New England Fishery Management Council's New Risk Policy Demonstration and Evaluation

Lisa Kerr, Roger Brothers, Alexandra Schneider

University of Maine

1. Executive Summary

As of 2025, the New England Fishery Management Council has implemented a new Risk Policy to inform fisheries management decisions. We evaluated the scoring rubric and applied the Risk Policy to the 22 stocks in the Northeast Multispecies Fishery Management plan. The resulting scores were weighted and summed to calculate Z-Scores (an intermediate quantity within the Risk Policy), which were used in a logistic function to calculate the recommended probability of management success (quantitative Risk Policy output) that correspond to each set of scores and weights. We demonstrated the sensitivity of the Risk Policy performance to various structural assumptions using multiple scenarios that altered factor weighting schemes (uniform or Council defined), the scaling of factor scores (maximum of 1, 2, or 4), the possible range of scores for each factor (some factors can have negative scores while others cannot), and the number of factors scored (including commercial and recreational fishery characterization and excluding fish condition). We also demonstrated the maximum and minimum score stocks could receive based on its assessment type.

This evaluation is the first phase of a project aimed at simulation testing the integration of the Risk Policy into harvest control rules. The Risk Policy will be used to incorporate uncertainty and risk into fisheries management decisions and potentially to inform catch advice by integrating the quantitative output with harvest control rules. The new Risk Policy aims to increase transparency and incorporate a broader set of considerations in decision-making, including climate uncertainty and socioeconomic factors. Some challenges remain, including finalizing the details of scoring, accurately capturing the influence of socioeconomic and climate conditions, and communication of the Risk Policy. This scoring demonstration and sensitivity analysis can inform the continued development of the Risk Policy and will be used to guide Phase II of the project, simulation testing the integration of the Policy into potential harvest control rules.

A brief summary of the recommendations from Phase I are as follows:

<u>Scoring Assumptions:</u> We recommend that 1) rubrics for factor scoring are specific and quantitative in nature, and 2) that the availability and recency of the data needed to score factors are considered in finalized rubrics.

Different score ranges across factors: We recommend that 1) potential implicit weighting, dependencies among factors, the difference in possible score ranges for empirically and analytically assessed stocks, and the unequal ranges of scores for each factor are considered in the context of the Council's goals, 2) Z-Scores be scaled to 4 so full access to the logistic curve is possible.

Harvest Control Rule Integration: The impact of a specific Z-Score and resulting recommended probability will be defined by how they are integrated into a harvest control rule. Therefore, we recommend that the shape of the logistic curve (e.g., the faster rate of change at low recommended probabilities) and the position of a neutral score for each factor and on the logistic curve more broadly are aligned with management goals.

Integrating changing environmental conditions into the Council's assessment of risk: We recommend that the stationarity of the environmental factors are considered (i.e., climate vulnerability scores are only updated with new climate vulnerability analyses) and the factors utility as a proxy for environmental conditions (e.g., fish condition) are clearly outlined and considered.

Developing a clear path to incorporate social and economic considerations: We recommend that the results of the fishery characterization scenario and the conclusions of the scoring demonstration (i.e., specific and quantitative rubrics and data sources used in the rubrics are available and updated) are considered for the continued development of the fishery characterization factors.

Transparency in Decision Making: We recommend that the language of the Risk Policy is streamlined for interpretability. Interpretability could be increased by keeping factor scores and stock condition intuitively aligned, using consistent and specific and consistent terminology to refer to risk (i.e., tolerance or aversion), and reframing "recommended probability of management success," which is difficult to conceptualize.

2. Background

In U.S. fisheries management, each Regional Management Council has an overarching Risk Policy that provides strategic guidance on how to consider risk and uncertainty in decision making. These policies prescribe approaches for identifying risks, analyzing their likelihood and potential consequences, and determining how risk averse or risk tolerant the Council should be in different situations. There are a host of different decisions made by the Councils that consider the Risk Policy at different scales (e.g., stock-specific catch advice and spatial restrictions to fisheries).

2.1. Updated NEFMC Risk Policy

The New England Fishery Management Council (NEFMC) recently developed a substantially revised Risk Policy which was approved in September 2024 and implemented on January 1, 2025. The purpose of the Risk Policy is to account for inherent uncertainty related to fisheries management during decision making. While the previous NEFMC Risk Policy (NEFMC 2016) was qualitative in nature, the revised policy considers both qualitative and quantitative information to generate a measure of recommended risk tolerance that can be applied in either a qualitative or quantitative manner. The implementation is being carried out in two phases: Alpha and Beta. The Alpha phase involves using the Risk Policy Statement in a qualitative manner to inform decisions, providing decision makers with a risk policy matrix for each stock based on the revised structure. The Beta phase, which will extend into 2026, involves updating and refining the quantitative details of the policy and simulation testing its performance.

A key feature of the Council's new Risk Policy are the factors used in characterizing risk. Seven factors were identified by the Risk Policy Working Group that span three categories: 1) stock status and uncertainty, 2) climate and ecosystem considerations, and 3) economic and community importance (Table 1). The broad category of *Stock Status and Uncertainty* includes three factors: biomass (i.e., spawning stock biomass, SSB), recruitment, and assessment type and uncertainty. The *Climate and Ecosystem* C*onsiderations* category includes two factors: climate vulnerability and fish condition. Lastly, the *Economic and Community Importance* category includes two factors, one each to characterize the commercial and recreational fisheries independently.

Factors are to be weighted and scored (NEFMC 2024). Weighting, which will be done by the Council, represents a policy choice to define the relative importance of each factor when determining risk. Scoring, which will be done by Plan Development Teams according to a rubric and prescribed information sources, is designed to objectively characterize the state of each stock. Currently, it is recommended that the Council weights the factors every three years (NEFMC 2025), while scoring would occur more frequently. The scores and weights will be multiplied for each factor individually and then aggregated to calculate an intermediate Z-Score.

The Z-Score is used to calculate a quantitative measure of risk aversion (Figure 1). This quantitative output of the Risk Policy is defined as the recommended probability of achieving a management goal (i.e., not overfishing) and reflects the intended degree of risk aversion needed when making management decisions: stocks with a high Z score require a high degree of risk aversion when making decisions (Figure 1). The function is truncated at 50% because the Magnuson-Stevens Act requires that the probability of overfishing does not exceed 50% (National Standard 1, 16 U.S.C. §§ 1801 et seq).

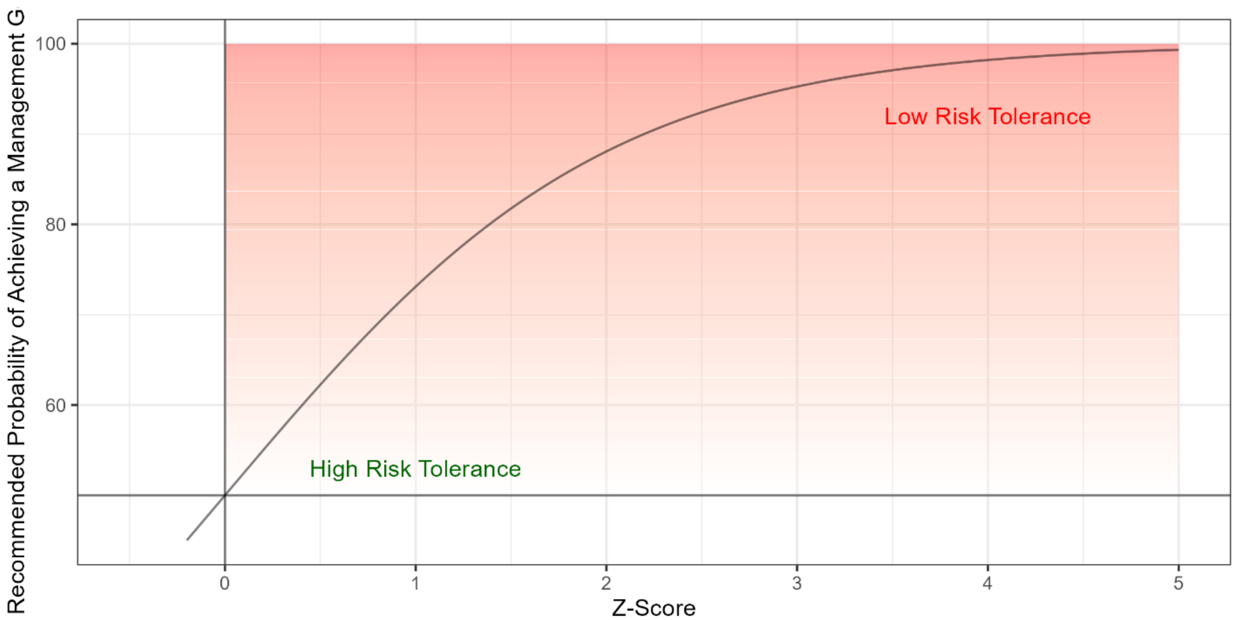


Figure 1. Relationship between the Z-Score and recommended probability of achieving a management goal.

Table 1. Factors included in the NEFMCs new Risk Policy and source materials identified to score factors.

Category	Factor	Representative of	Source Material
Stock Status and Uncertainty	Biomass	Current Productivity	Current stock assessment
	Recruitment	Future productivity	Current stock assessment
	Assessment type and uncertainty	Assessment performance	Current stock assessment
Climate and Ecosystem Considerations	Climate vulnerability	Vulnerability to environmental change	Climate vulnerability analysis (Hare et al. 2016)
	Fish condition	Ecosystem productivity	State of the Ecosystem report
Economic and Community	Commercial fishery characterization	Commercial fishery performance	Revenue, market value, lease value information from PDTs/NEFSC social science branch
	Recreational fishery characterization	Recreational fishery presence and performance	State of the Ecosystem report, Marine Recreational Information Program

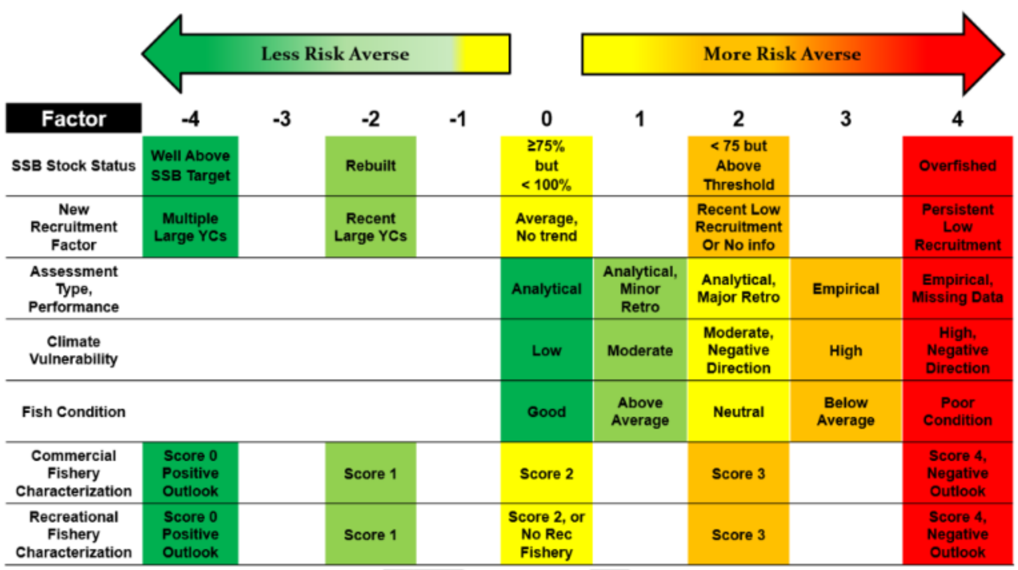
2.2 Goals and Objectives

Our goal was to evaluate the Council's new Risk Policy as the first phase of a project that will simulation test the performance of Risk Policy integrated harvest control rules. In this first phase, our specific objectives were to 1) qualitatively evaluate the Council's identified risk factors and proposed source materials for scoring, 2) execute factor scoring and risk tolerance calculations for groundfish stocks, and 3) explore the sensitivity of quantitative risk policy output to important decisions (e.g., weighting, scaling, the inclusion or exclusion of factors) and stock characteristics (e.g., empirical vs. analytical assessment). Our findings will be used to guide simulation testing of harvest control rules that integrate the Risk Policy and may inform refinement of the Risk Policy before quantitative implementation.

3. Methods

3.1 Factor Scoring

In the 2025 version of the NEFMC's Risk Policy, each factor had possible scores that ranged from either –4 to 4 or 0 to 4 (Figure 2) and defined criteria for scoring (NEFMC 2025). Negative scores represented characteristics of the stock or fishery that support more risk tolerance while positive scores represented situations that require more caution. For all stocks in the Northeast Multispecies (groundfish) Fisheries Management Plan (n = 22) we scored each factor using the most recent year of data and characterized them dynamically back in time, when possible. Aspects of the factor rubrics were open to interpretation; therefore, we made assumptions in certain cases to complete scoring. All assumptions and modifications are listed below by factor. We did not score the commercial or recreational fishery characterization factors because the rubrics were under development at the time of this analysis.



Factor	-4	-3	-2	-1	0	1	2	3	4
SSB Stock Status	Well Above SSB Target		Rebuilt		≥75% but < 100%		< 75 but Above Threshold		Overfished
New Recruitment Factor	Multiple Large YCs		Recent Large YCs		Average, No trend		Recent Low Recruitment Or No info		Persistent Low Recruitment
Assessment Type, Performance					Analytical	Analytical, Minor Retro	Analytical, Major Retro	Empirical	Empirical, Missing Data
Climate Vulnerability					Low	Moderate	Moderate, Negative Direction	High	High, Negative Direction
Fish Condition					Good	Above Average	Neutral	Below Average	Poor Condition
Commercial Fishery Characterization	Score 0 Positive Outlook		Score 1		Score 2		Score 3		Score 4, Negative Outlook
Recreational Fishery Characterization	Score 0 Positive Outlook		Score 1		Score 2, or No Rec Fishery		Score 3		Score 4, Negative Outlook

Figure 2. Range of scores for risk policy factors (NEFMC 2025)

To score factors that use recent stock assessment results (spawning stock biomass, recruitment, assessment type and uncertainty) we used the *stocksmart* package in R (Beet 2025) to retrieve assessment data from NOAA's Fisheries' Stock Status, Management, Assessment, and Resource Trends web tool (https://apps-st.fisheries.noaa.gov/stocksmart?app=homepage). As prescribed in the Risk Policy, we used the findings of Hare et al. (2016) to score climate condition and we used the results of the most recent State of the Ecosystem Report to score fish condition (NEFSC 2025a, 2025b). Spawning stock biomass, recruitment, and fish condition were characterized dynamically over time.

3.1.1 Spawning stock biomass:

The spawning stock biomass (SSB) factor was scored by comparing the terminal year SSB to the SSB reference points from the most recent stock assessment (i.e., SSB threshold = 0.5 SSB_{MSY} and SSB target = SSB_{MSY}, Figure 3). This factor can both increase and decrease risk tolerance, with 5 possible scores ranging from –4 to 4 (i.e., -4, -2, 0, 2, 4). Stocks with known status, including those with analytical assessments and some with empirical assessments can receive any of the 5 possible scores, whereas stocks with empirical assessments and unknown status because they lack SSB reference points were limited to scores of 0, 2, or 4. When the terminal SSB was ≥ 150% of the SSB target, the stock productivity supported lower risk aversion and received a score of –4. When a stock was considered overfished, because the terminal SSB was below the threshold reference point, or if the SSB status was unknown but the stock index had a negative 5-year trend, the stock received a score of 4.

For stocks with unknown SSB status, we used linear regression to determine trends. If the slope parameter was not significant (at $\alpha \geq 0.05$) the trend was considered neutral, and the stock received a score of 2. For significant slopes ($\alpha < 0.05$), the stock received a 0 for a positive slope parameter, or a 4 for a negative slope parameter.

Original Rubric:

Score	-4	-2	0	2	4
	Well above SSB Target	Rebuilt	≥ 75% but < 100%	< 75% but Above Threshold	Overfished
Description	≥ 150% SSB target	≥ SSB target but < 150% SSB target	≥ 75% SSB target but < 100% SSB target OR Unknown status and positive 5-year trend	< 75% SSB target OR Unknown status and 5-year trend is neutral/no trend	Below SSB threshold OR Unknown status and negative 5-year trend

Modified Rubric for scoring:

Score	-4	-2	0	2	4
	Well above SSB Target	Rebuilt	≥ 75% but < 100% Or unknown with	< 75% but Above Threshold	Overfished
Description	≥ 150% SSB target	≥ SSB target but < 150% SSB target	≥ 75% SSB target but < 100% SSB target OR Unknown status and positive, significant (α < 0.05) 5-year linear trend	< 75% SSB target OR Unknown status and 5-year linear trend is neutral/not significant (α ≥ 0.05)	Below SSB threshold OR Unknown status and negative, significant (α < 0.05) 5-year linear trend

Figure 3. Original and modified rubric guide for biomass factor (NEFMC 2025). The modified rubric was used for factor scoring.

3.1.2 Recruitment

The recruitment factor aims to assess the future productivity of the stock. Scoring was primarily based on comparing the magnitude of recruitment over the most recent 5 years to an average. This factor can both increase and decrease risk tolerance, with 5 possible scores ranging from –4 to 4 in increments of 2 (i.e., -4, -2, 0, 2, 4). A score of -4 represents multiple large year classes and a score of 4 represents persistent low recruitment in the past 5 years (Figure 4). We used the abundance of age 1 fish, as estimated by the most recent stock assessment, to characterize recruitment comparing the annual values in the most recent five years to the mean of the entire stock assessment timeseries. For stocks with empirical assessments, the score defaulted to 2 because these stock assessment methods do not estimate recruitment.

We had to modify the rubric slightly because the criteria for score of -2 and 2 were not mutually exclusive. The original rubric specified that over the last five years, if there were two years with above average recruitment the score should be -2 and that if there were two years with below average recruitment the score should be 2. Georges Bank haddock and Gulf of Maine haddock both met the criteria for a score -2 and 2 simultaneously. Therefore, we changed the criteria for a score of 2 to require lower than average recruitment in at least 4 of the last five years. We also added criteria for a score of 0 to further define which stock assessment decisions we assumed to account for recent recruitment changes in reference points and stock projections; we defaulted to 0 when reference points assumed recent average recruitment, projections assumed temporal autocorrelation in age-1 abundance, or projections assumed a stock-recruit relationship.

Based on the original rubric, and some necessary modifications, we assumed:

- Score of -4: There have been multiple large year classes, above the timeseries mean, in three or more of the last five years.
- Score of –2: There have been two large year classes, above the timeseries mean, in the last five years.
- Score of 0: Recruitment in the last five years is average or reference points assume recent average recruitment, projections assume temporal autocorrelation in age-1 abundance, or projections assume a stock-recruit relationship.
- Score of 2: There has been low recruitment, below the timeseries mean, in four or more of the last five years.
- Score of 4: There has been persistent low recruitment, i.e., recruitment below the timeseries mean, for the last 6 years

Original Rubric:

Score	-4	-2	0	2	4
	Multiple large Year Classes	Recent Large Year Classes	Average, No Trend	Recent Low Recruitment	Persistent Low Recruitment
Description	There have been multiple large (meaning above average) in the last five years	There has been two large (meaning above average) recruitment events in the last five years	Recruitment in the last five years is average OR recent changes in recruitment have been accounted for in reference points and or stock projections	Low (meaning below average) recruitment in at least two of the last five years OR there is no information on recruitment	Persistent low (meaning below average) recruitment for more than five years

Modified Rubric for scoring:

Score	-4	-2	0	2	4
	Multiple large Year Classes	Recent Large Year Classes	Average, No Trend	Recent Low Recruitment	Persistent Low Recruitment
Description	There have been multiple large (meaning above the timeseries average) in three or more of the last five years	There has been two large (meaning above the timeseries average) recruitment events in the last five years	Recruitment in the last five years is average **OR** recent reference points assume temporal autocorrelation in age-1 abundance, or projections assume a stock-recruit relationship	Low (meaning below the timeseries average) recruitment four or more of the last five years OR there is no information on recruitment	Persistent low (meaning below the timeseries average) recruitment for the last six years

Figure 4. Original and modified rubric guide for recruitment factor (NEFMC 2025). The modified rubric was used for factor scoring.

3.1.3 Assessment type and uncertainty

The assessment type and uncertainty factor considers the type of assessment (i.e., state-space, analytical or empirical) and the level of uncertainty as indicated by the magnitude of the retrospective pattern or the number of uncertainties noted in the assessment report. This factor can only decrease risk tolerance, with 5 possible scores ranging from 0 to 4 in increments of 1 (i.e., 0, 1, 2, 3, 4; Figure 5). Analytical stocks can have a score of 0, 1, or 2, while empirical stocks can have a score of 3 or 4. A score of 0 represents an analytical assessment with no retrospective pattern or a state-space model with limited sources of uncertainty listed in the assessment report. A score of 4 is assigned to a stock with an empirical assessment that is missing data in at least one of the three most recent years.

Scores 0 to 2 have an “OR” statement depending on whether the assessment is analytical or state-space. We modified portions of the original rubric to accommodate both non-state-space analytical assessments and state-space assessment types (e.g.., WHAM). For state-space assessments, the factor score was determined solely by the number of uncertainties in the assessment report. In contrast, for stocks with analytical assessments that are not state-space (e.g., ASAP), the score was based exclusively on the magnitude of retrospective pattern. The level of retrospective pattern (major or minor) was determined based on statements in the assessment reports.

Original Rubric:

Score	0	1	2	3	4
	Analytical	Analytical, Minor Retro	Analytical, Major Retro	Empirical	Empirical, Missing Survey Data
Description	Analytical assessment with no retrospective pattern, OR state-space model with limited sources of uncertainty as described in assessment report	Analytical assessment with minor retrospective pattern OR state-space model with at least two sources of uncertainty as described in assessment report	Analytical assessment with major retrospective pattern OR state-space model with at least three significant sources of uncertainty as described in assessment report	Empirical assessment approach	Empirical assessment approach with missing data in one of the three most recent years

Modified Rubric for scoring:

Score	0	1	2	3	4
	Analytical	Analytical, Minor Retro	Analytical, Major Retro	Empirical	Empirical, Missing Survey Data
Description	Non-state space analytical assessment with no retrospective pattern **OR** state space analytical assessment (e.g., WHAM) that may have any level of retrospective patterns, but only 0 or 1 source of uncertainty	Non-state space analytical assessment with minor retrospective pattern **OR** state space analytical assessment (e.g., WHAM) that may have any level of retrospective patterns, but 2 or more sources of uncertainty	Non-state space analytical assessment with major retrospective pattern **OR** state space analytical assessment (e.g., WHAM) that may have any level of retrospective patterns, but 3 or more sources of uncertainty	Empirical assessment approach with no missing data in any of the 3 most recent years	Empirical assessment approach with missing data in one of the three most recent years

Figure 5. Original and modified rubric guide for the assessment type and uncertainty factor (NEFMC 2025). The modified rubric was used for factor scoring.

3.1.4 Climate vulnerability

The climate vulnerability factor can only decrease risk tolerance, with 5 possible scores ranging from 0 to 4 in increments of 1 (i.e., 0, 1, 2, 3, 4). A score of 0 indicates low climate vulnerability and a score of 4 indicates high climate vulnerability with a negative directional effect of climate (Figure 6), thereby decreasing risk tolerance with increasing climate concerns. The scoring for this factor is based on species assessments from Hare et al. (2016).

Score	0	1	2	3	4
	Low Vulnerability	**Moderate Vulnerability**	**Moderate Vulnerability, Negative Direction**	**High Vulnerability**	**High Vulnerability, Negative Direction**
Description	"Low" vulnerability score	"Moderate" vulnerability score **OR** "Low" vulnerability score and negative climate directional effect	"Moderate" vulnerability score and negative climate directional effect	"Very high" or "high" vulnerability score	"Very high" or "high" vulnerability score and negative climate directional effect

Figure 6. Original rubric guide for the climate vulnerability factor (NEFMC 2025). The original rubric was used for factor scoring.

3.1.5 Fish condition

The purpose of the fish condition factor is to serve as a proxy for ecosystem productivity. Theoretically, fish condition can reflect overall ecosystem health and prey availability, providing a link between environmental change directly to fish performance.

To assess this factor we used data from the most recent State of the Ecosystem report (NOAA 2025),retrieved via the ecodata package in R (Beltz et al. 2025). The fish condition factor can only decrease risk tolerance, with 5 possible scores ranging from 0 to 4 in increments of 1 (i.e., 0, 1, 2, 3, 4; Figure 7). A score of 0 represented good condition and a score of 4 represented poor condition, thereby decreasing risk tolerance along with decreasing fish condition. A stock received a score of 0 when the state of the ecosystem reports "good condition" for a majority of the three most recent years. Condition factors of "above average", "neutral", "below average", and "poor condition", received scores of 1, 2, 3, and 4, respectively. If fish condition data are not available, the stock receives a score of 2.

The State of the Ecosystem condition ratings are based on Ecological Production Units (EPU), as defined in the report (NEFSC 2025a, b), rather than stock area. Therefore, we made assumptions about the overlap between stocks and EPUs to score fish condition. We used stock areas and distribution maps via NOAA DisMAP (https://apps-st.fisheries.noaa.gov/dismap/DisMAP.html) to qualitatively assign stocks to an EPU, or multiple EPUs if appropriate.

Score	0	1	2	3	4
	Good Condition	Above Average	Neutral	Below Average	Poor Condition
Description	In three most recent years of available data, majority of boxes are scored as "good condition"	In three most recent years of available data, majority of boxes are scored as "above average condition"	In three most recent years of available data, majority of boxes are scored as "neutral condition" **OR** no information on fish condition	In three most recent years of available data, majority of boxes are scored as "below average condition"	In three most recent years of available data, majority of boxes are scored as "poor condition"

Figure 7. Original rubric guide for the fish condition factor (NEFMC 2025). The original rubric was used for factor scoring, with assumptions made about the assignment of stocks to ecological production units.

3.1.6 Commercial fishery characterization

The criteria for scoring the commercial fishery characterization had not been formalized at the time of this report, so we did not demonstrate empirical scoring. Instead, we used scenarios (described below) to explore the sensitivity of Risk Policy calculations to the proposed possible scores, which include 5 scores ranging from -4 to 4 in increments of 2 (i.e., -4, -2, 0, 2, 4) allowing this factor to both increase and decrease risk tolerance (Figure 2).

3.1.7 Recreational fishery characterization

The criteria for scoring the recreational fishery characterization had not been formalized at the time of this report, so we did not demonstrate empirical scoring. Instead, we used scenarios (described below) to explore the sensitivity of Risk Policy calculations to the proposed possible scores, which include 5 scores ranging from -4 to 4 in increments of 2 (i.e., -4, -2, 0, 2, 4) allowing this factor to both increase and decrease risk tolerance (Figure 2).

3.2 Weighting:

In addition to being scored, the factors are weighted. As prescribed by the Risk Policy (NEFMC 2024), the Council members define factor weights by assigning a relative importance to each factor, on a scale of 0 to 4, with 0 representing the lowest level of importance and 4 representing the highest level of importance. Council members conducted a mock weighting exercise in April, 2025 and we used these values in our demonstration. Individual Council members rated the importance of each factor but could not use a rating of 4 for more than three factors. To calculate the weight of each factor, the mean rating was calculated for each factor and normalized so that the weights of all factors sum to one. These normalized values represent the council weightings for each factor (Table 2). Note, because not all factors were included in the risk policy demonstration (i.e. recreational and commercial fishery characterizations) the council weights were renormalized and scaled to 1 for the subset of factors included in each scenario.

3.3 Z-Scores, and Recommended Probabilities

The Z-Score was calculated as $Z = \Sigma(w_i s_i)$, where w_i represented factor weights and s_i represented factor scores. This Z-Score was then used in a logistic function to calculate the recommended probability of management success as: $p(Z) = \frac{1}{1+e^{-Z}}$.

3.4 Baseline scenario and sensitivity explorations

We explored aspects of the Risk Policy by scoring a baseline scenario and then exploring the sensitivity of the Z-Scores and recommended probabilities to a variety of structural assumptions. These included the scaling of maximum potential Z-Scores, alternative factor weightings, the range of possible scores for each factor scores, the inclusion and exclusion of factors, and the type of assessment a stock has.

3.4.1 Five Factor Scenario-Baseline:

The baseline scenario implements the risk policy and rubrics as outlined and incorporated scoring of the five factors with well-defined rubrics: SSB, recruitment, stock assessment and uncertainty, climate vulnerability and fish condition. The baseline scenario weighted factor scores using values from the Council's mock weighting exercise.

3.4.2 Impact of Z-Score Scaling with Five Factors:

We explored two additional scaling scenarios to test the impact the scale of scores has on the Z-Scores and recommended probabilities. In these scenarios, the five factors in the baseline scenario were: 1) divided by 2 so that factor scores ranged from -2 to 2 or 0 to 2, and 2) divided by 4 so that factor scores ranged from -1 to 1 or 0 to 1. These scaling scenarios produced maximum Z-Scores of 2 and 1, respectively.

3.4.3 Impact of Weighting with Five Factors:

To test the sensitivity of the Risk Policy to factor weightings this analysis considered a uniform weighting option where all factors are considered equally import, rather than the Council weighting option.

3.4.4 Impact of Potential Score Ranges:

To explore the impacts of unequal scoring ranges for factors we tested a scenario where all five baseline factors had the same range of possible scores and equal increments between them. In this scenario all scores increased by one unit and ranged from 0 to 4, rather than some factors ranging from 0 to 4 and increasing by increments of one and other factors ranging from -4 to 4 and increasing increments of two. We applied each weighting option (uniform or council) to this scenario.

3.4.5 Impact of Assessment Type:

We also conducted sensitivity analyses to examine the possible range of scores for stocks with empirical or analytical assessments. Multiple factors have restrictions on possible scores based on the assessment type. Therefore, these scenarios demonstrate the maximum and minimum potential score for stocks under each assessment type. To do this, we used the same five factors as the baseline scenario to calculate the potential Risk Policy output for analytical and empirical stocks separately. For each factor we used the highest and lowest possible score available to stocks with each stock assessment type and applied both weighting options (uniform or council).

3.4.6 Impact of including fishery factors:

Final factor scoring criteria were still being developed for commercial and recreational fishery characterizations at the time of this report. Therefore, we scored the five factors with established rubrics and explored the sensitivity of the Risk Policy output to including the two additional fishery factors. To understand the potential impact of including all seven factors, we considered scenarios that included minimum (-4), intermediate (0), and maximum (4) scores for both fishery factors (Figure 2). For these scenarios, we considered the Council and uniform weighting options and scaling with a maximum Z-Score of 4. We compared the outcome of the seven factor scenarios to the five factor baseline scenario with a maximum Z-Score of 4.

3.4.7 Impact of excluding fish condition:

Lastly, we demonstrated a reduced factor scenario that excluded fish condition. Fish condition was removed because the link between fish condition and ecosystem condition is not well understood (see 4.2.5 for a full discussion of factor limitations). This scenario includes the four factors with established rubrics: SSB, recruitment, assessment type and uncertainty, and climate vulnerability. For this scenario, we applied the Council and uniform weighting options and scaling with a maximum Z-Score of 4.

3.4.8 Six Factor Scenarios (fishery factors included & fish condition factor excluded):

We also explored a six factor scenario that excluded fish condition but includes the maximum, intermediate and minimum scores for the commercial and recreation fishery characterization factors. For these scenarios, we applied the Council and uniform weighting options and scaling with a maximum Z-Score of 4.

Table 2. Council defined factor weights. SSB = spawning stock biomass.

	Risk Policy Mock Factor Weights (NEFMC mean)						
Scale	**SSB**	**Recruits**	**Assessment**	**Climate**	**Fish Cond.**	**Commercial**	**Recreational**
Mean (0-4)	3.29	3.00	2.86	2.95	2.00	3.24	1.90
Normalized, Sum(weights) = 1	0.17	0.16	0.15	0.15	0.10	0.17	0.10

4. Results and Discussion:

4.1 Demonstrated Factor Scores:

We scored 22 groundfish stocks within the Northeast Multispecies Fishery Management Plan across the five factors that have established rubrics.

4.1.1 Spawning Stock Biomass Status

Of the 22 groundfish stocks, 17 have a known stock status. Four stocks were well above the target and received a score of -4 (Figure 8). Two stocks were rebuilt and received a score of -2. Five stocks were less than the SSB target but above the overfished threshold to varying degrees, receiving a score of either 0 or 2. Six stocks were classified as overfished and received a score of 4. The remaining five stocks have unknown status. Two of these had a statistically significant, negative recent trend, receiving a score of 4, while the other three did not have a statistically significant trend and received a score of 2.

Scores of the SSB stock status factor over time revealed that a few stocks had consistently poor stock status over time (e.g., Atlantic halibut, ocean pout, northern windowpane, Figure 9), while other stocks have highly variable scores (e.g., northern white hake, and Georges Bank winter flounder). Trends in SSB over time were generally associated with an improvement in SSB status as indicated by an increasingly negative score (e.g., Gulf of Maine haddock, Acadian redfish, American plaice). In some cases, SSB scores fluctuated substantially between assessments (e.g., fluctuating by up to 8 points for Georges Bank haddock between assessments).

4.1.2 Recruitment

Half of the groundfish stocks (n =11) were scored as having either recent low or persistent low recruitment (i.e., score 2 and 4), although seven of the 11 stocks defaulted to a score of 2 because they were assessed empirically. Only 3 stocks (Acadian redfish, American plaice, and Cape Cod/Gulf of Maine yellowtail flounder) had multiple or recent high recruitment leading to scores of -2 and -4 (Figure 8).

Seven stocks showed no change in recruitment score over time because they were assessed empirically and defaulted to a score of 2 (Figure 10). An additional six stocks were relatively stable over time. No stocks showed a change in score of more than 6 points between time periods. Recruitment scores may be more stable than SSB scores because recruitment scores are based on the most recent 5 years of data so a large change in recruitment, either positive or negative, is buffered by this grouping of recruitment data, whereas SSB scoring is based on the SSB estimate from just the terminal year of the assessment. For example, Georges Bank haddock received four different SSB scores (-4, 2, 2, and 4) over time but only two different recruitment scores (-2 or 2). Similarly, southern windowpane flounder received all five possible scores for SSB, but over the same time period the stock received a recruitment score of 2 for all years except one.

4.1.3 Assessment Type, Performance

Assessment performance scores ranged from 0 to 4. Most stocks, 73%, received a score of 1, 2, or 3. The empirical stocks could only receive a score of 3 or 4 and out of the seven empirically assessed stocks, four received a score of 3 and three stocks had a score of 4. Only three analytically assessed stocks received a score of 0: western Gulf of Maine cod, Gulf of Maine haddock, and American plaice.

4.1.4 Climate Vulnerability

Most groundfish stocks had climate vulnerability levels between moderate (score =1) and high with a negative direction (score =4). Seven stocks were scored very high or high vulnerability with a negative directional change, and seven stocks were scored as moderate vulnerability with a negative direction. Only two stocks, southern windowpane flounder and northern windowpane flounder, were scored as having low vulnerability (score =0).

4.1.5 Fish Condition

For fish condition, nine groundfish stocks had either good or above average fish condition (scores of 0 and 1), while 10 stocks had either below average or poor condition (scores of 3 and 4). Most groundfish stocks, 33%, were scored as below average. Fish condition scores were more variable over time than recruitment or SSB scores. These scores were rarely stable and differed substantially between years and EPUs (Figure 11).

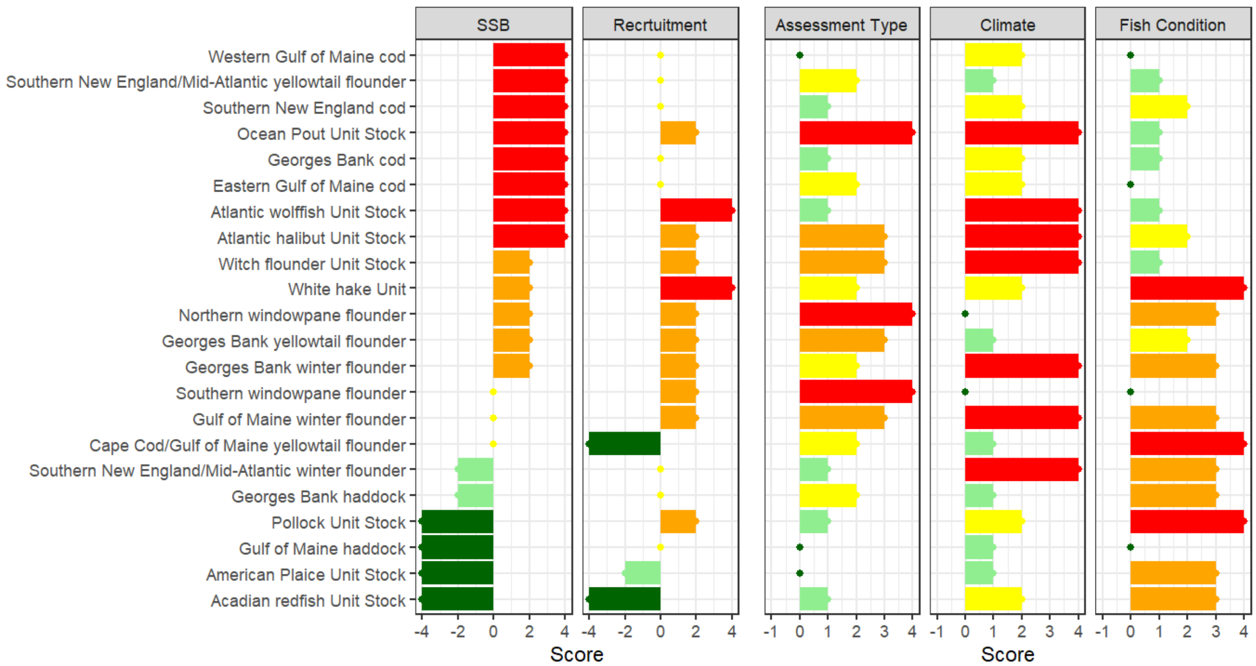


Figure 8. Factor scores for groundfish stocks in New England. Dots represent a score of 0. Note that spawning stock biomass (SSB) and recruitment are scored from a minimum score of -4 to a maximum score of 4, while assessment type, climate, and fish condition are scored from a minimum score of 0 to a maximum score of 4.

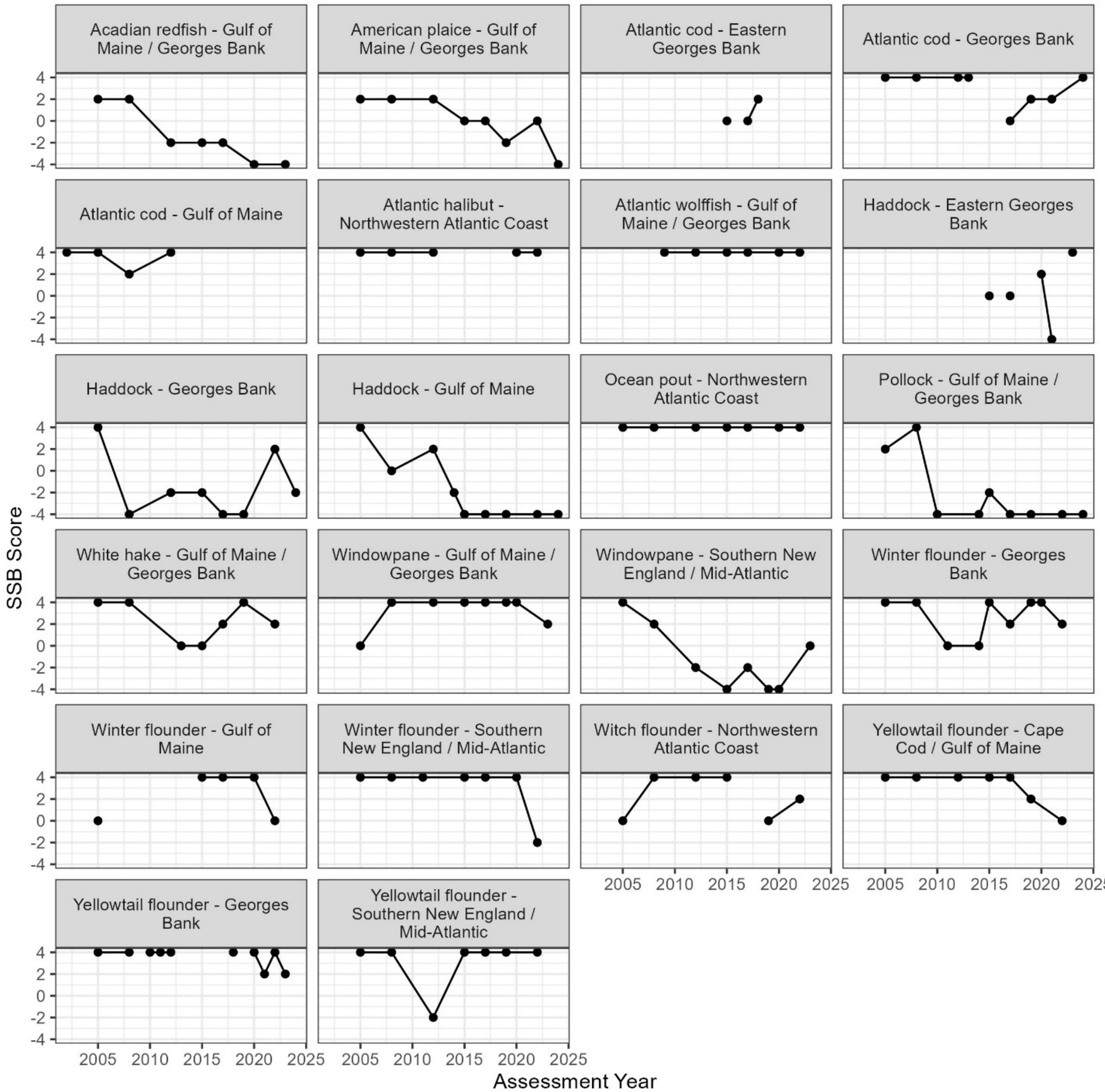


Figure 9. Scores for spawning stock biomass (SSB) over time.

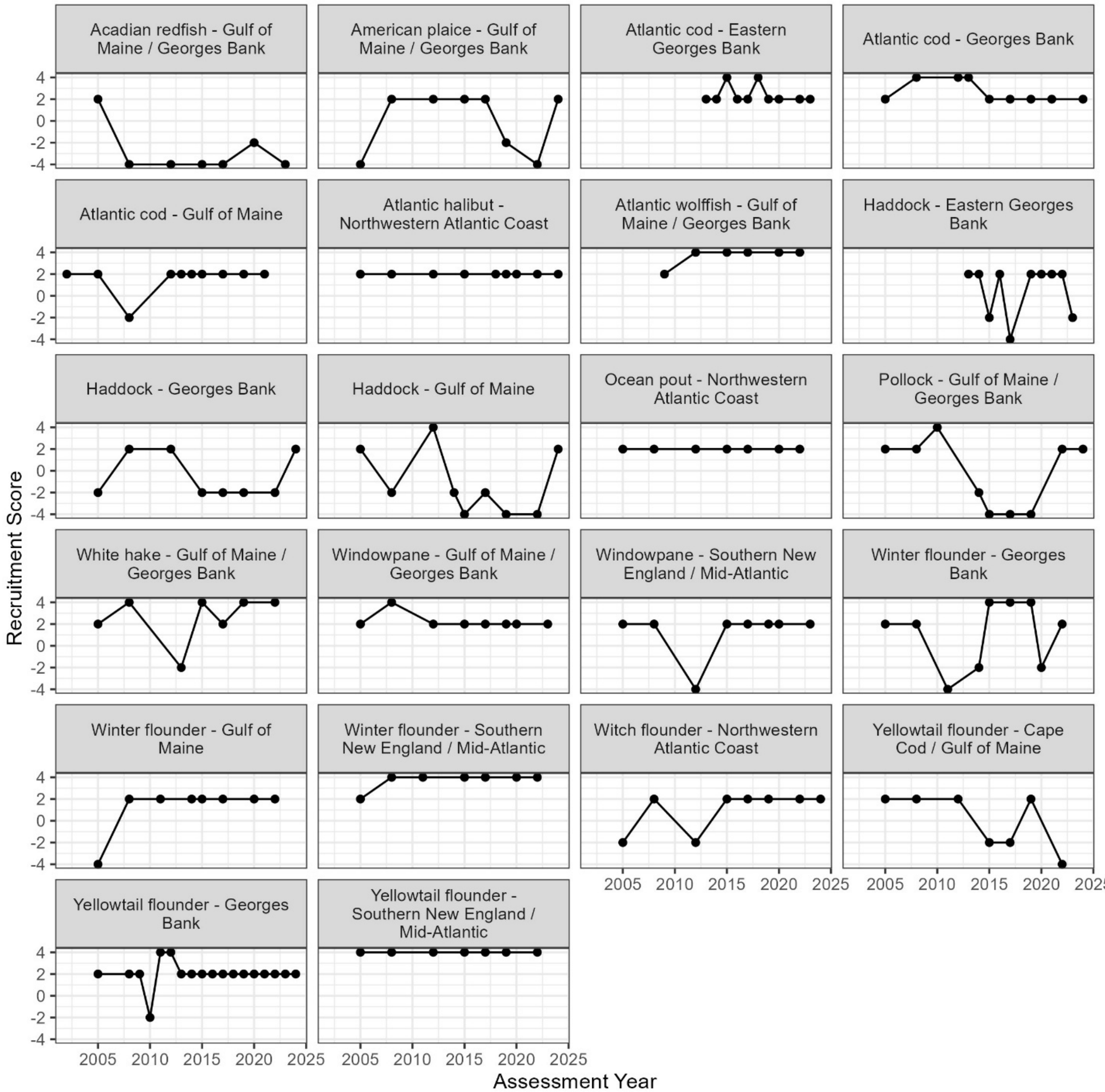


Figure 10. Scores for the recruitment factor over time.

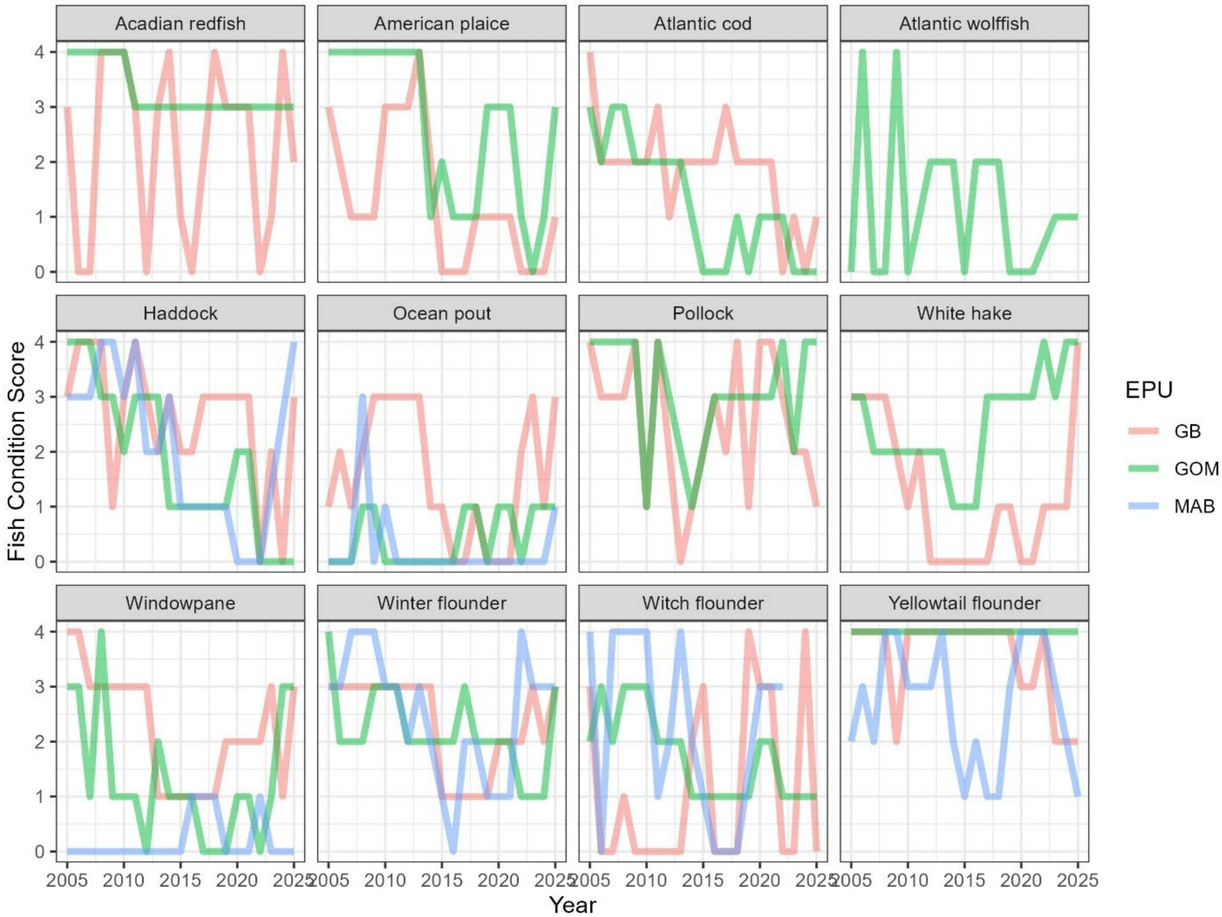


Figure 11. Scores for the fish condition factor over time by ecological production unit.

4.2 Feedback on Scoring Guidelines:

Factors were scored according to the defined rubrics, with assumptions applied only when necessary to generate scores. These assumptions are detailed above in the methods for scoring each factor. We provide both a technical evaluation of the scoring rubric, and a conceptual evaluation of each factor for further consideration.

Many considerations are specific to individual factors, but several common themes exist across factors. First, factor scoring relies on external data (e.g., stock assessment output, climate vulnerability analyses), but the reliability and uncertainty of these data are not considered within the scoring rubric. For example, recruitment estimates from an assessment may have varying degrees of uncertainty or error associated with them. Second, factor scores may fluctuate substantially between years when a metric falls close to a threshold in the rubric (e.g., SSB close to 150% of the SSB reference point). This could produce large inter-annual changes in risk tolerance. This is likely more impactful for factors that separate potential scores in increments of two (i.e., SSB, recruitment, commercial fishery, and recreational fishery) than factors that use increments of one (i.e., assessment type, climate vulnerability, fish condition). Third, the reliability of scores is directly related to the availability of data or a stock assessment.

Conceptually, factors have inherent overlaps and dependencies. For example, whether a stock has an analytical or empirical assessment impacts scoring in three different factors (stock assessment and uncertainty, spawning stock biomass, recruitment). Similarly, inherent dependencies exist between spawning stock biomass and recruitment that can affect the scoring of multiple factors. If the stock has a significant stock-recruit relationship and sensitivity to climatic conditions, these dynamics can potentially affect biomass, recruitment, climate vulnerability and fish condition factors. These interactions should be considered when specifying scoring guidelines, potential score ranges, and weighting so that the desired impact of each factor is captured.

4.2.1 Spawning stock biomass

The rubric for SSB is well defined for stocks with known SSB status, however, the rubric would benefit from specifying how to characterize trends for stocks with unknown status. For this scoring demonstration, we characterized trends in abundance using the significance and directional effect of a simple linear regression at $\alpha = 0.05$. Due to the low sample size ($n = 5$), a higher α may be needed to detect ecologically relevant changes in SSB, such as $\alpha = 0.1$. Other criteria may also be considered, such as nonlinear trends, correlation coefficients, comparisons to a mean or defined bounds (e.g., Rumble strip approach of MAFMC). It is important to note that status is determined for some stocks with empirical approaches and not others, without strong rationale.

More broadly, the SSB factor scores for stocks with known status may be impacted by drivers beyond estimated biomass. The model platform, model specification, and reference points may be updated or revised and alter the perception of SSB relative to reference points. For example, in the 2022 Southern New England winter flounder assessment, the SSB reference point was lowered as a result of restricting the time series of recruitment estimates (NEFSC 2022), which aligns with the large change in SSB score from 4 to -2 in 2022 (Figure 9).

4.2.2 Recruitment

We made necessary assumptions to complete scoring for the recruitment factor (see 3.1.2), but there are also further challenges. The threshold for above, below, and equal to average is not defined by the rubric and specifying these criteria is needed for more consistent scoring across stocks. As written, recruitment for stocks with known estimates will rarely be equal to the average and score a 0. Moreover, recruitment estimates that are slightly above or below the mean might provide a more optimistic or pessimistic view of recruitment than is realistic. A quantile approach may be a more systematic way to identify observations that are above or below a specified percentile value.

For this demonstration and to be consistent across stocks, we estimated mean recruitment using the entire time-series. However, this approach poses a risk if recruitment consistently declines and reduces the average recruitment over time. With this type of trend the probability that poor or average recruitment is scored positively would increase. A more responsive approach might limit the mean recruitment to a time period representative of prevailing conditions.

We also recommend that explicit criteria be developed to identify when "recent changes in recruitment have been accounted for in reference points and/or stock projections" for an average or no trend recruitment score (score of 0). We defaulted to 0 when reference points assumed recent average recruitment (e.g., Southern New England Mid Atlantic yellowtail flounder), projections assume temporal auto-correlation in age-1 abundance (e.g., GOM haddock) or if projections assume a stock-recruit relationship (e.g., SNE Atlantic cod).

Conceptually, recruitment is difficult to generalize and score due to its inherent variability and volatility. Periodic high recruitment may be part of a species life history strategy. Alternatively, high recruitment may be inconsequential if predation or environmental conditions reduce abundance between ages one and two. Currently, sporadic and variable recruitment is not fully considered in scoring.

4.2.3 Assessment Type and Uncertainty

The Risk Policy scores the assessment type and uncertainty factor based on three elements: the assessment type, the magnitude of the retrospective pattern, and the number of uncertainties identified in the assessment report. However, for certain assessment types only the uncertainties or retrospective patterns can be considered, not both (see 3.1.3). The rubric does not explicitly define what is meant by significant sources of uncertainty. Establishing specific, quantitative criteria for classifying uncertainty could improve consistency of scoring across stocks. The number and type of uncertainties listed in stock assessment reports can vary greatly between stocks. A more standardized approach to how assessment uncertainty is characterized within the Risk Policy, could improve the reliability of scoring. While stock assessment reports typically characterize the retrospective pattern as minor or major, it is not clear how one would score an assessment that applied a retrospective adjustment.

Lastly, empirical stock assessments are automatically scored as having high uncertainty (i.e., score a 3 or 4), while analytical stocks can only score from low to intermediate uncertainty scores (0, 1, or 2). However, it is possible to have an empirical model that meets management needs.

4.2.4 Climate Vulnerability

The rubric provided clear guidance for scoring climate vulnerability. We made no additional assumptions, but caution that no species scored a 3, "high vulnerability." We also note that the basis for scoring, the Hare et al. (2016) Climate Vulnerability Analysis (CVA), is almost ten years old and potentially outdated. Updated climate vulnerabilities are forthcoming, but the associated scorings are ultimately static until new analyses are produced. The Council might consider the frequency at which they may need updated information in the future as ecosystems continue to change. It is also important to note that CVA's are based upon a structured process to elicit expert opinion. While this approach is rooted in scientific consensus, there is a qualitative nature to the synthesis. Given recent advances in the region to characterize climate impacts, there may be opportunities to use more quantitative indicators with mechanistic linkages to stock performance.

4.2.5 Fish Condition

Conceptually, the relationship between fish condition and fisheries risk is not well understood. For the Risk Policy, fish condition aims to be a proxy for ecosystem productivity, however, it is also impacted by cohort sizes, prey availability, growth rates, and spawning, which vary within and among years.

A key challenge for scoring fish condition is the spatial extent of the State of the Ecosystem fish condition data compared to stock boundaries. The State of the Ecosystem reports in ecological production units, which do not align with the spatial domain of stocks (e.g., statistical units, survey strata). For our scoring demonstration, we weighted the fish conditions from multiple regions equally. However, a stock's distribution is unlikely to be uniform throughout all regions. Producing fish conditions by stock area would require additional resources but would improve the reliability of this factor.

4.2.6 Commercial and Recreational Fishery Characteristics

The scoring rubrics for commercial fishery characterization and recreational fishery characterization are still underway. The stocks without a recreational fishery will include the recreation fishery characterization in Z-Score calculations but will likely default to a score of 0. We recommend that during rubric development scoring considerations and factor specific difficulties are considered to ensure robust scoring that accomplishes the Council's goals.

4.3 Scenarios:

4.3.1 Five Factor Scenario-Baseline:

The baseline scenario produced Z-Scores that ranged from -0.74 to 3.15, with the final quantitative output of the Risk Policy, defined as the recommended probability of management success, ranging from 33 to 96%. Only three stocks received Z-Scores that were less than zero: Acadian redfish, American plaice, and Gulf of Maine haddock. Most stocks, 60%, received recommended probabilities above 80% (Figure 12).

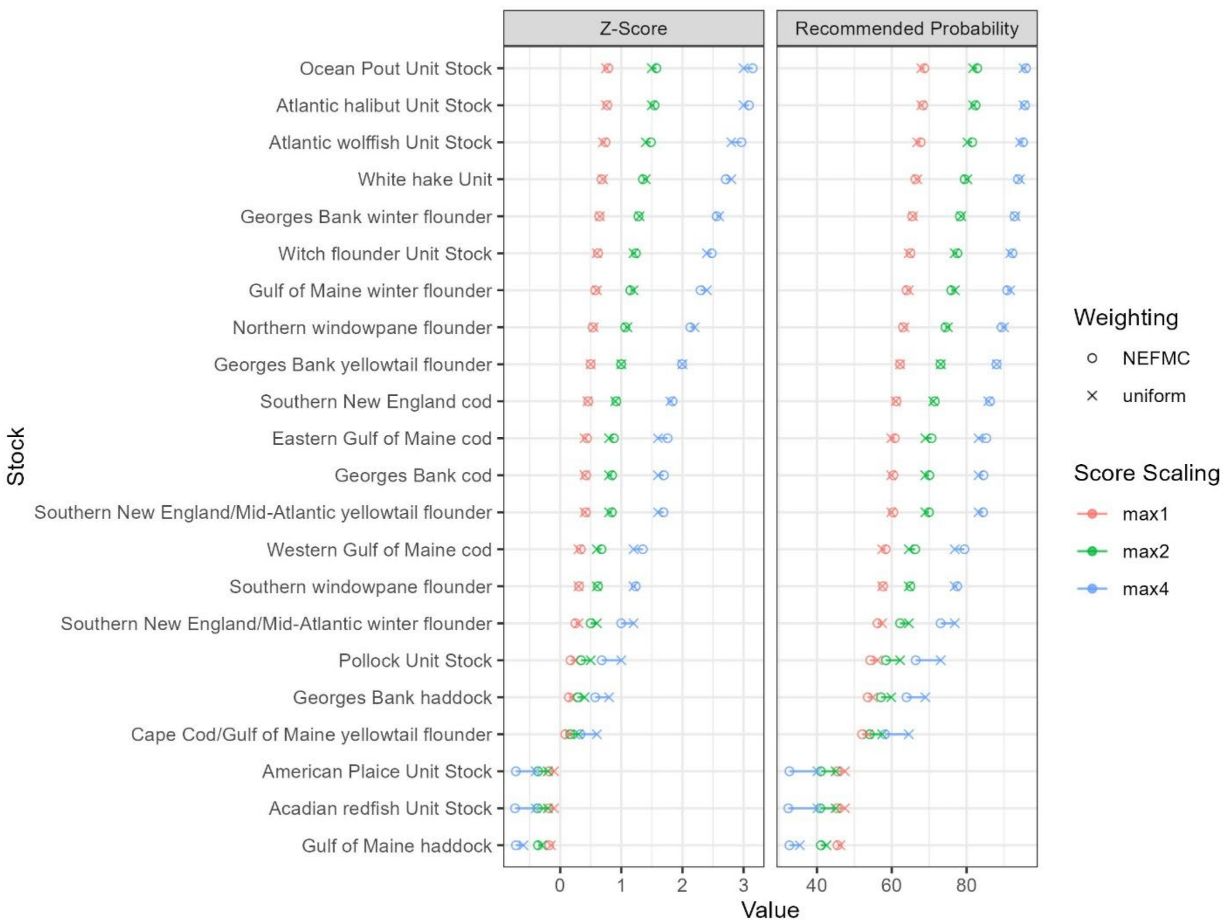


Figure 12. Range of recommended probabilities and Z-Scores for individual NEFMC groundfish stocks under three scaling scenarios (maximum Z-Score of 1, 2, or 4) and using a uniform or council-based weighting scenario.

4.3.2 Five Factor Z-Score Scaling Scenarios:

The final scaling of the factor scores (4, 2, or 1) had a notable impact on the resulting Z-Scores and recommended probabilities. Factor scores scaled to a maximum of 4 resulted in a greater range of potential Z-Scores and, therefore, the recommended probabilities of management success that would come out of the Risk Policy. For example, the groundfish stock with the highest Z-Score (i.e., ocean pout) had a recommended probability of 96% when scaled to 4 but a recommended probability of 69% when scaled to 1 (Figure 13). Under a maximum scale of 1, groundfish stocks are limited to recommended probabilities below 70%.

The different scaling scenarios effectively change the functional form of the Risk Policy by limiting the range of possible Z-Scores and recommended probabilities. The Risk Policy implements a logistic function to translate Z-Scores into the recommended probability of management success, which includes a linear portion at Z-Scores close to zero with an asymptote at higher Z-Scores. This means that the same magnitude of change in Z-Score has a larger influence on the resulting recommended probabilities along the linear portion when stock conditions promote increased risk tolerance (i.e., low Z-Scores) and a smaller influence along the asymptote when stock conditions require caution (i.e., high Z-Scores). Scaling scenarios that limit the range of possible Z-Scores (e.g., maximum = 1, Figure 13) only use the linear portion of the logistic curve and the asymptote is only accessible when higher Z-Scores are achievable.

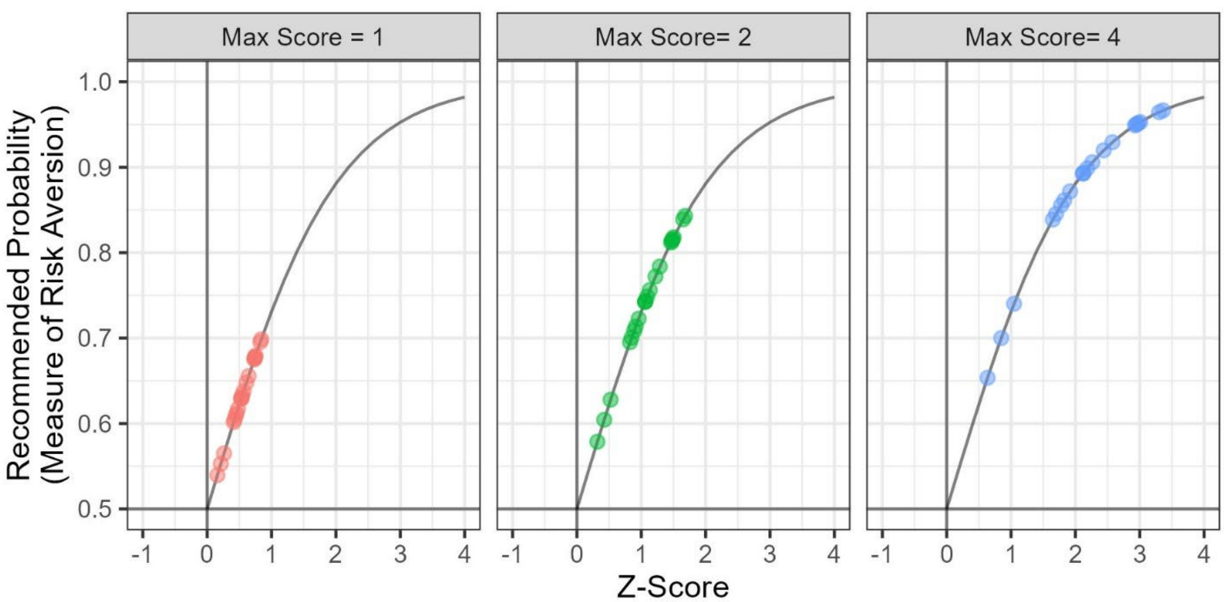


Figure 13. Range of recommended probabilities and Z-Scores for NEFMC groundfish stocks under three scaling scenarios: max factor score 1, 2, or 4.

4.3.3 Five Factor Weighting Scenarios:

The alternative weighting scenarios, uniform versus Council defined weights, were applied to the three scales of Z-Scores. The Council weighting process put the greatest emphasis on SSB and commercial fishery characterization and assigned the lowest emphasis on fish condition and recreational fishery characterization (Table 2). Compared to uniform weighting, this increased the influence of SSB and commercial factors and reduced the influence of fish condition and recreational inputs (Figure 14).

For most groundfish stocks, the choice of weighting schemes had a negligible effect on the Z-Scores. However, differences emerged for stocks with risk averse scoring (i.e., low Z-Scores; Figure 12). This effect stems from the higher importance assigned to SSB and recruitment under the Council weighting (Figure 14), combined with the greater range of possible scores for these factors under positive conditions.

Both SSB and recruitment can achieve a minimum score of -4, compared to the other three factors, which have minimum scores of 0. Therefore, if a stock receives the minimum score for either SSB or recruitment, there is a greater impact on the final Z-Score than for other factors; this is amplified under Council weighting. With a maximum scale of 4 the weighting scheme had the greatest impact on Z-Scores for Acadian redfish, American plaice, pollock, and Gulf of Maine yellowtail flounder. All four stocks have the minimum score for either SSB, recruitment, or both (Figure 8), highlighting the disproportionate impact of minimum SSB and recruitment scores and the compounding impacts of the Council weights.

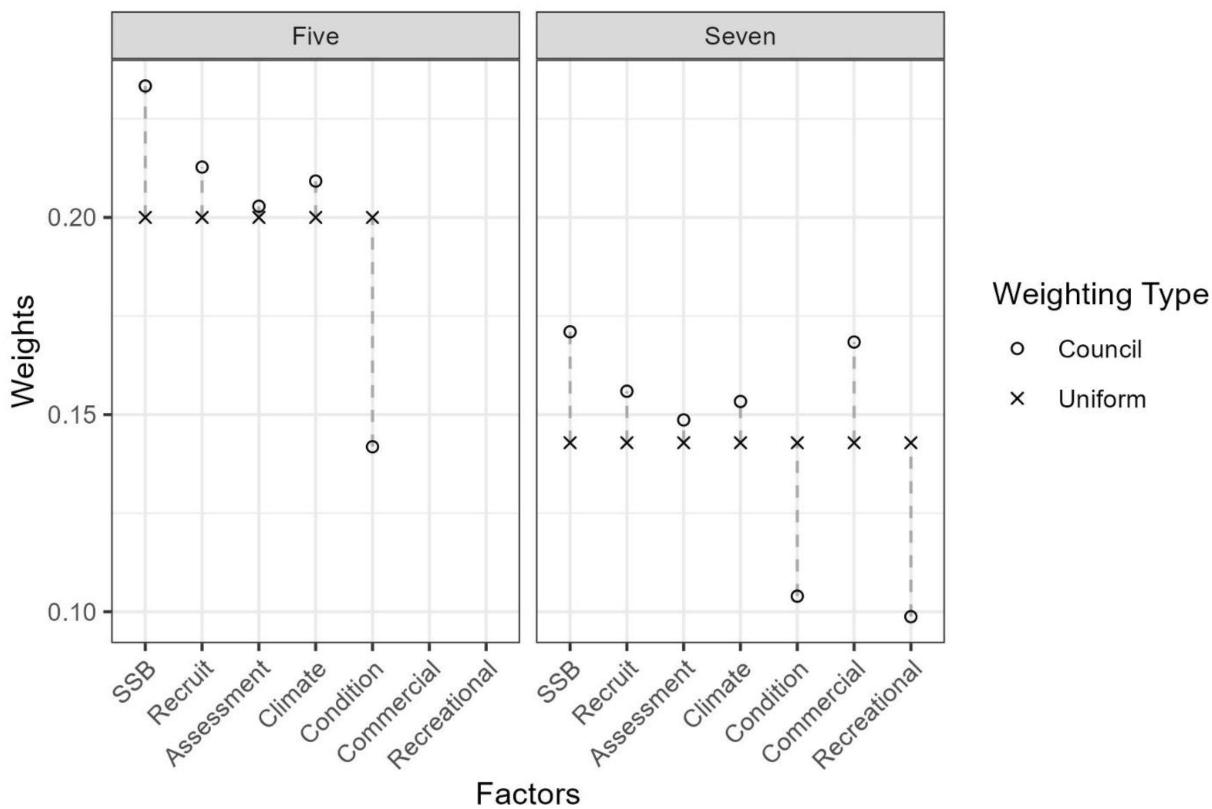


Figure 14. Weights applied to factor scores for the Council weighting scheme and the uniform weighting scheme for scoring demonstrations with 5 factors (left) and 7 factors (right).

4.3.4 Five Factor Equal Score Ranges Scenario:

We demonstrated the impact of rescaling the SSB and recruitment factors from 0 to 4, rather than -4 to 4. The difference in Z-Scores and probabilities between the Council and uniform weighting scenarios was reduced for stocks with low Z-Scores (Figure 15), indicating that differences between the weighting scenarios are exacerbated by the ranges of factor scores. The Council weighting scheme, coupled with a greater scoring range, implicitly places a higher degree of importance on SSB and recruitment under positive stock status. While some difference in recommended probabilities is to be expected with the shift in factor ranges, the difference is greater for low Z-Scores (Figure 15), due to the greater steepness of the logistic curve for Z-Scores below 2.

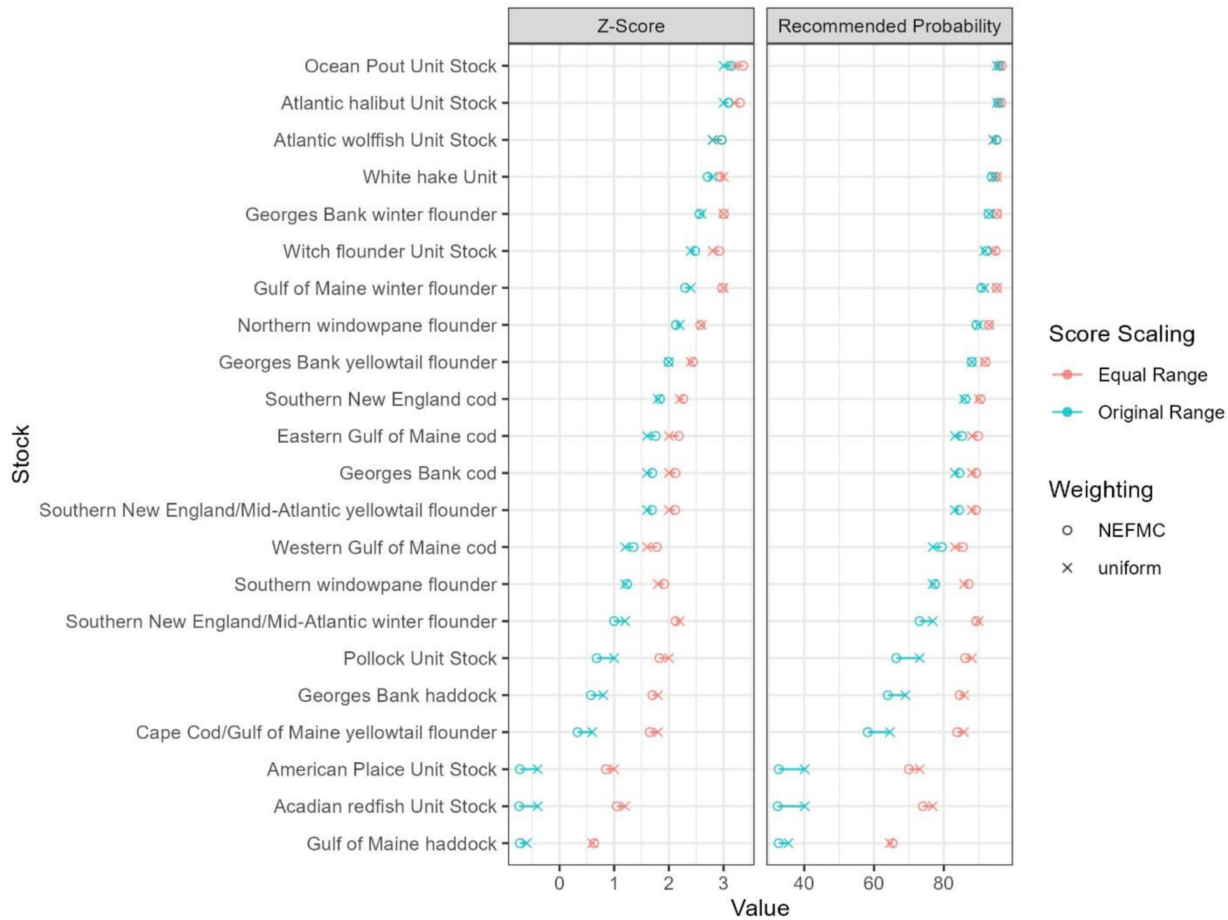


Figure 15. Range of Z-Scores and recommended probabilities for individual groundfish stocks using either the uniform or council defined weighting scenario and either the original factor ranges or equal ranges. The scenarios use a maximum Z-Score of 4.

4.3.5 <u>Assessment Type / Uncertainty Scenarios:</u>

The range of potential scores differed substantially between stocks with empirical and analytical assessments (Table 3). The lowest potential total score, summed across factors, for an empirical stock was 5 whereas the lowest potential score for an analytical stock was -8. This disparity translated into substantial differences in Z-Scores and recommended probabilities, with similar results across weighting schemes. For analytical stocks, the lowest score resulted in a recommended probability of 0.14 and for empirical stocks the recommended probability associated with the lowest potential score was 0.74, using the Council weights (Figure 16).

The difference between assessment types was less pronounced at the high end of the scoring range. A stock with an empirical assessment and high scores across all factors reached a total score of 20, compared to 18 for a stock with an analytical assessment. This resulted in nearly identical recommended probabilities (Table 3).

The two groundfish stocks with the highest Z-Scores, associated with the lowest risk tolerance, were ocean pout and Atlantic halibut, both of which are assessed empirically. The empirically assessed stock with the lowest Z-Score and highest risk tolerance was southern windowpane flounder (Z-Score = 1.24). Eight analytical stocks scored within the same range as the empirical stocks, while seven had lower Z-Scores than the lowest Z-Score received by southern windowpane flounder.

Table. 3 Extreme scoring scenarios for hypothetical stocks assessed with an analytical or empirical assessment. The "highest" scenario represents the maximum possible score achievable for each assessment type, while the "lowest" scenario represents the minimum possible score achievable for each assessment type. Z-Scores (Z) and recommended probabilities (P) are based on the baseline scale (4) and Council defined weights.

Empirical Assessment	SSB	Recruit	Assessment	Climate	Condition	$Z_{Council}$	$P_{Council}$
Lowest	0	2	3	0	0	1.03	0.74
Highest	4	4	4	4	4	4	0.98

Analytical Assessment	SSB	Recruit	Assessment	Climate	Condition	$Z_{Council}$	$P_{Council}$
Lowest	-4	-4	0	0	0	-1.78	0.14
Highest	4	4	2	4	4	3.59	0.97

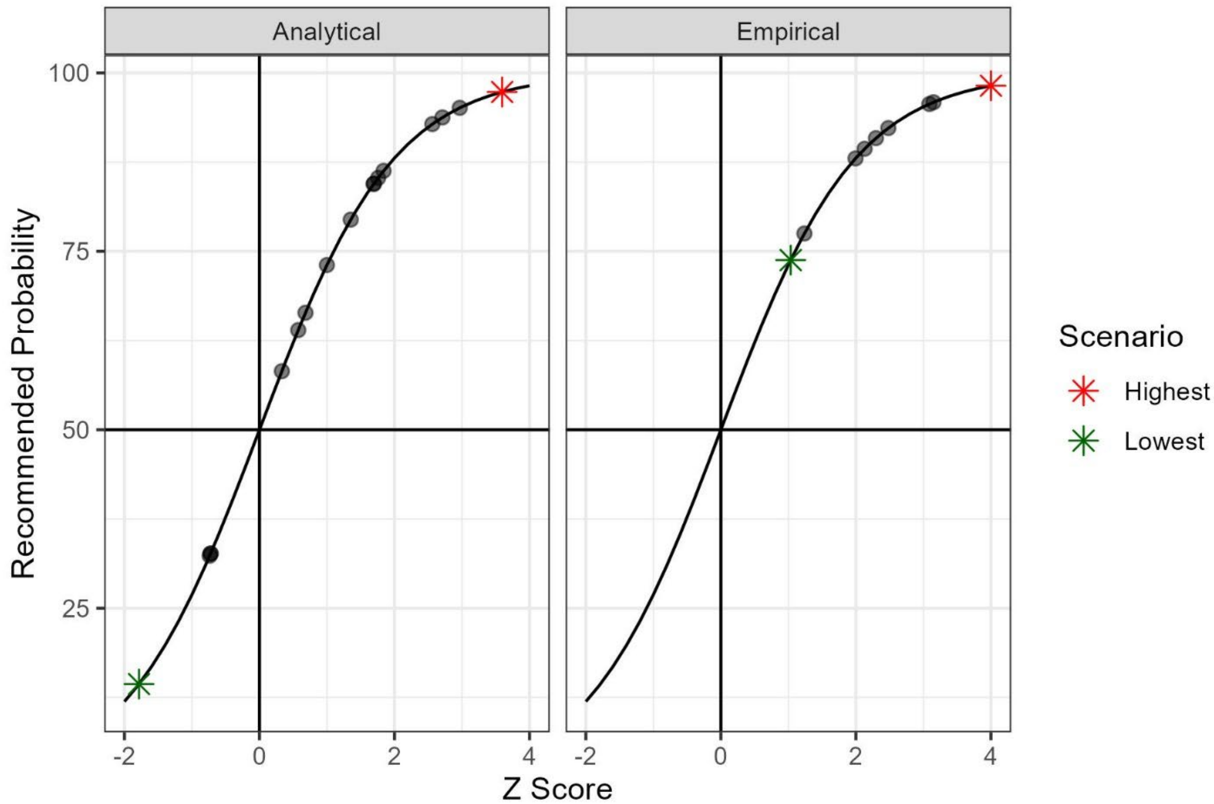


Figure 16. Z-Scores and recommended probabilities for stocks with analytical and empirical assessments (black dots) along with the highest and lowest possible values for each assessment type (colored stars). This demonstration uses a maximum scale of 4 and the Council defined weights.

4.3.6 <u>Seven Factor Scenarios (fishery factors included):</u>

We also explored the sensitivity of Z-Scores and corresponding recommended probabilities to including the two fishery-factors. We did not demonstrate scoring for these factors, but instead explored the implications of incorporating the maximum, intermediate, and minimum scores. As expected, including the maximum fishery factor scores increased the Z-Scores and recommended probabilities compared to the five-factor baseline scenario, while including the minimum fishery factor scores decreased them. Interestingly, the difference from the baseline scenario was greater when adding the minimum scores than when adding the maximum scores, particularly for stocks with high Z-Scores (Figure 17). This is likely a result of the disproportionate impact of including negative numbers into the Z-Score for stocks that otherwise have primarily positive scores under the baseline scenario (Figure 8).

The influence of incorporating the intermediate fishery factor scores varied by stock. For all stocks, incorporating the intermediate fishery scores (i.e., score of 0) moved the Z-Score closer to zero, but the direction of change depended on the stock-specific Z-Score in the baseline scenario. Stocks with baseline Z-Scores above 0, and corresponding recommended probabilities greater than 50%, saw declines in both metrics, while stocks with baseline Z-Scores below 0 and recommended probabilities less than 50% saw increases in both metrics. The rubrics for the fishery characterizations have not been defined, but stocks without recreational components will likely default to a score of 0. It is important to note that the Risk Policy specifies that a recommended probability below 50% cannot be applied in management. Therefore, stocks with negative z-Scores will always default to a recommended probability of 50% in practice.

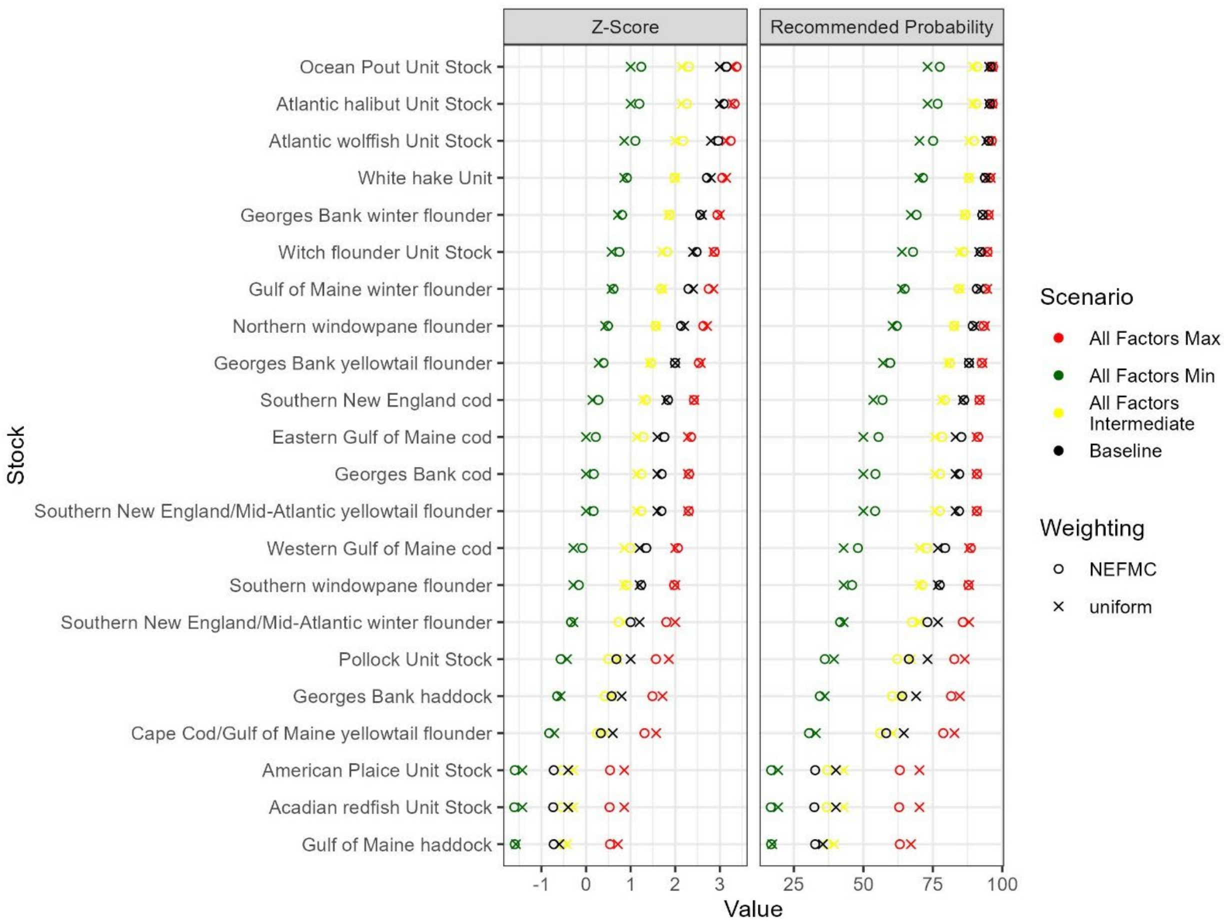


Figure 17. The range of Z-Scores and recommended probabilities for individual NEFMC groundfish using either the uniform or Council defined weights and scored fishery factors (commercial and recreational fishery characterization). "All factor" scenarios calculate the Z-Scores and recommended probabilities based on all seven factors, using the maximum, minimum, or intermediate score for the commercial and recreational fishery scores. The

“Baseline” scenario uses the five factors with a defined scoring rubric. All scenarios use a maximum Z-Score of 4.

4.3.7 Four Factor Scenarios (fish condition factor excluded) & Six Factor Scenarios (fishery factors included & fish condition factor excluded):

The impact of excluding the fish condition factor on Z-Scores and recommended probabilities depended on the stock-specific condition score (Figure 18). For example, the three stocks with the highest overall Z-Scores and recommended probabilities—ocean pout, Atlantic halibut, and Atlantic wolffish—showed further increases because their fish condition scores were only 1 or 2. Conversely, for stocks such as American plaice and Acadian redfish, which had relatively higher fish condition scores, removing the factor led to a decrease in their Z-Scores and recommended probabilities. Removing fish condition also decreased the difference in Z-Scores and recommended probabilities between the Council and Uniform weighting scenarios, because across the 5 factors the Council placed the least importance on fish condition (Figure 14).

Including the fishery characterization factors while excluding fish condition produced results similar to scenarios where both fishery characterization and fish condition were included (Figures 17, 19).

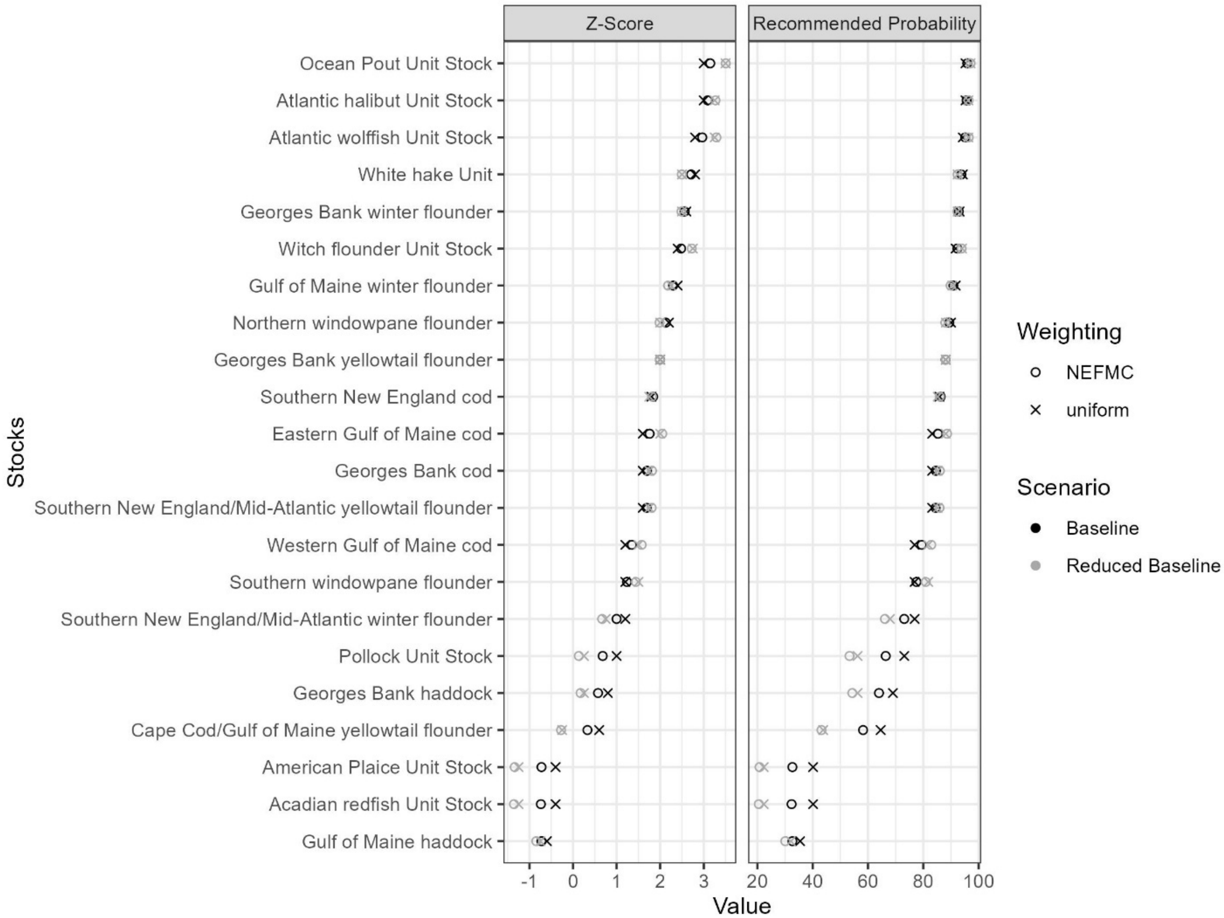


Figure 18. The range of Z-Scores and recommended probabilities for individual NEFMC groundfish stocks using either the uniform or council defined weighting scenario while excluding fish condition scores. This "Reduced Baseline" scenario calculates the Z-Scores and recommended probabilities based on four factors: SSB, recruitment, assessment type, and climate vulnerability. The Baseline scenario includes all five factors with a defined scoring rubric. Both scenarios used a maximum Z-Score of 4.

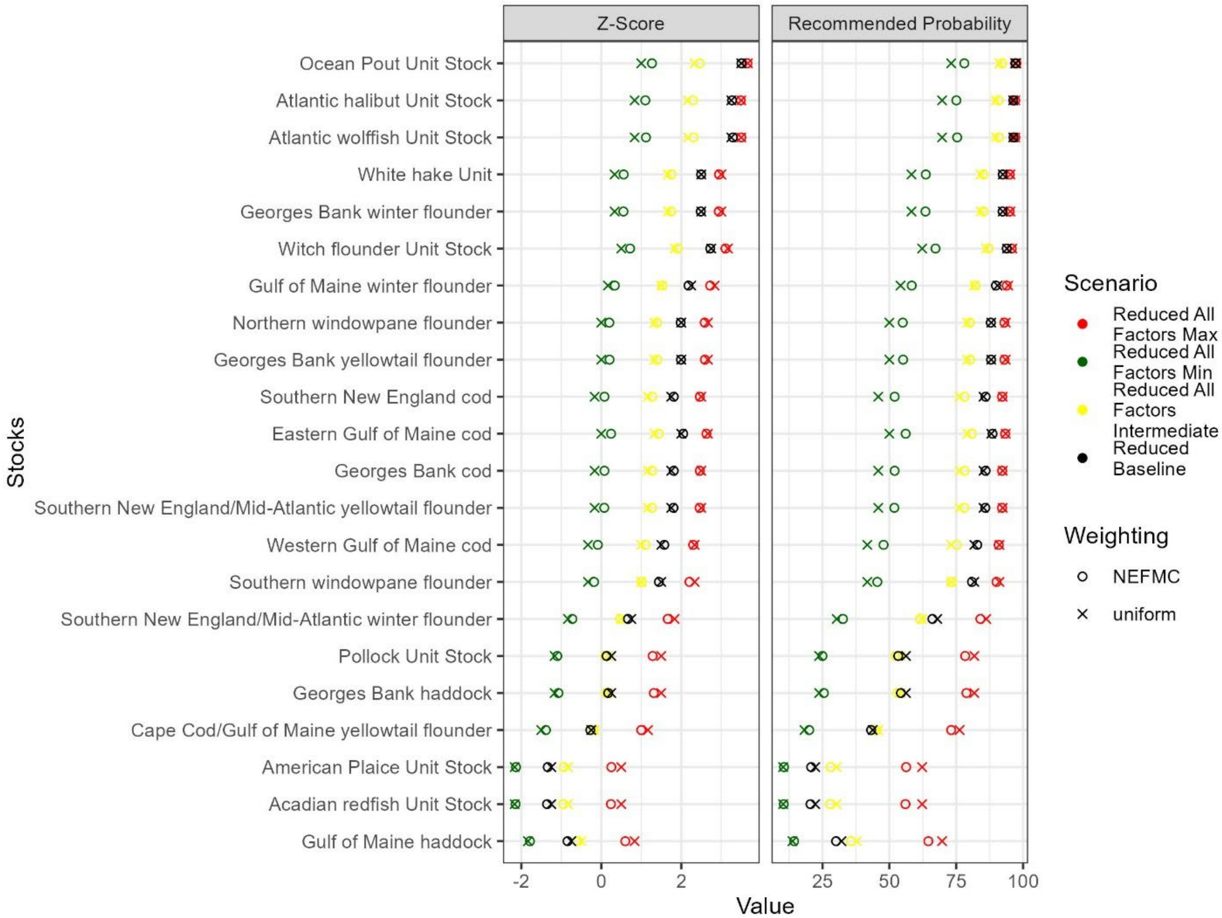


Figure 19. The range of Z-Scores and recommended probabilities for individual NEFMC groundfish stocks using either the uniform or council defined weighting scenario and fishery factors (commercial and recreational fishery characterization) while excluding fish condition scores. These "Reduced All factor" scenarios calculate the Z-Scores and recommended probabilities based on six factors, using either the maximum, minimum, or intermediate score for the commercial and recreational fishery factors. The "Baseline" scenario uses four factors (SSB, recruitment, assessment type, and climate vulnerability). All scenarios use a maximum Z-Score of 4.

5. Takeaways & Recommendations

5.1 How well does the Risk Policy Meet its Stated Goals?

The stated goals of the new Risk Policy for the NEFMC are to: 1) provide guidance to the Council and its subordinate bodies on taking account of risk and uncertainty in Fishery Management Plans and specification setting; 2) clearly communicate the priorities and preferences of the Council regarding risk and uncertainty, including using a common set of terms and definitions so it is accessible to a wide variety of audiences; and 3) make the discussion of risk tolerance in the Council's decisions a more forward and fundamental aspect of the management process (NEFMC 2024).

The new Risk Policy seeks to meet these goals by identifying factors of importance, incorporating the Council's perspective on the relative importance of each factor via weights, and using the outlined policy to explicitly and transparently define the amount of risk intended with each management decision (e.g., setting catch advice). Key features motivating this revision are described below and we have provided our perspectives on how these have been addressed in the Risk Policy implementation thus far. We also note potential challenges to implementation and recommendations for consideration.

5.1.1 Integrating changing environmental conditions into the Council's assessment of risk

Factors in the revised risk policy allow for consideration of climate change and the dynamic environment in the Council decision process. The direct incorporation of climate and ecosystem risk into the risk assessment is novel to the region and identification of appropriate indicators has been challenging. The climate vulnerability and fish condition factors seek to account for species specific and ecosystem conditions related to climate change, but these effects are complex, multidimensional and uncertain.

It is important to note that one of the two factors, climate vulnerability, will be relatively stationary over time, only changing when a new climate vulnerability analysis is conducted. The Council should consider the rate at which updated climate vulnerability data may be needed moving forward (the last one was conducted in 2016, but an update is underway and scheduled to be completed by fall 2026). On the other hand, the fish condition factor will be dynamic, likely updated annually. There are questions, however, regarding the appropriateness of this as a proxy for ecosystem condition. Finally, it is important to note that there is ongoing work by the Risk Policy Working Group to reconsider climate and ecosystem factors and scoring details to ensure they are meeting the goals of the risk policy.

5.1.2 Developing a clear path to incorporate social and economic considerations

The inclusion of recreational and commercial fishery characterization factors within the Risk Policy aims to more explicitly integrate social and economic considerations into Council decision-making. Fishery dynamics for both the recreational and commercial sectors are difficult to capture given the intersecting impacts of social, economic, and biological drivers. Directly incorporating the outlook of the fishery into decision-making is new for the region and capturing the complexities in a representative and systematic way may be difficult.

Although there is ongoing work by the Risk Policy Working Group to define fishery factors and scoring details to ensure they are meeting the goals of the policy, a formal rubric had not been developed at the time of this report. This highlights the difficulty of identifying appropriate indicators for these fisheries and their relationship to risk tolerance. Therefore, we demonstrated scenarios that use maximum, minimum and intermediate scores to explore the impact of including these factors. We used a maximum score of 4, a minimum score of -4 and an intermediate score of 0. However, the range of output may also be used to inform potential rescaling of the factor ranges, such as from -4 to 0 as has been discussed by the Risk Policy Working Group. In formulating the final rubrics for these factors, our overarching recommendations about factor scoring should be considered: quantitative and specific criteria, uncertainty of input data, the frequency at which input data is produced, and any interdependence among factors. Moreover, including these factor scores had a notable influence on the range of potential Z-Scores and recommended probabilities, suggesting that defining these rubrics and finalizing the potential range of factor scores is essential.

5.1.3 Transparency in Decision Making

The outlined approach to scoring and weighting aims to increase transparency in how the Council assesses risk and incorporates it in decision-making. Transparency in fisheries management is critical for fostering trust with the industry and stakeholders. A publicly available and methods-based Risk Policy concept, with examples (NEFMC 2024), helps accomplish this goal. However, the new Risk Policy is technical in nature and it includes several steps (e.g., weighting, scoring, scaling) that have distinct purposes but can be easily confused because they are inter-related and the differences are nuanced. This makes clear communication challenging so fully understanding the process and interpreting the outcomes may be difficult for non-technical audiences.

One foundational challenge is that some aspects of the process, including factor scoring and calculating risk outcomes, have non-intuitive directionality. For example, in factor scoring a high and positive number corresponds to decreased risk tolerance and a low or negative number corresponds to increased risk tolerance. This leads to counter-intuitive scoring within each factor whereby stocks that are “doing well” by the factor definitions receive negative scores (e.g., high biomass receives a score of -4). One way to resolve this would be to invert the scoring rubric (e.g., high biomass becomes a score of 4 and low biomass becomes a score of -4), but this would have implications for subsequent steps.

Another key challenge is the terminology used to describe the measure of risk tolerance, which is calculated by inputting the Z-Scores into a logistic function (i.e., Y-axis of Figure 1). This value is defined as the “recommended probability of management success.” This terminology has some benefits (e.g., aligning the Risk Policy with National Standard 1), but is also prone to mis-interpretation. This wording can lead one to interpret the Risk Policy output as the true empirical probability that management will succeed under a given set of factor scores. In reality, the metric is intended to reflect the level of importance managers need to place on ensuring that management is successful (i.e., how risk averse they should be). These are fundamentally different concepts, and the distinction is critical for effective communication.

Lastly, throughout the Risk Policy the terms “risk aversion” and “risk tolerance” are used interchangeably, often with qualifiers such as “less” and “more”. This likely reflects the fact that different terms feel more intuitive at different steps in the process. However, the multi-step structure of the policy and different directionality at each step make it challenging to apply consistent terminology. As a result, the reader has to repeatedly reconsider the definition of each metric and the implications with respect to risk and precaution in management. This, coupled with the unintuitive relationships whereby a positive stock status leads to a lower Z-Score, but a higher recommended probability, and then a lower catch advice decreases the interpretability of the Risk Policy. We recommend consistent and specific use of risk terminology to improve interpretability.

5.2 Potential Performance Issues

The new Risk Policy will be used to provide guidance on risk in Council decision-making and has the potential to quantitatively inform the catch advice setting process (i.e., buffer between overfishing limits (OFL) and acceptable biological catch (ABC)). Therefore, the mechanics of scoring and calculating the Z-Scores and recommended probabilities could ultimately impact the performance of ABC control rules. Understanding the sensitivity of the Policy to different weightings, scaling, scoring, and factors, is critical to ensuring that the Policy is functioning as intended.

5.2.1 <u>Different score ranges across factors</u>

Across the scenarios we explored, the largest differences in Z-Scores and recommended probabilities were related to implicit weighting associated with different score ranges, weighting options, and how many factors of each range were included in the Z-Score. Currently, factor scores vary in range to limit the impact a specific factor can have on risk (e.g., climate vulnerability cannot increase risk tolerance, it can only decrease risk tolerance). However, the ultimate impact of a score on risk and therefore management decisions is dependent on how the baseline or neutral condition interacts with the logistic curve. Within scenarios, constraints imposed on the potential scores for different types of assessments had a notable impact on the range of Z-Scores and recommended probabilities. Therefore, we recommend that the Council consider the desired range of recommended probabilities they would like stocks to be able to access, the level of importance they aim to place on specific factors (i.e., implicit weighting), and what scores are associated with neutrality.

5.2.2 Scoring Assumptions

Factor scoring required multiple assumptions and alterations to the rubrics. Scoring would be more consistent with specific, quantitative criteria rather than qualitatively or visually defining trends. We recommend that the Risk Policy Working Group consider our assumptions and alter or adopt them as they see fit. Second, the quality of the input data, the time since the last assessment, the rate at which a stock is assessed, and the acceptance of the assessment by peer-review may impact the reliability of scoring for SSB, recruitment, and assessment type and uncertainty. These underlying details are subject to change as a function of stock assessment priorities, staff workloads, and managing jurisdiction, therefore contingencies for changes in data quality and assessment frequency may be beneficial to consider.

5.2.3 Harvest Control Rule Integration

Ultimately, the relative importance of different Z-Scores and recommended probabilities are related to how they will be integrated into ABC control rules. For example, in a tiered approach, stocks with factor scores that lead to recommended probabilities between 80 to 100% may have the same buffer applied to their OFL. However, if a more dynamic approach to integrating the Risk Policy into ABC control rules is used, the impact of different recommended probabilities may have a more pronounced impact on catch advice. It is important to note that the relationship between intended risk tolerance and actual management success is complex. While a recommended high probability of management success implies a need to be more cautious when setting catch advice, identifying specific management measures (e.g., an exact buffer between the OFL and ABC) that result in the intended probability of management success (i.e., not overfishing) could be operationally challenging and would benefit from detailed simulation testing.

In addition, due to the shape of the logistic curve the same change in Z-Score has a different impact on risk tolerance depending on the location on the curve. For example, at high Z-Scores a change in factor score will have a small impact and at low Z-Scores a change in factor score will have a large impact on the resulting recommended probability.

With the current directionality of scoring, stock conditions that allow more risk tolerance (low factor scores and low Z-Scores) fall on the steepest part of the logistic curve and stock conditions that require more caution (high factor scores and high Z-Scores) fall on the asymptote of the logistic curve. This means that the Risk Policy output can move more quickly when risk tolerance is high than when risk tolerance is low. In practice, this may mean that catch advice changes more rapidly when the ABC is close to the OFL and more slowly when the ABC is far from the OFL, which may not align with management intentions.

7. References

Beltz B, Beet A, Bastille K, Hardison S. 2025. *ecodata: Documentation of Ecosystem Indicator Reporting*. R package version 6.0, https://github.com/NOAA-EDAB/ecodata, https://noaa-edab.github.io/ecodata/.

Beet A. 2025. *stocksmart: Provides access to NOAAs stock SMART data*. R package version 0.6.45, https://github.com/NOAA-EDAB/stocksmart.

Hare JA, Morrison WE, Nelson MW, Stachura MM, Teeters EJ, Griffis RB, Alexander MA, Scott JD, Alade L, Bell RJ, Chute AS, Curti KL, Curtis TH, Griswold CA . 2016. A Vulnerability Assessment of Fish and Invertebrates to Climate Change on the Northeast U.S. Continental Shelf. PLOS ONE 11(2): e0146756. https://doi.org/10.1371/journal.pone.0146756

Magnuson-Stevens Fishery Conservation and Management Act, 16 U.S.C. § 1801 et seq. (2006); 50 C.F.R. Part 600 (2024).

New England Fishery Management Council (NEFMC). 2025. Risk Policy Statement and Concept (2025). Version 1. https://d23h0vhsm26o6d.cloudfront.net/Risk-Policy-Statement-and-Concept-Overview-for-posting-v1-final.pdf

New England Fishery Management Council (NEFMC). 2016. Risk Policy Road Map. https://d23h0vhsm26o6d.cloudfront.net/Risk.Policy.Road.Map_Final_063016.pdf

Northeast Fisheries Science Center (NEFSC). 2022. Southern New England Mid-Atlantic winter flounder 2022, Management Track Assessment Report.

Northeast Fisheries Science Center (NEFSC). 2025a. State of the Ecosystem 2025: New England. Northeast Fisheries Science Center. https://doi.org/10.25923/zr75-a788

Northeast Fisheries Science Center (NEFSC). 2025n. State of the Ecosystem 2025: Mid Atlantic. Northeast Fisheries Science Center. https://doi.org/10.25923/23nx-qf59