the two NEFSC surveys

included a single time invariant selectivity vector with selectivity-at-age being freely estimated from age 1 to age 5 and fixed at age 6 and older. The MADMF spring survey was fit using a double logistic function to account for the sharp declines in selectivity-at-age observed in the VPA results. The descending slope of the double logistic function experienced boundary problems in preliminary runs and was subsequently fixed at 10 in the base model.

The effective sample size (ESS) estimated for both the fishery and survey catch-at-age data (which are treated as multinomial) was compared to the input effective sample size in an iterative fashion until the effective sample size specified more or less matched the model estimated value, or until no further improvement in trying to match the estimated value could be made. Additionally, following Francis (2011), minor adjustment in the effective sample sizes were informed by the overall fit between the predicted and observed mean age of the catch. The final ESS for the fishery was set to 75, the two NEFSC surveys set to 30 and the MADMF spring set to 15. The CVs on the surveys were initially set equal to the bootstrapped CVs presented in Tables A.47 and A.52). The bootstrapped CVs characterize the sampling error, but additional process error may be present in the survey indices that are not reflected in the bootstrapped CVs. Subsequent examination of the model fits to the survey indices resulted in adjustments to the survey CVs by adding the following constants to each of the survey CV vectors to account for additional process error: 0.2 (NEFSC spring), 0.1 (NEFSC fall), 0.3 (MADMF spring). It should be noted that these minor adjustments offered slight improvements to the statistical fit of the model but had little impact on the model results (e.g., see earlier models presented in Appendix 2 where survey CV vectors were not adjusted).

An annual CV of 0.05 was assumed for the fishery catch. This was a trade-off in forcing an exact fit to the catch (as in a VPA-like formulation) versus accounting for some of the uncertainty in catch owing to the uncertainty in stock allocation, discard estimation and hindcasting procedure. Commercial landings in the assessment time period are assumed to be very precise. There is a limited amount of error introduced in the allocation procedure and through VTR misreporting, but generally, these uncertainties are low. CVs on commercial discards are in the range of 0.11 – 0.38 and recreational catch PSEs are in the vicinity of 20%. Given the overall uncertainties, the assumption of a constant catch CV=0.05 was not unreasonable. Model sensitivities to alternate CV assumptions are explored in Appendix 2, but overall, the model results are robust to alternate estimates of catch precision.

ASAP base model (BASE) diagnostics

ASAP BASE model fits to the fishery catches were good, with no strong patterning of residuals over time and generally good agreement between modeled and observed catches (Fig. A.109). A ESS of 75 on the fishery catch-at-age appeared reasonable (Fig. A.110), and achieved reasonable fits to the observed catch-at-age (Fig. A.111.a-d) with no large residual runs or obvious year class effects apparent in the residual patterning (Fig. A.112). Model fits to the observed mean catch-at-age are good, with a root mean square error (RMSE) of 1.28 (Fig. A.113). Fishery selectivities were moderately domed in both blocks (Fig. A.114). The selectivity patterns in block 1 are somewhat noisy and not well explained by biological or management-based mechanisms.

The overall fits to the survey indices were good, with the relationship of observed to predicted survey indices generally falling around the 1:1 equality line (Fig. A.115). Fits to the NEFSC spring survey index exhibited no strong residual patterning (Fig. A.116). It is notable that the ASAP model did not fit the 2007 and 2008 index values well, with the model fits being influenced by the high CVs in these years. The input ESS value of 30 were generally supported by the modeled estimates (Fig. A.117) and decent fit of observed to predicted age compositions (Fig. A.118). There was no strong residual patterning to the index age composition fits (Fig. A.119), although there are some transient year class effects in the early to

mid-1990s. Fits to the mean age were comparable to the fishery mean ages (Fig. A.120, RMSE=1.47) lending additional support to the input ESS.

Models fits to the NEFSC fall survey were better than the spring fits, with stronger coherence between the observed index and modeled estimate (Fig. A.121). ESS values of 30 are generally supported by the modeled estimates, though there is some suggestion of decreased ESS more recently in the time series (Fig. A.122). The fit to the age composition was good, with observed to predicted indices-at-age, generally falling around the 1:1 equality line (Fig. A.123) and very little patterning to the survey indices age composition residuals (Fig. A.124). The overall fit to the mean catch-at-age is reasonable, though there is some indication of reduced fit in the most recent period (Fig. A.125) as suggested by the comparison of the input ESS to the modeled ESS values.

Similar to the fits to the NEFSC surveys, the fit to the MADMF spring survey is reasonably good with the model tracking the observed index values moderately well, with no strong residual patterning (Fig. A.126). The modeled ESS is noisy, but overall, the input ESS appears reasonable (Fig. A.127). The MADMF spring age compositions were not fit as well as the NEFSC surveys (Fig. A.128), with the magnitude of residuals being somewhat larger for this survey relative to the others, though no long runs of residuals (either positive or negative) are observed (Fig. A.129). Estimated mean ages were fairly close to the observed mean ages, with a RMSE of 1.32 (Fig. A.130).

The NEFSC fall survey exhibits higher selectivity at younger ages relative to the spring survey (Fig. A.131). Survey catchabilities (q) are presented in Figure A.132. The q CVs were less than 20%. The NEFSC spring survey $q=0.92$ which would appear to suggest that the NEFSC spring is close to 100% efficient. Considering the calibration coefficients applied to the Bigelow survey years, this would suggest greater than 100% efficiency over the last two years. This is not necessarily a valid assumption and caution needs to be taken when interpreting the area-swept converted values of q . A full exploration of the survey q estimates is provided in Appendix 2 along with model independent estimates of total stock biomass which support the general scale of biomass estimated by the BASE model.

Additional ASAP sensitivity runs

Over ten different sensitivity runs were explored to evaluate the sensitivity of the ASAP model to alternate assumptions. A full documentation of the range of sensitivity runs is presented in Appendix 2. Four specific sensitivity runs that were critical to the final formulation of the BASE model are presented: sensitivity to the age of the plus group (BASE_11, a plus group at 11 instead of 9), assumptions about survey selection (flat top vs. dome; BASE_DOME), model starting points (e.g., including data before age composition information was available). Two different starting point assumptions were investigated: 1970 (BASE_1970), which extends the time series back to the start of the time series where survey age composition information is available; and, 1964 (BASE_1964), back to the start of modern landings statistics.

In all sensitivity runs the model configurations were kept identical to the BASE model except where noted. For the BASE_DOME run, survey selectivity on age 6 was fixed with the model allowed to freely estimate selectivity at all other ages. With the historical runs, the average weights-at-age from the period 1982 to 1990 (block 1) were extended backward to the beginning of the time series. Additionally, since hindcasted time series only extend as far back as 1982 for commercial discards and 1981 for recreational discards, a 25% 'bump-up' factor was applied to the 'Total catch (mt)' column in Table A.6 in the years prior to 1981. A summary of all sensitivity model configurations is provided in Table A.61.

The BASE model was insensitive to the plus group specification; the BASE and BASE_11 models

achieved nearly identical results throughout the time series with respect to SSB (Fig. A.133), F (Fig. A.134) and age-1 recruitment (Fig. A.135). Fits to the total catch and aggregate survey indices were nearly identical between the two runs (Table A.62). The survey selectivities of ages 10 and 11 were poorly estimated as evidenced on the large CVs on these ages in both fishery blocks 1 and 2 (Table A.63). Selectivity of age 10 in block 1 hit a boundary at 1. Given the insensitivity of model results to the choice of the plus group and the poorly estimated selectivities on older ages, the base model configuration using age 9 as the plus group is supported.

Relative to the BASE model, the influence of allowing survey selectivities to be domed resulted in a positive rescaling of SSB (e.g., 21% increase in 2010 SSB) and a decrease in F, particularly in the second fishery block (1991-2010). There was virtually no change in estimated recruitment. The majority of the increase in SSB was driven by increases in the older ages (e.g., age 9⁺, Fig. A.136) due to more severe doming of fishery selectivities (Table. A.63). Based on the evidence presented earlier, there is little biological or scientific evidence to support such strong doming, additionally, there was little model support for this with an increase of 6 parameters and an improvement of only 3 objective points.

The historical runs, BASE_1970 and BASE_1964, did not alter the perception of the stock. Nearly identical trends were observed in F (Fig. A.137) and SSB (Fig. A.138). The small differences in F and SSB observed at the end of the series are being driven almost exclusively by differences in recruitment (Fig. A.139), as fleet and index selectivities are almost identical between the BASE run and the two historical runs. With respect to evaluating the current condition of the stock, the choice in starting year has little impact. Where the starting year *does* make a difference is in establishing reference points. There is a high degree of uncertainty in the recruitment estimates pre-1982 since they are driven solely off of survey age compositions run. Given the experience of the GARM III VPA update, caution should be taken in placing too much weight on recruitment estimates driven entirely off of survey information that cannot be corroborated with catch-at-age information.

ASAP base (BASE) model results

The ASAP BASE model configuration reflects the consensus opinion of the NDMBRPWG as the best model with which to evaluate stock status and provide catch advice. The assessment indicates that total SSB has ranged from 7,270 mt to 23,675 mt during the assessment time period, with current SSB in 2010 estimated at 11,868 mt (Table A.64, Fig. A.140). The base model estimates SSB in 2007 at 12,561, 37% of the 33,877 mt estimated at GARM III. Total biomass in 2010 is estimated at 20,589 mt and F's at the end of the time series are near historic highs (Fig. A.140) with the 2010 fully recruited, $F_{full} = 1.14$ and $F_{5-7} = 1.10$ (Table A.65). Fishing mortalities-at-age are presented in Table A.66. The low fishing mortality on ages 1 through 3 is notable given that the maturity $A_{50\%}$ is between ages 2 and 3. The current fishery selectivity allows one to two spawning events on average prior to entering the fishery. These patterns partly explain the persistence of the population in the presence of the high Fs over the past decade.

Recruitment over the past decade has been poor despite modest increases in SSB (Fig. A.141 and A.142). Age-1 recruitment has not exceeded 10 million fish since 1999 and has exceeded that threshold only twice in the past twenty years (Table A.67). While there is an absence of a well defined stock-recruit relationship there is some indication of a relationship. The five highest recruitment events in the time series were spawned during a six year period from 1982 to 1987 where the SSB was near the highest observed in the time series, averaging over 15,000 mt annually. The current population structure is comprised primarily of fish that have not yet recruited to the fishery (fish age 1-3), with approximately 25% of the population age 4 and older (Table A.67 and Fig. A.143).

MCMC simulation was performed to obtain posterior distributions of the SSB, total B, F_{full} and F_{5-7} time

series. Two MCMC chains of initial length 1 million were simulated with every 100th value saved. The trace of each chain's saved draws suggests good mixing (Fig. A.144). The lagged autocorrelations showed decreasing correlation with increased lag with correlations < 0.1 beyond lag 6. Ultimately, a subsequent thin was applied by saving every 10th value to create an MCMC chain with a length of 1000. Finally, the Gelman-Rubin potential scale reduction factor (psrf) was calculated for the time series of F_{5-7} and SSB. All psrf were between 1.0 and 1.01, which again, suggests convergence of the chains. As the MCMC simulations appear to have converged, 90% probability intervals (PI) were calculated to provide a measure of uncertainty for the model point estimates. Time series plots of the 90% PIs as well as plots of the posterior for B_{2010} , SSB_{2010} and $F_{5-7(2010)}$, F_{full} are shown in Figures A.145 through A.148. ASAP point estimates and the 90% PIs are reported below:

Metric	ASAP point estimate	90% probability interval
SSB_{2010} (mt)	11,868	9,479 - 16,301
B_{2010} (mt)	20,589	17,638 - 25,996
F_{full}	1.14	0.79 - 1.54
F_{5-7}	1.10	0.74 - 1.46

Retrospective analysis for the 2003-2010 terminal years indicates retrospective error in both F and SSB with the tendency for the model to underestimate F and overestimate SSB (Fig. A.149 and Fig. A.150). The F retrospective error ranged from -0.10 in 2009 to -0.52 in 2003 (Table A.68). SSB retrospective error ranged from 0.09 in 2009 to 0.90 in 2003. Retrospective error in age 1 recruitment varied from -0.07 in 2005 to 4.32 in 2003. It is worth noting the decreased retrospective pattern in Age 1 recruitment in the ASAP BASE run (Fig. A.151), relative to the updated VPA run (Run 10, Fig. A.97). The ASAP model does not exhibit nearly as severe a retrospective pattern in the recent period, particularly in the 2008 assessment peel (coinciding with the timing of the GARM III assessment). This suggests that had ASAP been used as the base model in GARM III, the assessment results would not have been as susceptible to the uncertainty in the 2007 and 2008 NEFSC spring survey indices. Retrospective statistics calculated using both seven year peels and five year peels are presented in Table A.68. However, the NDMBRPWG noted that there was a notable shift in the retrospective pattern such that retrospective statistics (Mohn's rho) calculated using a five year peel (back to 2005) more accurately capture the current retrospective patterns.

Historical assessment retrospective

A comparison between the results of the current assessment (including the updated VPA for perspective) and the four previous assessment (SARC 53, GARM I, GARM II and GARM III) is provided in Figures A.152-A.155. This historical "retrospective" examination of past model performance illustrates the general tendency of updated models to achieve higher estimates of F and lower estimates of SSB, total biomass and overall stock size over the last decade. These patterns are in addition to the intra-model retrospective patterns that are present in the existing ASAP model as well as past VPA models. Given the major changes in data that have occurred in the most recent update, the current assessment is not entirely comparable with previous assessments. Much of the scale differences between the current assessment and previous assessments are driven by changes to the underlying data (e.g., weights-at-age) and not as a result of the assessment or choice of model. It is important to note that the updated VPA and ASAP BASE model achieve nearly identical results; however, given the capacity of the ASAP BASE model to better account for data uncertainty, it is considered the preferred model on which to base fisheries management advice.

Sensitivity analysis to assessment model (Butterworth & Rademeyer SCAA)

An additional statistical catch-at-age (SCAA) assessment model was considered by the NDMBRPWG (mathematical details of which are provided in Appendix 4). In the course of the NDMBRPWG meeting, attempts were made to bring the two models (based on an assessment time series of 1982-2010) into as close agreement as possible. The following list of items was identified as methodological differences between the two models.

- Equilibrium age structure under estimated F parameter (SCAA) versus freely estimated age structure (ASAP).
- Likelihood to fit indices (SCAA estimates an additional variance when fitting survey indices; described in Appendix 4)
- Likelihood for age compositions (SCAA adjusted lognormal, ASAP multinomial)
- Use of biomass (SCAA) versus abundance survey indices (ASAP) for tuning
- Use of Baranov (ASAP) versus Pope’s approximation (SCAA) under high F conditions (model F’s are near 1)

The NDMBRPWG was able to ascribe most of the differences between model estimates as likely due to the following three items: different estimates of selectivity (arising from likelihood form for age composition data), use of Pope’s approximation rather than Baranov to estimate F, and the time of the year when SSB was calculated (0.25 in ASAP versus 0.1667 in SCAA). Of these three items, the only one that would require further research is the form of the likelihood. For the estimation of F, Baranov is preferred when fishing mortality rates are high.

A comparison of the results of the base ASAP (BASE) results to the SCAA results are presented below. In an effort to address one of the differences highlighted above, SCAA results are presented using both Baranov and Pope’s approximation to estimate F. While the SCAA Baranov results were not reviewed by the NDMBRPWG, they do help address the difference noted above.

Biomass estimate	ASAP (BASE)		SCAA Pope		SCAA Baranov	
SSB ₁₉₈₂ (mt)	23,675	(20,760 - 26,958)	31,549	(19,831 - 43,267)	30,294	(19,642 - 40,946)
SSB ₂₀₁₀ (mt)	11,868	(9,479 - 16,301)	17,373	(13,713 - 21,033)	16,481	(11,695 - 21,267)
SSB ₀ (mt)	171,417	(136,351 - 218,992)	214,258	(7,481 - 421,035)	188,342	(59,499 - 317,181)
SSB _{MSY} (mt)	54,247	(41,394 - 72,462)	68,118	(59,626 - 76,609)	65,943	(53,936 - 71,446)
MSY (mt)	10,691	(8,012 - 14,687)	10,250	(8,891 - 11,609)	10,107	(8,462 - 10,754)

**Note that ASAP reference points were not estimated internally within the model but estimated through long term projections described in TOR. Also, confidence intervals (CI) presented for ASAP are 90% CI, while the SCAA are 95% CI.*

At the close of the NDMBRPWG meeting, the group was comfortable that despite the structural differences between the two models, they were capable of producing similar results when configured similarly. The scale of the SCAA model is slightly higher than the ASAP (BASE) model, though the trends are similar. Thus, the SCAA model provided valuable feedback regarding model sensitivity to assumed error distributions, estimation of starting conditions, and selectivity fitting.

TOR A.4. Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (TOR-3).

Historically, the recreational fishery has been split between Georges Bank and Gulf of Maine. Since 1999, recreational landings of Atlantic cod have been predominately in the Gulf of Maine region (NEFSC 2008). The potential for misallocation of recreational landings is unknown, however, given the behavior of the recreational fleet operating in the Gulf of Maine, the magnitude of the impacts is likely to be small. The issue is misallocation of commercial landings is likely to be larger and have a greater impact on model performance. With respect to Gulf of Maine Atlantic cod, the allocation procedure itself does not contribute additional uncertainty as indicated by the low CVs on the allocated landings (Table A.7). A more likely source of allocation uncertainty arises from the misreporting of statistical area on VTRs. The previously discussed work of Palmer and Wigley (2007, 2008, and 2010) suggests that these impacts are likely to be small (<5%), but consistently unidirectional (under-reporting of total Gulf Maine cod catch).

Sensitivity runs were conducted to bound the potential impacts of mis-allocation. Two sensitivity runs were conducted, one which inflated landings by 5% and another which decreased landings by 5%. Spawning stock biomass changed +/- 5% with no change in F. The 2010 estimates of SSB were within the 95% confidence intervals achieved from the MCMC estimate of uncertainty (9,479 – 16,301 mt; Fig. A.156).

TOR A.5. If time permits, consider the small-scale distribution of cod (e.g., spawning sites, resource distribution, fishing effort) in the Gulf of Maine and advise on its management implications.

Discussion related to resource distributions occurred throughout the NDDWG meeting as both surveys (NEFSC, MADMF, MENH, IBS) and fleet activity were reviewed. Given the full agenda, and extent of reanalysis of data, there was not an abundance of time available to delve into this TOR. The NDDWG did attempt to review as much with the time available. The main points relating to Gulf of Maine cod distributions discussed by the NDDWG are summarized below as bullet points:

- There is a body of work that has attempted to investigate small-scale distributions of Gulf of Maine cod. This work includes collaborative work between University of Massachusetts School for Marine Science and Technology (SMASST) and MADMF in the Cod Conservation Zone (CCZ) in the western Gulf of Maine; University of New Hampshire (UNH) research around the Whaleback Closure; and a longline sentinel survey from Downeast Maine.
- The studies in the western Gulf of Maine confirmed that many of the fish on the spawning aggregations show site fidelity; that the timing of the closures is appropriate; and that when fishing resumes at the end of the closure it can be very disruptive to the cod (interrupts any residual spawning because the fish rapidly disperse from the spawning grounds). Wandering from spawning grounds was detected with the aid of acoustic tags and arrays. It was suggested to evaluate the size of fish on the spawning ground as a function of when they arrive to see if large fish enter first with smaller fish moving in only towards the end of the spawning area closure. Feeding patterns could also be examined to see if that is the reason for wandering.
- Recreational fishermen are aware of the spawning sites but it is unclear whether they have always known about them, or whether they have just starting going there since the commercial vessels stopped. It would be interesting to plot VTR information for recreational data on a map of habitats to try to identify any patterns that might indicate the existence of other important spawning areas. It would also be interesting to identify whether there were physical, ecological characteristics that make these areas preferred habitat.

- UNH studies confirmed that spawning sites exist off the coast of New Hampshire and the Whaleback Closure encompasses the majority of the density identified in those studies.

- The Downeast Maine sentinel survey has completed some pilot field work. The longline survey sets approximately 2000 hooks/day for 30 days in summer with the goal of establishing a baseline of cod abundance so that any rebuilding or recolonization of the Maine coast can be detected. The low abundance observed to date in the survey confirms distributions seen in annual plots for the MENH survey. These results are also consistent with the Northeast Regional Cod Tagging Program, which suggests that there few cod in the Downeast Maine region from 2003-2005.

- The MADMF IBS survey distribution data confirm the patterns seen in MADMF and NEFSC surveys, with cod moving offshore in the fall compared to the spring.

TOR A.6. State the existing stock status definitions for “overfished” and “overfishing”. Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY} , $B_{THRESHOLD}$, F_{MSY} , and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the “new” (i.e., updated, redefined, or alternative) BRPs.

The existing MSY reference points are based on a spawning potential ratio (SPR) of 40%. The overfishing definition is $F_{MSY} = F_{40\%} = 0.237$. A stock is considered to be overfished if spawning biomass is less than half of SSB_{MSY} . The existing overfished definition is $\frac{1}{2} SSB_{MSY} = \frac{1}{2} SSB_{40\%} = 0.5 \cdot 58,248 \text{ mt} = 29,124 \text{ mt}$. A history of Gulf of Maine cod reference point values since 2001 is provided in Table A.2.

The existing MSY reference points were derived from a VPA model with a plus group at age 11. There are a number of reasons why new reference points are needed for the proposed base model for the current assessment including: the number of age classes modeled in the BASE model is 9 instead of 11 (this changes the weight and selectivity in the plus group), commercial and recreational discards are included (this changes the weights and selectivities at all ages), the parameters of the LW equation were re-estimated (this also affects weights at all ages), and the time elapsed before spawning was increased from 0.1667 to 0.25 (this affects discounting in YPR calculations).

The ASAP model has the capability to estimate a stock recruit function within the model; however, initial model runs attempting to fit a Beverton-Holt function were unsuccessful. Analytic model-based reference points are not estimable because there is insufficient contrast in the ASAP base model time series of estimated SSB and recruitment (1982-2010). There was consensus among the NDMBRPWG that a proxy reference point approach was the preferred method to estimate updated reference points given an assessment time series of 1982 to 2010. Yield per recruit (YPR) analysis was performed with a 3-year average of weights-at-age. The remaining YPR inputs were time invariant (maturity-at-age) or were constant in the most recent time block of the assessment model (selectivity). YPR inputs are summarized in Table A.69. The NDMBRPWG evaluated the sensitivity of YPR estimates to the number of years in the average weight calculation by comparing the results from the 3-year average approach to those of a 10-year average. The YPR estimates were insensitive to alternate averaging time blocks.

Despite the inability to estimate a stock recruit function, there was consensus that F_{MAX} was not a sensible overfishing reference point for the Gulf of Maine cod. Use of F_{MAX} implies that there is no relationship between spawners and recruits. In the context of the current Gulf of Maine cod assessment, not having contrast in the data series to reliably estimate a stock recruit function is not saying that there is no

relationship between spawners and recruits. Given the consensus that F_{MAX} was not acceptable as a reference point, the working group debated what would be an appropriate %SPR for the resource.

The current reference points were derived at GARM-III, and are based on $F_{40\%}$. The decision to use $F_{40\%}$ as a proxy for F_{MSY} was endorsed by the independent reviewers at the GARM III meeting, who wrote that *“If the recruitment and spawning stock biomass derived from the assessments are not informative about a relationship, the Panel recommended use of $F_{40\%MSP}$ as a proxy for F_{MSY} (NEFSC 2002) and a B_{MSY} proxy computed using the stochastic projection approach (herein termed the ‘non-parametric’ approach)”* (NEFSC 2008, p979). Furthermore, it was noted that $F_{40\%}$ is supported by published studies on sustainability (NEFSC 2008; Overholtz et al. 1986; Gabriel et al. 1989; Clark 1991; Clark 1993; Goodyear 1993; Clark 2002). It was pointed out by a member of the NDMBRPWG that the published studies focused on F_{MSY} proxies that emphasized sustainability while minimizing yield loss rather than the implications for rebuilding and that the use of $F_{40\%}$ does not fully consider the biomass implications of the overfishing proxy. There were different views within the NDMBRPWG as to the relative priorities of focusing on sustainability and minimization of yield loss, versus implications for biomass targets and rebuilding. With respect to the yield minimization argument, the updated estimate of $F_{40\%}$ was nearly the same as $F_{0.1}$ (0.20 versus 0.21 respectively). The amount of SSB that corresponds to $F_{40\%}$ is 61,218 mt, whereas the 1982-2010 time series of spawning biomass estimates from the preferred ASAP model is 7,270 mt – 23,675 mt. While the SSB_{MSY} reference point is outside the range of SSB that has been seen in model estimates, it should be noted that the model begins in 1982 while the Gulf of Maine cod stock has been exploited for centuries and may already be quite depleted. If the stock is highly depleted within the years modeled, one would not expect to have observed SSB on the scale of estimated SSB_{MSY} . Given the limited contrast in model estimates from the past 30 years there are few data to support estimation of unexploited conditions. Nevertheless, there was consensus that extrapolation beyond the range of ASAP estimates of SSB was necessary to define SSB_{MSY} . This decision, and the observation that reference points would be beyond abundance levels observed since 1982, is consistent with the conclusions from the working group that re-evaluated biological reference points for New England groundfish at GARM II (NEFSC 2002a).

Survey data were examined to determine if there was support for a positive relationship between spawners and recruits. There was a weak trend for higher age 1 fall survey indices to be associated with larger fall survey biomass indices (Fig. A.157). The working group agreed that this analysis provided some additional support that recruitment is higher when spawning abundance is higher, however the question of an appropriate %SPR could not be resolved from this work. An alternative exploratory analysis to address this question considered historical catch and survey data. Although the ASAP preferred model begins in 1982, sensitivity models were conducted during the working group meeting that began in either 1970 or 1964 that could potentially provide more contrast in SSB and recruitment. The working group decided to look at the 1970 run rather than the 1964 run, because there is survey age composition beginning in 1970 from which recruitment fluctuations could be estimated. The 1970 sensitivity run provides some evidence that larger recruitment was associated with higher spawning biomass (Fig. A.158). A Beverton-Holt stock recruit relationship was fit within ASAP for the model that began in 1970 as an exercise to determine whether there was sufficient contrast with the additional data to inform the group about productivity and an appropriate %SPR (Fig. A.159). The 1970 ASAP sensitivity model was able to estimate a Beverton-Holt stock recruit relationship, and the residual diagnostics were not unreasonable (Fig. A.160). The estimate of steepness was 0.89 and the implied unexploited conditions were 315,152 mt. The estimate of F_{MSY} , and corresponding %SPR $_{MSY}$, from this exercise informed the decision about an appropriate $F_{\%SPR}$ proxy. The estimate of F_{MSY} from the 1970 ASAP run was 0.24, which corresponds to a %SPR in the YPR analysis of about 35%.

The proxies for F_{MSY} that were debated were $F_{22\%}$ (F_{MAX}), $F_{35\%}$ (F_{MSY} in the 1970 ASAP sensitivity run), and $F_{40\%}$ (status quo). Ultimately, the SARC Panel did not feel that there was sufficient justification for

the $F_{35\%}$ approach. An $F_{40\%}$ approach will be used for reference point determination.

To arrive at estimates for $SSB_{40\%}$ and corresponding MSY, long term projections were run, sampling from the empirical distribution of recruitment estimates from the preferred ASAP model (recruitment estimates from 1982-2008, final two years excluded). Based on suggestions made by the SARC 53 Panel, the modeling approach used to estimate reference points in GARM III was modified to better account for uncertainty in projections at low stock sizes. Identical to the modeling used in GARM III, the revised projection model samples from a cumulative density function derived from estimated age-1 recruitment. However, the revised model adjusts projected recruitment when SSB falls below some specified spawning biomass threshold based on a linear function that declines to zero at zero spawning stock biomass. For all projections, the threshold SSB was set at 7.3 thousand mt, which coincides with the lowest observed SSB in the time series. To approximate the distribution of the SSB and MSY distributions, the long term projections were made from 1000 estimates of NAA in 2011, which were estimated by performing MCMC simulation of the ASAP base model (described above under TOR 3). The resulting reference points and their 90% confidence intervals corresponding to $F_{MSYproxy}=F_{40\%}=0.20$ are $SSB_{MSY} = 61,218$ mt (46,905 – 81,089 mt), $MSY = 10,392$ mt (7,825 – 14,146 mt). All projections were conducted with the AGEPRO software (Age Structured Projection Model v4.1).

TOR A.7. Evaluate stock status with respect to the existing model (from the most recent accepted peer reviewed assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.

TOR A.7.a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

The existing peer reviewed assessment model is a VPA. A meticulous bridge was built from the existing VPA model structure to the updated VPA model structure. The updated VPA model, which includes changes to the catch (inclusion of discards), weights-at-age, etc., estimates that in SSB_{2010} is 12,270 mt. This is less than the existing overfished threshold of 29,124 mt; therefore, the stock is overfished. The updated VPA estimate of average fishing mortality on ages 5-7, $F_{(5-7)2010}$ is 1.48, while the fully recruited F from the VPA is $F_{full}=2.46$. These are both greater than the overfishing limit, and therefore, overfishing is occurring.

TOR A.7.b. Then use the newly proposed model and evaluate stock status with respect to “new” BRPs (from Cod TOR-6).

The revised reference points are $F_{MSYproxy}=F_{40\%}=0.20$ and $SSB_{MSY} = 61,218$ mt ($0.5 \times SSB_{MSY} = 30,609$ mt). The proposed ASAP base model 2010 estimate of SSB is 11,868 mt. This is less than the overfished threshold of 30,609 mt; therefore, the stock is overfished. The estimate of 2010 average fishing mortality on ages 5-7 from ASAP is $F_{5-7}=1.10$, while the fully recruited F_{2010} is 1.14. This is greater than the overfishing limit of 0.20, and therefore, overfishing is occurring.

The NDMBRPWG reached consensus that the stock status determination offered by the ASAP base model was preferred. However, given the retrospective pattern for the base model, alternative stock status determinations were conducted based on retrospective adjustments to F_{full} and SSB_{2010} to account for the relative model bias observed in the retrospective patterns over the past 5 years. Retrospective adjustments were accomplished using Equations 13 and 14.

$$(13) \quad SSB_{2010adjusted} = SSB_{2010} / (1 + \rho_{SSB})$$

$$(14) F_{full2010adjusted} = F_{mult2010} / (1 + \rho_F)$$

where:

ρ_{SSB} = Mohn's rho for spawning stock biomass (from Table A.68)

ρ_F = Mohn's rho for F_{full} (from Table A.68)

Accounting for the retrospective bias does not result in a change of stock status (Table A.70), though the revised stock status phase plot (Fig. A.161) shows that the revised point lies just inside the confidence intervals of the unadjusted point. The precedence established at GARM III (NEFSC 2008) was to only make retrospective adjustments when the adjusted point fell outside the confidence intervals of the unadjusted point. Based on the GARM III precedence, the SARC 53 Panel recommended that stock status determination should not be based on adjusted estimates of SSB and F.

For both the existing VPA model with respect to existing reference points and the new proposed ASAP base model with respect to updated reference points, the stock is overfished and overfishing is occurring. Consequently, for both models and reference point sets, the stock is not rebuilt.

TOR A.8. Develop and apply analytical approaches to conduct single and multi-year stock projections to compute the pdf (probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

TOR A.8.a. Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

Short term projections of future stock status were conducted based on the current assessment results without accounting for retrospective bias. This rationale was identical to that of stock status determination. Numbers-at-age in 2011 were derived from 1000 different vectors of numbers-at-age produced from the MCMC chain. Short term projections have assumed catch in 2011 to be equal to the catch in 2010. The NDMBRPWG concluded that this was a reasonable assumption given that the total ACLs in these two years were similar (2010=8,088 mt, 2011=8,545 mt).

Recruitment was sampled from a cumulative density function (CDF) of estimated age 1 recruitment from 1982 to 2008. The same AGEPRO model used for reference point determination was used to conduct short-term projections (i.e., model adjusts projected recruitment based on a linear function that declines to zero at zero SSB when SSB falls below 7.3 thousand mt). The NDMBRPWG did not support the use of hindcasted recruitment for the same reasons they rejected the historical ASAP sensitivity runs; recruitment estimates based solely on survey information have proven unreliable to use as the basis for stock determination. Projections were run under three different F assumptions: $F_0 = 0.00$, $F_{MSYproxy} = F_{40\%} = 0.20$, and $F_{75\%FMSY} = 0.15$.

Projection results are summarized in terms of median SSB and fishery catch (yield) under all three scenarios outlined above in Table A.71. Under even the most optimistic scenario in terms of rebuilding (F_0), the stock cannot rebuild to SSB_{MSY} by the current rebuilding date of 2014. Plots showing the most optimistic (F_0 , unadjusted) and pessimistic ($F_{40\%}$) scenarios in terms of rebuilding are shown in Figure A.162.

TOR A.8.b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

The major uncertainties are the moderate retrospective patterns that have been observed over the last five years. Given these patterns, there is additional uncertainty in catch advice based on these projections. Moreover, the projections will be sensitive to realized recruitment. Recent recruitment has been weak with no strong recruitment observed in the last twenty years. Continued weak recruitment will impede the ability for this stock to rebuild. Given the poor performance of past projections beyond a time period of two to three years, the longer term projections presented in this report should be considered highly uncertain.

TOR A.8.c. Describe this stock’s vulnerability (see “Appendix to the SAW TORs”) to becoming overfished, and how this could affect the choice of ABC.

Uncertainties that were not accounted for by assessment and reference point models were evaluated using model diagnostics. Standard model diagnostics (e.g., residual analyses, retrospective analyses) were used for model validation. Vulnerabilities that were not accounted for by assessment and reference point models were evaluated using exploratory modeling, habitat observations and preliminary results from studies conducted in the spawning closure areas. Those studies indicate strong site fidelity to the spawning grounds, and the almost immediate disruption of spawning activity when those areas are opened. This would suggest that area closures to protect spawning grounds is beneficial and could reduce vulnerability. Additional considerations of vulnerability and productivity are the implications of shifts in distribution, recruitment dynamics and increased natural mortality. Consumption of Atlantic cod by other fishes and mammals may be increasing as predator populations increase, however empirical evidence is lacking to support testing this hypothesis directly. A considerable source of additional vulnerability is the continued weak recruitment and low reproductive rate (e.g., recruits per spawner) of Gulf of Maine cod. If weak recruitment and low reproductive rate continues, productivity and rebuilding of the stock will be less than projected.

TOR A.9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments and review panel reports. Identify new research recommendations.

Previous from GARM I (October 2002)

- Explore a VPA formulation where autumn tuning indices are adjusted back to Jan 1, instead of shifted forward one year and one age.
 - *Unknown whether this was explicitly addressed during GARM II. This will not be explored in this benchmark, but alternate models (e.g., ASAP) which allow for explicit definition of survey timing will be explored.*
- Given the overall truncation in the age composition, investigate possible trends in size/age composition of the inshore versus offshore areas.
 - *Unknown whether this was explicitly addressed during GARM II. The size/age composition of the present period has expanded relative to the size/age composition observed during the mid/late-1990s.*
- Request the Methods Working Group to investigate means of deriving an appropriate sampling intensity for commercial landings.

- *NOAA Toolbox Biostat software includes an option to estimate CVs associated with the landings-at-age. This provides a precision-based approach to determining the sufficiency of the commercial biosampling effort.*
- Explore the use of the state of Maine – New Hampshire Inshore Trawl Survey as tuning indices.
 - *These surveys have not historically been used. There is no explicit age information available for this survey, and as such, no age-specific indices. The survey information was examined by the NDDWG, and specific avenues for further exploration are listed as a new research recommendation (see below).*

Previous from GARM II (August 2005)

- For the 2008 benchmark assessment use biological data from the Cod Industry Based Survey (IBS) in the Gulf of Maine.
 - *The previous assessment applied the ALK information to the recreational fishery; however, the age data are limited in their temporal coverage and the timing of the IBS does not coincide well with the recreational fishery. For this reason, these data were not used in the updated assessment.*
 - *Additionally, sampling of the commercial discards and landings was largely sufficient during the 2004 – 2007 period, such that the augmented information from the IBS has little utility.*
 - *The NDDWG did review the IBS data to corroborate the general presumptions on spawning activity in the Gulf of Maine. The IBS collected spawning condition male and females in the western Gulf of Maine during the March-May time period.*

Previous from GARM III (August 2008)

- As with Georges Bank cod, the Panel recommended that historical data be used to hindcast recruitments as far back in time as possible for use in the estimation of reference points and projections.
 - *This research recommendation was discussed by the Northern Demersal Models and BRP Working Group (NDMBRPWG). For the same reasons the group recommended against extending the base ASAP model out beyond years when age information was available, the group concluded that it was not appropriate to hindcast the recruitment time series.*

New from SAW 53

- Further pursue the incorporation of the Maine – New Hampshire Inshore Trawl Survey in future assessments. The unavailability of age information and short time series have precluded this survey from being used in past assessments. While age structures are currently collected from this survey, they have not been aged. The Data Working Group suggested exploration of the maturity information collected by this survey to examine agreement with the NEFSC maturity ogives.
- Examine the reproductive information collected from the Maine/New Hampshire inshore trawl survey for the early years (e.g., where Downeast Maine stations were sampled to evaluate whether any of the fish were mature and if it could possibly suggest the presence of a spawning aggregation.
- Examine historical and contemporary estimates of cod catch in the lobster fishery. Preliminary discussions with Maine DMR suggest that the lobster bycatch may be relatively small proportional to other fishery removals.

- Examine the impacts of excluding the Commercial LPUE index from the assessment. The Commercial LPUE index exists for the year 1982 – 1993 and is no longer updated. Regulations implemented since 1994 (e.g., trip limits, area closures) limit the utility of a LPUE index that extends beyond these years. Initial modeling to explore this recommendation indicated no impact to the updated VPA and negligible impact to the ASAP base model if the Commercial LPUE index is excluded. The NDMBRPWG therefore decided to drop the Commercial LPUE index from this, and all future assessments of Gulf of Maine cod.

References

- Ames EP. 2004. The stock structure of Atlantic cod in the Gulf of Maine. *Fisheries*, 29:10-28.
- Applegate A, Cadrin S, Hoenig J, Moore C, Murawski S, Pikitch E. 1998. Evaluation of existing overfishing definitions and recommendations for new overfishing definitions to comply with the Sustainable Fisheries Act. Final Report, Overfishing Definition Review Panel. New England Fishery Management Council, Newburyport, Massachusetts. 179 p.
- Beamish FWH. 1966. Vertical migration of demersal fish in the Northwest Atlantic. *Journal of the Fisheries Research Board of Canada*, 23: 109–139.
- Begg GA, Hare JA, Sheehan D. 1999. The role of life history parameters as indicators of stock structure. *Fisheries Research* 43:141-163.
- Brooks EN, Miller TJ, Legault CM, O'Brien L, Clark KJ, Garvaris S, Eeckhaute LV. 2010. Determining Length-based Calibration Factors for Cod, Haddock and Yellowtail Flounder. Transboundary Resource Assessment Committee (TRAC) Reference Document 2010/08. 23 p.
- Brown R. 2009. Design and field data collection to compare the relative catchabilities of multispecies bottom trawl survey conducted on the NOA ship Albatross IV and the FSV Henry B. Bigelow. NEFSC Bottom Trawl Survey Calibration Peer Review Working Paper. NEFSC. 19 p.
- Butterworth DS, Rademeyer RA. 2008a. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. *ICES J. Mar. Sci.* 65(9): 1717-1732.
- Butterworth DS, Rademeyer RA. 2008b. Further SCAA/ASPM Assessments of Gulf of Maine Cod Including Data for 2007 and Exploring the Impact of Age-Dependence in Natural Mortality. Groundfish Assessment Review Meeting (GARM) III Working Paper 1.F.b. 17 p.
- Forrester JRS, Byrne CJ, Fogarty M, Sissenwine MP, Bowman EW. 1997. Background papers on USA vessel, trawl, and door conversion studies. SAW-24 SARC Working Paper Gen. 6.
- Francis RICC. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68:1124-1138.
- Gunderson DR. 1997. Trade-off between reproductive effort and adult survival in oviparous and viviparous fishes. *Can. J. Fish. Aquat. Sci.* 54:990-998.
- Hart D, Miller T. 2008. Analyses of tagging data for evidence of decreased fishing mortality for large Gulf of Maine Cod, *Gadus morhua*. Groundfish Assessment Review Meeting (GARM) Biological Reference Point Meeting. April 29 – May 2, 2008. Working Paper 4.7. 12 p.
- Hewitt DA, Hoenig JM. 2005. Comparison of two approaches for estimating natural mortality based on longevity. *Fishery Bulletin*. 103(2):433-437.
- Hoenig JM. 1983. Empirical use of longevity data to estimate mortality rates. *Fishery Bulletin*. 82(1):898-903.
- Hoffman WS, Correia SJ, Pierce DE. 2006. Implementing the Industry-Based Survey for Gulf of Maine Cod Pilot Study. Final Report for NOAA Contract #: EA133F-03-CN-0109.

- Jensen AL. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. *Can. J. Fish. Aquat. Sci.* 53:820-822.
- King JR, Camisa MJ, Manfredi VM. 2010. Massachusetts Division of Marine Fisheries Trawl Survey Effort, Lists of Species Recorded, and Bottom Temperature Trends, 1978-2007. Massachusetts Division of Marine Fisheries Technical Report TR-38. 157 p.
- Legault, CM, Restrepo VR. 1998. A flexible forward age-structured assessment program. *ICCAT. Col. Vol. Sci. Pap.* 49:246-253.
- Legault CM. 2008. Technical Documentation for ASAP Version 2.0 NOAA Fisheries Toolbox (<http://nft.nefsc.noaa.gov/>).
- Loehrke J, Cadrin S. 2007. A review of tagging information for stock identification of cod off New England. Groundfish Assessment Review Meeting (GARM) Data Meeting. October 29-November 2, 2007. Working Paper 3b. 17 p.
- Mayo RK, O'Brien L, Buxton N. 1992. Discard estimates of American plaice, *Hippoglossoides platessoides*, in the Gulf of Maine northern shrimp fishery and Gulf of Maine-Georges Bank large-mesh otter trawl fishery. Appendix to NEFSC Ref. Doc. 92-07. 40 p.
- Mayo RK, Hesler TE, O'Brien L, Sosebee KA, Figuerido BF, Hayes DB. 1994. Estimation of standardized otter trawl effort, landings per unit effort, and landings-at-age for Gulf of Maine and Georges Bank cod. NMFS NEFSC Ref. Doc. 94-12. 17 p.
- Mayo RK. 1995. Assessment of the Gulf of Maine Cod Stock for 1993. NMFS NEFSC Ref. Doc. 95-02. 74 p.
- Mayo RK. 1998. Assessment of the Gulf of Maine Cod Stock for 1997. NMFS NEFSC Ref. Doc. 98-08. 88 p.
- Mayo RK, O'Brien L, Serchuk FM. 1993. Assessment of the Gulf of Maine Cod Stock for 1992. NMFS NEFSC Ref. Doc. 94-04. 54 p.
- Mayo RK, O'Brien L, Wigley SE. 1998. Assessment of the Gulf of Maine Cod Stock for 1998. NMFS NEFSC Ref. Doc. 98-13. 88 p.
- Mayo RK, Thunberg EM, Wigley SE, Cadrin SX. 2002. The 2001 Assessment of the Gulf of Maine Atlantic Cod Stock. NMFS NEFSC Ref. Doc. 02-02. 154p.
- Mayo RK, Col L. 2006. The 2005 Assessment of the Gulf of Maine Atlantic Cod Stock. NEFSC Ref Doc. 06-02. 109 p.
- Mayo RK, Sheperd G, O'Brien L, Col LA, Traver M. 2009. The 2008 assessment of the Gulf of Maine Atlantic cod (*Gadus morhua*) stock. NMFS NEFSC Ref. Doc. 09-03. 128 p.
- McIntyre TM, Hutchings JA. 2003. Small-scale temporal and spatial variation in Atlantic cod (*Gadus morhua*) life history. *Can. J. Fish. Aquat. Sci.* 60:1111-1121.

- Miller TJ, Das C, Politis PJ, Miller AS, Lucey SM, Legault CM, Brown RW, Rago PJ. 2010. Estimation of Albatross IV to Henry B. Bigelow calibration factors. NMFS NEFSC Ref. Doc. 10-05. 233 p.
- Nies T. 2011. Management Regulations. Stock Assessment Workshop 52 Southern Demersal Working Group 52 Working Paper 6.
- NEFSC (Northeast Fisheries Science Center). 1989. Report of the Seventh NEFC Stock Assessment Workshop (7th SAW). NMFS NEFSC Ref. Doc. 89-04. 108 p.
- NEFSC (Northeast Fisheries Science Center). 1991. Report of the Twelfth Northeast Regional Stock Assessment Workshop (12th SAW). NMFS NEFSC Ref. Doc. 91-03: 187 p.
- NEFSC (Northeast Fisheries Science Center). 1992. Report of the Thirteenth Northeast Regional Stock Assessment Workshop (13th SAW). NMFS NEFSC Ref. Doc. 92-02. 183 p.
- NEFSC (Northeast Fisheries Science Center). 1993. Report of the Fifteenth Northeast Regional Stock Assessment Workshop (15th SAW). Stock Assessment Review Committee (SARC) Consensus Summary of Assessments. NMFS NEFSC Ref. Doc. 93-06. 108 p.
- NEFSC (Northeast Fisheries Science Center). 1995. Report of the Nineteenth Northeast Regional Stock Assessment Workshop (19th SAW). Stock Assessment Review Committee (SARC) Consensus Summary of Assessments. NMFS NEFSC Ref. Doc. 95-08. 221 p.
- NEFSC (Northeast Fisheries Science Center). 1997. Report of the Twenty-fourth Northeast Regional Stock Assessment Workshop (24th SAW). Stock Assessment Review Committee (SARC) Consensus Summary of Assessments. NMFS NEFSC Ref. Doc. 97-12. 291 p.
- NEFSC (Northeast Fisheries Science Center). 1998. Report of the Twenty-seventh Northeast Regional Stock Assessment Workshop (27th SAW). Stock Assessment Review Committee (SARC) Consensus Summary of Assessments. NMFS NEFSC Ref. Doc. 98-15. 350 p.
- NEFSC (Northeast Fisheries Science Center). 2000. Assessment of 11 Northeast Groundfish stocks through 1999. Report of the Northern Demersal Working Group, Northeast Regional Stock Assessment Workshop. NMFS NEFSC Ref. Doc. 00-05. 173 p.
- NEFSC (Northeast Fisheries Science Center). 2001a. Assessment of 19 Northeast Groundfish stocks through 2000. Report of the Northern Demersal Working Group, Northeast Regional Stock Assessment Workshop. NMFS NEFSC Ref. Doc. 01-20. 217 p.
- NEFSC (Northeast Fisheries Science Center). 2001b. Northeast Fisheries Science Center. 2001. Report of the 33rd Northeast Regional Stock Assessment Workshop (33rd SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments NMFS NEFSC Ref. Doc. 01-18. 281 p.
- NEFSC (Northeast Fisheries Science Center). 2002a. Final report of the Working Group on re-evaluation of biological reference points for New England groundfish. NMFS NEFSC Ref. Doc. 02-04. 123 p.
- NEFSC (Northeast Fisheries Science Center). 2002b. Assessment of 20 Northeast groundfish stocks through 2001: a report of the Groundfish Assessment Review Meeting (GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts. October 8-11, 2002. NMFS NEFSC Ref.

Doc. 02-16. 521 p.

- NEFSC (Northeast Fisheries Science Center). 2005. Assessment of 19 Northeast groundfish stocks through 2004. 2005 Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts. August 15-19, 2005. NMFS NEFSC Ref. Doc. 05-13. 508 p.
- NEFSC (Northeast Fisheries Science Center). 2008. Assessment of 19 Northeast groundfish stocks through 2007. Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts. August 4-8, 2005. NMFS NEFSC Ref. Doc. 08-15. 884 p.
- O'Brien L, Burnett J, Mayo RK. 1993. Maturation of Nineteen Species of Finfish off the Northeast Coast of the United States, 1985-1990. NOAA Tech. Rep. NMFS 113: 66 p.
- O'Brien L, Esteves C. 2001. Update Assessment of American plaice in the Gulf of Maine-Georges Bank Region for 2000. NMFS NEFSC Ref. Doc. 01-02. 114 p.
- O'Brien L, Lough RG, Mayo RK, Junt JJ. 2005. Gulf of Maine and Georges Bank (NAFO Subareas 5 and 6). In Brander K. Spawning and life history information for North Atlantic cod stocks. ICES Cooperative Research Report No. 274. 162 p.
- Palmer MC, Hawkins AE, Traver M, Brooks EN. 2011. A review of factors affecting the survival of Gulf of Maine Atlantic cod (*Gadus morhua*). SAW 53 Gulf of Maine Atlantic Cod Data Working Group. Working Paper 1. 29 p.
- Palmer MC, Wigley SE. 2007. Validating the stock apportionment of commercial fisheries landings using positional data from Vessel Monitoring Systems (VMS). NMFS NEFSC Ref. Doc. 07-22; 35 p.
- Palmer MC, O'Brien L, Wigley S, Mayo R, Rago P, Hendrickson L. 2008. A brief overview of discard estimation methods where observer coverage is unavailable. Groundfish Assessment Review Meeting (GARM) Reference Points Meeting. April 29 – May 2, 2008. Working Paper 4.5. 13 p.
- Palmer MC, Wigley SE. 2009. Using positional data from vessel monitoring systems to validate the logbook-reported area fished and the stock allocation of commercial fisheries landings. N. Amer. J. Fish. Manag. 29:928-942.
- Palmer MC, Wigley SE. 2011. Using positional data from vessel monitoring systems to validate the logbook-reported area fished and the stock allocation of commercial fisheries landings, 2004-2008. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 9.
- Pauly D. 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. J. Cons. Int. Explor. Mer. 39(2):175-192.
- Penttila JA, Gifford VM. 1976. Growth and mortality rates for cod from the Georges Bank and Gulf of Maine areas. International Commission for the Northwest Atlantic Fisheries Research Bulletin. 12:29-36.
- Pope JA, Margetts AR, Hamley JM, Aykuz EF. 1975. Manual of methods for fish stock assessment Part III: Selectivity of fishing gear. FAO Fish. Tech. Pap. (41) Rev. 1.
- Rivard D, 1980. APL programs for stock assessment. Can. Tech. Rep. Fish. Aquat. Sci. 953: 103 p.

- Rivard D, 1982. APL programs for stock assessment (revised). Can. Tech. Rep. Fish. Aquat. Sci. 1091. 146 p.
- Serchuk FM, Wigley SE. 1986. Assessment and status of the Georges Bank and Gulf of Maine Atlantic cod stocks – 1986. NMFS NEFSC Ref. Doc. 86-12. 84 p.
- Serchuk FM, Grosslein MD, Lough RG, Mountain DG, O’Brien L. 1994. Fishery and environmental factors affecting trends and fluctuations in the Georges Bank and Gulf of Maine Atlantic cod stocks – an overview. ICES Mar. Sci. Symp. 198:77-109.
- Sherman SA, Stepanek K, Sowles J. 2005. Maine – New Hampshire Inshore Groundfish Trawl Survey: Procedures and Protocols. Available from: <http://www.maine.gov/dmr/rm/rawl/reports/proceduresandprotocols.pdf>. 42 p.
- Sherman SA, Stepanek KL, Gowen AM, Sowles J, Tetrault R, Michael R, Zobel R. 2010. Annual Report on the Maine-New Hampshire Inshore Trawl Survey: January 1, 2009-December 31, 2009. NOAA Contract # NA07NMF4720357.
- Tallack S. 2007. A description of tagging data from the Northeast Regional Cod Tagging Program (WPA3A) and preliminary applications of weighting and mixing analysis (WP3C). Groundfish Assessment Review Meeting (GARM) Data Meeting. October 29-November 2, 2007. Working Paper 3a and c. 17 p.
- Tallack S. 2009. Proceedings from a workshop to identify future research priorities for cod tagging in the Gulf of Maine. NMFS NEFSC Ref. Doc. 09-09. 76 p.
- Wigley, S.E. 1996. The Lorenz curve method applied to NEFSC bottom trawl survey data. NMFS NEFSC Ref. Doc. 96-05f. 11 p.
- Wigley SE, McBride HM, McHugh NJ. 2003. Length-weight relationships for 74 fish species collected during NEFSC research vessel bottom trawl surveys, 1992-9. NOAA Tech Memo NMFS NE 171. 26 p.
- Wigley SE, Rago PJ, Sosebee KA, Palka DL. 2007. The Analytic Component to the Standardized Bycatch Reporting Methodology Omnibus Amendment: Sampling Design, and Estimation of Precision and Accuracy (2nd Edition). NMFS NEFSC Ref. Doc. 07-09. 156 p.
- Wigley SE, Hersey P, Palmer JE. 2008. A description of the allocation procedure applied to the 1994 to 2007 commercial landings data NMFS NEFSC Ref. Doc. 08-18. 61 p.
- Wigley SE, Palmer MC, Legault C. 2011. A comparison of discard rates derived from at-sea monitoring and observer trips. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 11.
- Wileman DA, Ferro RST, Fonteye R, Millar RB (eds.). 1996. Manual of methods of measuring the selectivity of towed fishing gears. ICES Cooperative Research Report. No. 215.
- Wood SA. 2001. [Report to MERI: Summary of Harbor Seal \(*Phoca vitulina concolor*\) Food Habits in Mid-Coast Maine: Summer 2000](#). Marine Environ. Research Institute, Blue Hill, ME, 12 pp.