Report of an Independent Peer Review of Japanese Amberjack, Pacific Cod, and Pointhead Flounder Stock Assessments Conducted by the Japan Fishery Research and Education Agency (FRA) in 2024

By

Steven L. H. Teo

Research Fishery Biologist

Southwest Fisheries Science Center

National Oceanic and Atmospheric Administration (NOAA)

8901 La Jolla Shores Drive

La Jolla, California 92037-1508

USA

December 18th, 2024

Table of Contents

1. Summary....5
2. Background....9
3. Review Process....9
4. Japanese Amberjack (JA)....10
    4.1 Data....10
        4.1.1 Catch-at-age....11
        4.1.2 Relative abundance indices....12
        4.1.3 Recommendations on data....12
    4.2 Biology....13
        4.2.1 Stock structure and distribution....13
        4.2.2 Natural mortality....13
        4.2.3 Maturity....13
        4.2.4 Growth....14
        4.2.5 Recommendations on biology....14
    4.3 Estimation Model....14
        4.3.1 VPA model....14
        4.3.2 Model diagnostics....16
        4.3.3 Uncertainty....16
        4.3.4 Recommendations on estimation modeling....16
    4.4 Projections....17
        4.4.1 Stock-recruitment relationship (SRR)....17
        4.4.2 Short-term projections....18
        4.4.3 Long-term projections....18
        4.4.4 Recommendations on projections....19
5. Pacific Cod – Honshu Northern Sea of Japan Stock (PC-HNSJ)....20
    5.1 Data....20
        5.1.1 Catch-at-age....20
        5.1.2 Relative abundance indices....21
        5.1.3 Recommendations on data....22
    5.2 Biology....23
        5.2.1 Stock structure and distribution....23
        5.2.2 Natural mortality....23
        5.2.3 Maturity....23
        5.2.4 Growth....24
        5.2.5 Recommendations on biology....24

5.3 Estimation Model.......... 24
5.3.1 VPA model.......... 24
5.3.2 Model diagnostics.......... 25
5.3.3 Uncertainty.......... 25
5.3.4 Recommendations on estimation modeling.......... 25
5.4 Projections.......... 26
5.4.1 Stock-recruitment relationship (SRR).......... 26
5.4.2 Short-term projections.......... 27
5.4.3 Long-term projections.......... 27
5.4.4 Recommendations on projections.......... 27
6. Pacific Cod – Honshu Northern Pacific Stock (PC-HNP).......... 28
6.1 Data.......... 29
6.1.1 Catch-at-age.......... 29
6.1.2 Relative abundance indices.......... 30
6.1.3 Hatchery releases.......... 30
6.1.4 Recommendations on data.......... 31
6.2 Biology.......... 31
6.2.1 Stock structure and distribution.......... 31
6.2.2 Natural mortality.......... 32
6.2.3 Maturity.......... 32
6.2.4 Growth.......... 32
6.2.5 Recommendations on biology.......... 32
6.3 Estimation Model.......... 33
6.3.1 VPA model.......... 33
6.3.2 Model diagnostics.......... 33
6.3.3 Uncertainty.......... 33
6.3.4 Recommendations on estimation modeling.......... 34
6.4 Projections.......... 34
6.4.1 Stock-recruitment relationship (SRR).......... 34
6.4.2 Short-term projections.......... 35
6.4.3 Long-term projections.......... 36
6.4.4 Recommendations on projections.......... 36
7. Pointhead Flounder – Northern Hokkaido (PF-NH).......... 36
7.1 Data.......... 36
7.1.1 Catch.......... 37
7.1.2 Relative abundance indices.......... 37
7.1.3 Other data.......... 38
7.1.3 Recommendations on data.......... 38

7.2 Biology 39
7.2.1 Stock structure and distribution 39
7.2.2 Biological parameters 39
7.2.5 Recommendations on biology 39
7.3 Estimation Model 40
7.3.1 State-space surplus production model 40
7.3.2 Model diagnostics 41
7.3.3 Uncertainty 41
7.3.4 Recommendations on estimation modeling 41
7.4 Projections 41
7.4.1 Recommendations on projections 42
8. References 42
9. Appendix 1 – Terms of Reference 42

1. Summary

Assessments of Japanese Amberjack (*Seriola quinqueradiata*), Japanese Cod (*Gadus macrocephalus*) and Pointhead Flounder (*Cleisthenes pinetorum*) stocks were conducted by the Japan Fishery Research and Education Agency (FRA) in 2023. These four stock assessments were on the: 1) Japanese Amberjack (JA) ; 2) Pacific Cod - Honshu Northern Sea of Japan (PC-HNSJ); 3) Pacific Cod - Honshu Northern Pacific (PC-HNP); and 4) Pointhead Flounder – Northern Hokkaido (PF-NH). An independent peer review meeting of these assessments was organized by the Secretariat of Peer Review of the FRA in November 2024. This report is an independent review of the four assessments based on my personal opinions of the submitted documents and clarifications during the meeting. The primary aims for this review were to evaluate if the current stock assessments were based on the best scientific information available (BSIA) and recommend improvements for future stock assessments.

Based on the assessment documents and discussions during the meeting, this review concluded that the results of the PC-HNSJ, PC-HNP, and PF-NH stock assessments were based on BSIA. However, the JA stock assessment was considered to be not based on BSIA. The four assessments could be grouped into three groups, with both Pacific cod assessments being relatively similar but the JA and PF-NH assessments being very different.

The PC-HNSJ and PC-HNP assessments were both considered BSIA and were generally similar in terms of data availability, data processing, assumptions made, and modelling approaches. Both assessments used virtual population analysis (VPA) models (i.e., cohort analysis) with catch-at-age data that were based on adequate sampling and aging, and matched the basic assumptions of a VPA model. In addition, the model fitted abundance indices from scientific surveys or standardized catch-per-unit-effort (CPUE) of commercial fisheries. Although both assessments were considered BSIA, this review provided some recommendations that could improve future assessments as well as identified some issues with the current assessments that would require additional work (e.g., natural mortality, uncertainty).

Similar to the PC-HNSJ and PC-HNP, the JA also took a VPA modelling approach. However, in the JA case, the catch-at-age data were not based on adequate sampling and aging, and substantially violated the basic assumptions of a VPA model. Therefore, the results of the JA stock assessment were considered to be not based on BSIA. Critically, a VPA model assumes that the catch-at-age is known without error. Developing the catch-at-age time series depends largely on two components: 1) the total catch; and 2) the age composition of the catch. Both these components of the JA stock assessment were not considered to be BSIA. The JA assessment used a total catch time series that was a complex of three species but did not separate out the JA stock even though there was data to do so. In addition, the JA assessment assumed that the recreational catch was negligible even though there was data suggesting otherwise. The fisheries on the JA stock were

numerous and complex. This may have led to the clearly inadequate sampling and aging of the catch. Instead, the age compositions were largely based on the landings in commercial size categories, which had varying definitions between fisheries as well as over time. Each commercial size category was assumed to contain a single age but an examination of the available age-length data clearly showed that this assumption was invalid. The age-length relationship was also thought to be variable through time and likely space. There was also a lack of an abundance index but there appeared to be data available to develop one from potential sources. There are numerous problems with the JA stock assessment and it was clear that a VPA model was inappropriate for the assessment. A substantial amount of work is required to improve all aspects of this assessment. It would be important to re-evaluate the overall approach of this assessment so that the modelling matches the available data and biological parameters, rather than making strong but invalid assumptions.

The PF-NH stock assessment used a production model approach, which was substantially different from the other three assessments in this review. A production model approach was used because the only data available were total catch and an abundance index based on standardized CPUE from a commercial fishery. Given the available data, the data preparation (e.g., index standardization), biological parameters (e.g., r prior), and modelling were all consistent and considered to be BSIA. However, there was one strange aspect of this assessment that was considered to be inappropriate and not BSIA. This was the inclusion and fitting of the results (i.e., estimated N-at-age) of an alternative or competing assessment of the same stock from a different scientific organization. From the viewpoint of this review, these estimated N-at-age were an undocumented and unreviewed (i.e., beyond the scope of the ToRs) source of information that were used inappropriately as data for this assessment. If information from this alternative assessment was necessary to be included, it would be critical to obtain the data used in the alternative assessment rather than using the results as data. If these two assessments used independent data, one could have used the estimated posteriors from one model as priors for the second but in this case, the data were clearly not independent. Nevertheless, an examination of models without information from the alternative assessment indicated that the results would have been more uncertain but otherwise similar. Therefore, this review considered the results of the PF-NH assessment to be BSIA, albeit with some reservations. One important recommendation for this assessment would be to investigate the importance of age structure on the results. One valuable consequence of using this modelling approach rather than the VPA models of the other assessments is that the projections were done with the same model and uncertainties are propagated appropriately. In contrast, the projections from other three assessments were done in an awkward manner and have largely inappropriate uncertainties.

2. Background

Assessments of Japanese Amberjack (*Seriola quinqueradiata*), Japanese Cod (*Gadus macrocephalus*) and Pointhead Flounder (*Cleisthenes pinetorum*) stocks were conducted by the Japan Fishery Research and Education Agency (FRA) in 2023. These four stock assessments were on the: 1) Japanese Amberjack (JA) ; 2) Pacific Cod - Honshu Northern Sea of Japan (PC-HNSJ); 3) Pacific Cod - Honshu Northern Pacific (PC-HNP); and 4) Pointhead Flounder – Northern Hokkaido (PF-NH). All four assessments were led by the Fisheries Stock Assessment Center, Fisheries Resources Institute of the FRA, in collaboration with other scientific organizations.

The four assessments could be grouped into three groups, with both Pacific cod assessments being relatively similar but the JA and PF-NH assessments being very different. In general, the assessments consisted of development of data and estimation models, estimation of management quantities (e.g., SSB, reference points, stock status), short-term projections, and long-term projections. The short-term projections were used to bring forward the estimated population from the terminal year of the estimation model to the current year for management purposes. The long-term projections were used to test the robustness of the harvest control rules (HCRs) to reach management objectives.

3. Review Process

An independent peer review of these assessments was organized by the Secretariat of Peer Review of the FRA under the specified Terms of Reference (ToR, Appendix 1). The review panel consisted of one reviewer from Japan, who was appointed by the Japanese Society of Fisheries Science, and two overseas reviewers from the National Oceanic and Atmospheric Administration of the USA, including myself. The in-person meeting was held during November 12-14, 2024, and included English-Japanese translators to ease language difficulties.

Assessment documents were translated from Japanese to English, and submitted to reviewers before the meeting. In order to help FRA scientists prepare for the meeting, initial questions on the assessments were submitted prior to the meeting, with follow up questions during the meeting. The documentation for each assessment consisted of a main assessment document, and depending on the assessment, several appendices.

This report is an independent review of the five assessments based on my personal opinions of the submitted documents and clarifications during the meeting. There was no communication between the other reviewers and I, with regards to this report. The submitted assessment documents were not detailed enough to conduct a thorough technical 'desktop' review. The lack of detailed documentation and time made it difficult to thoroughly review all aspects of the assessments. Nevertheless, this review attempts to cover the specified ToRs. It should be noted that there was no opportunity to review and recommend alternative model configurations for the models. Instead, the primary aims for this review were to evaluate if the current data and stock assessments were based on the best scientific information available (BSIA), and recommend improvements for future stock assessments.

This report is subdivided into separate sections for each stock and a section summarizing the overall findings. As specified in the ToRs, the data, biology (i.e., information and parameters), methodology, and results from the assessment and projection models of each stock were reviewed and evaluated. The assessment of the Japanese Amberjack (JA) stock was reviewed first, followed by the Pacific cod stocks (PC-HNSJ and PC-HNP), and the Pointhead Flounder (PF-NH) stock. If a comment was applicable to multiple stocks, the reader will be referred to earlier sections whenever possible to reduce the report length and repetitiveness. However, the recommendations for each stock were kept separate for the sake of clarity.

4. Japanese Amberjack (JA)

The JA stock assessment took a VPA modelling approach. Critically, a VPA model assumes that the catch-at-age is known without error. However, the catch-at-age data in this assessment were not based on adequate sampling and aging, and substantially violated the basic assumptions of a VPA model. Therefore, the results of the JA stock assessment were considered to be not based on BSIA. There were numerous problems with this stock assessment and it was clear that the current modelling approach was inappropriate. A substantial amount of work will be required to improve all aspects of this assessment. It would be important to re-evaluate the overall approach of this assessment so that the modelling matches the available data and biological parameters, rather than making strong but invalid assumptions.

4.1 Data

A VPA model (i.e., cohort analysis) without tuning indices was used to estimate the population dynamics of the JA stock during the historical period. A VPA model assumes that the catch-at-age is known without error. Therefore, it was critical to examine the preparation of the catch-at-age data and determine if this assumption was substantially violated. This review found that the current data quality and data preparation for this assessment were relatively poor and would require substantial improvements for future assessments (see Section 4.1.3).

4.1.1 Catch-at-age

Developing the catch-at-age time series depends largely on two components: 1) the total catch in numbers; and 2) robustly sampling the age composition of the catch. Substantial discussion during the meeting was focused on this part of the assessment. Overall, this review found that the catch-at-age data used for the JA stock assessment were not BSIA because both the total catch and the age composition estimates were considered to be not BSIA. The development of the catch-at-age data for the JA assessment was highly complex and not well documented.

This review found that the total catch estimates of JA were not BSIA because: 1) the total catch estimates were of a non-negligible mix of three species; and 2) non-negligible amounts of recreational catch were assumed to be negligible. The Japanese national catch statistics of amberjack are currently a mix of three species (Japanese Amberjack, Yellowtail Amberjack, and Greater Amberjack). The assessment currently assumes that 100% of the catch are made up of JA but recent estimates found that approximately 10% of the catch are of Yellowtail Amberjack and Greater Amberjack. The authors of the assessment reported there were sufficient data in recent years to separate the catch by species but the algorithm to do so was still preliminary. This review found that 10% of the catch being other species was non-negligible and neglecting this discrepancy was not appropriate. Sensitivity runs with alternative catch scenarios show substantial changes in estimated SSB. It was recommended to continue the development of the algorithm to estimate JA-specific catch, and use JA-specific catch for future assessments. These algorithms may have to be region-specific depending on the region-specific data. If there continues to be substantial uncertainty in the JA catch, it was recommended to develop catch scenarios to encompass the uncertainty of the assessment. In addition, there were occasional surveys of JA recreational catch, which found that JA was a popular species for recreational fishing. It was estimated that the JA recreational catches were approximately 5-6% of the total estimated catch annually. Although there have been insufficient surveys to provide annual estimates of JA recreational catches, it was recommended to use the survey results to develop scenarios of JA recreational catch for future assessments. It was also recommended to develop and conduct regular surveys of recreational catch of this and other species.

There was discussion on the potential JA catches by fisheries of neighbouring countries. It was noted that the JA catches by South Korea were included in the assessment. However, there were also amberjack catches by Taiwanese fisheries but the amounts were considered to be negligible (<1% of the total JA catches) for this assessment, especially given that the species identification of the catch was uncertain. Other neighbouring countries either reported zero JA catches or their catches were unknown. Therefore, this review found that the treatment of non-Japanese JA catches were BSIA but recommended that work continue on identifying potential sources of non-Japanese JA catches and collaborating with neighbouring countries to quantify these catches. One interesting aspect of the JA assessment was the treatment of age-0 fish (mojako) caught and used as feedstock in aquaculture operations. While it was not documented clearly, the assessment authors clarified that although previous assessments did not include mojako catches, mojako catches have been included in the JA total catch since the 2021 assessment.

This review found that the age compositions of the JA catch were not BSIA because: 1) there was insufficient sampling of the JA catch; and 2) the use of cohort slicing to assign single ages to a single size or market category. The development of the age compositions of the JA catch was relatively complex because there were low levels of size sampling of the catch. Most JA catch was grouped into market categories but large fish were auctioned as individual fish. Different prefectures may also have different definitions of market categories, and it was also unclear if the definitions of market categories remained consistent over time within the same prefecture. It was recommended to conduct more size sampling of the JA catch, especially within market categories to examine if the market categories were consistent and reasonable proxies for fish size. The catch in each market category and size class was subsequently converted into age compositions for each category, and compiled into age compositions for the entire stock. The stock assessment assumed that each size class and /or market category was equivalent to a single age class. However, an examination of the available size and age data showed that this assumption was not valid, with single size classes clearly containing several age classes. If the size sampling continues to be inadequate, there will be continued dependence on market categories as proxies for size classes. If this is the case, it was recommended to develop age-length keys, age-weight keys, and age-market category keys for each appropriate fishery, area, and time period, depending on the variability of these keys through space, time, and fishery.

4.1.2 Relative abundance indices

No tuning indices were used in the JA assessment. Several potential sources of data for relative abundance indices were discussed. After much discussion, it was recommended to develop a relative abundance index for adult JA from the logbook data of the purse seine fishery targeting JA spawning aggregations. In addition, it was recommended to examine the mojako abundance survey data to potentially develop a recruitment index for the stock. It was also reported that there was substantial amounts of tagging and recapture data available for this stock. It was noted that tagging and recapture experiments and data, if done well, can be highly informative of the stock's population dynamics and/or used to develop abundance indices. It was therefore recommended that work be done on developing an abundance index from the tagging and recapture data, and/or fitting the tagging and recapture data directly in the assessment model.

4.1.3 Recommendations on data

Based on the above findings, there were several recommendations for improving the data used for the JA assessment:

1) It was strongly recommended to continue the development of estimates of JA-specific total catch from the Japan national catch statistics of amberjack, which currently consists of the catch of three amberjack species. These estimates may have to be region-specific depending on the region-specific data.

2) It was strongly recommended to use JA-specific catch for future assessments.

3) It was strongly recommended to develop catch scenarios to encompass the uncertainty of the assessment, if there continues to be substantial uncertainty in the JA catch.

4) It was also recommended to develop and conduct regular surveys of recreational catch of this and other species.

5) It was strongly recommended to use the recreational survey results to develop scenarios of JA recreational catch and include these catches in future assessments.

6) It was recommended that work continue on identifying potential sources of non-Japanese JA catches and collaborating with neighbouring countries to quantify these catches.

7) It was strongly recommended to conduct more size sampling of the JA catch.

8) It was strongly recommended to examine if the market categories were consistent and reasonable proxies for fish size for all prefectures and fisheries, where catch in market categories is the primary data and size sampling is insufficient.

9) It was strongly recommended to develop age-length keys, age-weight keys, age-market category, length-market category, and weight-market category keys for each appropriate fishery, area, and time period, depending on the available data and the variability of these keys through space, time, and fishery.

10) It was strongly recommended to develop a relative abundance index for adult JA from the logbook data of the purse seine fishery targeting JA spawning aggregations.

11) It was recommended to examine the mojako abundance survey data to potentially develop a recruitment index for the stock.

12) It was recommended that work be done on developing an abundance index from the tagging and recapture data, and/or fitting the tagging and recapture data directly in the assessment model.

4.2 Biology

4.2.1 Stock structure and distribution

The stock structure for this assessment was considered BSIA, given the current available information. However, it was encouraged to continue research on this topic.

4.2.2 Natural mortality

There was substantial discussion on the natural mortality (M) schedule of the JA stock. For this assessment, it was assumed that M was fixed at 0.6 y^{-1} for the mojako stage (first 6 months of the age-0 fish) and at 0.3 y^{-1} for all ages after that. The fixed M of 0.3 y^{-1} for all post-mojako ages was based on the observation that maximum age for this stock was approximately 7-9 years, and using a meta-analytic relationship of 2.5/Amax (Tanaka 1960), that this translates to an M of 0.357 - 0.278 y^{-1}, which was subsequently rounded to an M of 0.3 y^{-1}. The authors reported that they have investigated alternative meta-analytical relationships that have been recently developed (e.g., Then et al. 2015; Thorson 2020a; Cope and Hamel 2022), which suggested alternative M schedules, but the authors decided to maintain a fixed M of 0.3 y^{-1} for all ages after the mojako stage. The fixed M of 0.6 y^{-1} for the mojako stage was an assumption that the M of the mojako stage was double that of the other ages. It was reported that sensitivity runs with these alternative M schedules showed that the assessment results were highly sensitive to the assumed M schedule. As noted in Section 4.1, it was reported that there is substantial tagging and recapture data for this stock. The authors reported that some scientists consider that the M for the JA stock was likely to be low due to the high number of recaptures. However, no tagging data was provided to support this claim. After considering the importance of the M schedule and the availability of alternative meta-analytical relationships for M that are more recent and based on many more stocks than Tanaka (1960), this review considered that the justification for using the current M schedule was insufficient and not BSIA. In addition, the current assessment approach of using a fixed M, with arbitrarily chosen M schedules for sensitivity runs, were also found to be inappropriate.

It was strongly recommended that the M schedule for this stock to be re-evaluated and improved. It would be recommended to perform a meta-analysis using more recent empirical relationships or incorporating the metadata from Tanaka (1960) into more recent metadata (Then et al. 2015), and developing a prior for M. It was recommended that future assessments consider estimating M within the assessment model, with a prior for M. If estimating M within the assessment model is not possible, it was recommended to perform a series of sensitivity runs that represent the estimated uncertainty of the M values (i.e., the previously mentioned M prior) and reasonable, alternative M patterns (e.g., average M for all ages; Lorenzen size-specific M). It was also recommended that work be done to develop M priors from the tagging and recapture data and/or estimating M inside the assessment model by fitting the tagging and recapture data directly in the assessment model.

4.2.3 Maturity

The maturity of the JA stock was assumed to be 0% at age-0 and 1, 50% at age-2, and 100 % at age-3+. Current research is still ongoing with regards to the maturity schedule of the JA stock but the available information suggests that this maturity assumption is currently BSIA. It was recommended that work continue on the maturity schedule of this stock, as well as the variability through time and the relationship with growth.

4.2.4 Growth

Given the lack of size sampling for the JA stock (see Section 4.1.1) and the importance of market categories as proxies for size and age, the growth model was not important for this assessment. Nevertheless, it was recommended to include growth studies in the future to help with interpolating data gaps in the various keys (see Recommendation #9 in Section 4.1.3).

4.2.5 Recommendations on biology

Based on the above findings, there are several recommendations for improving the biological assumptions and parameters used for the JA assessment:

1) It was strongly recommended that the natural mortality for the JA stock be re-evaluated and improved.

2) It was recommended to perform a meta-analysis for the M of the JA stock using more recent empirical relationships or incorporating the metadata from Tanaka (1960) into more recent metadata, and developing a prior for M.

3) It was recommended that future assessments consider estimating M within the assessment model, with a prior for M.

4) If estimating M within the assessment model is not possible, it was recommended to perform a series of sensitivity runs that represent the estimated uncertainty of the M values (i.e., the previously mentioned M prior) and reasonable, alternative M patterns (e.g., average M for all ages; Lorenzen size-specific M).

5) It was also recommended that work be done to develop M priors from tagging and recapture data and/or estimating M inside the assessment model by fitting the tagging and recapture data directly in the assessment model.

6) It was recommended that work continue on the maturity schedule of this stock, as well as the variability through time and the relationship with growth.

7) It was recommended to include growth studies in the future to help with interpolating data gaps in the various recommended keys (see Recommendation #9 in Section 4.1.3).

4.3 Estimation Model

In Section 4.1, this review found that the primary data inputs for the JA stock assessment were not BSIA. Given that the estimation model was a VPA, which makes the critical assumption that the catch-at-age is known without error, this review also found that the results of the VPA were not BSIA. This review did not spend much time on the VPA methodology during the meeting because it was clear that the data issues identified precluded the VPA results from being considered as BSIA. Nevertheless, this review briefly examined the methodology of the VPA model. Overall, it would be important to re-evaluate the overall approach of this assessment so that the modelling matches the available data and biological parameters, rather than making strong but invalid assumptions.

4.3.1 Estimation model

The estimation model was a mostly standard cohort analysis without tuning indices. The key difference is the inclusion of two 6 month periods in the age-0 period (one for mojako, and one for older age-0 fish) but the model adhering to an annual time-step. This led to an awkward "double-counting" of age-0 fish in the N-at-age calculations, which is not considered good practice and is not BSIA. Instead, the model should perhaps be based on a 6-month or seasonal time-step for the age-0 or all ages. In general, this type of cohort analysis without tuning indices is inflexible and is largely dependent on the assumption that the catch-at-age is known without error and that the assumed M-at-age is appropriate. As stated above, both assumptions were likely violated.

Two data-driven modelling choices were made that potentially could be improved in future assessments. First is the choice of age-3+ as the plus group, which means that the F for age-3+ is assumed to be the same as age-2. However, a back of the envelope calculation with the catch suggests that this assumption may not be valid, given the change in fishery targeting at these ages. This modelling choice was reported to be largely due to the lack of catch-at-age data at ages beyond age-2. Second, the VPA model starts in 1994 but the catch data can be extended back to at least the early 20th century and the historical period show much greater contrast in catch (Fig 3-3 in the assessment document) than the current modelling period. It was reported that the start year of 1994 was dictated by the availability of sales slip data, which allows for the development of catch-at-age data. The pre-1994 data consists only of total catch. The greater contrast in the past allows for a better estimation of the population productivity parameters and population projections, if the productivity of the stock has remained relatively stable. It is unclear if the JA stock productivity varies with the environment like sardines and anchovies, or is relatively stable like groundfish species. Nevertheless, the development of a longer model encompassing the high contrast periods would allow for the testing of stock productivity hypotheses and a better quantification of projection uncertainty over the long term.

The authors of the assessment reported there is interest to change the estimation model from a cohort analysis to an integrated modelling approach like Stock Synthesis (Methot and Wetzel 2013) or Woods Hole Assessment Model (Stock and Miller 2021). However, the authors also reported that this change is only likely to happen over the long run. This idea of using an integrated modelling approach was discussed at length during the meeting, and this review would strongly recommend this change in approach for this stock and would substantially improve the stock assessment. In general, it would likely be easier and less costly to meet the data requirements for an integrated model than a VPA model. For example, the fisheries for JA are highly complex and a lot of work will be required to develop the catch-at-age data adequately from the market categories data in order for the assumptions of a VPA model not to be violated. On the other hand, as long as the length or weight range of a market category are known, the catch in weight or numbers of each category can simply be used in an integrated model without conversion to catch-at-age beforehand. The requirement for an integrated model would be for the development of length-weight and age-length (growth) relationships. However, these relationships are often easier and less expensive to develop than sampling and aging each market category appropriately. In addition, using an integrated model would allow for an experimental model to start earlier with only total catch data because the pre-1994 selectivity of each fishery could be assumed to have been the same as post-1994. An integrated model would also allow for a relaxation of the assumption that the F of age-3+ is the same as age-2 because the selectivity of each fishery could be more flexibly modeled to fit the data. In addition, the auction data of the larger fish could also be easily added to the model without much additional work, so that the selectivity of the larger fish could be better modelled. In addition, an integrated model would be able to estimate M within the estimation model by including M priors and/or tagging data.

Regardless of whether future assessments continue using a VPA model or an integrated model, developing an abundance index from fishery and/or tagging data would be critical. However, an integrated model will be able to incorporate the uncertainties of the abundance index and other sources of data, and parameters, and estimate uncertainties with model parameter estimates and management quantities. In addition, the integrated model is a forward-simulating model and can thus include all these uncertainties into the projections, if needed, into the forward projections.

4.3.2 Model diagnostics

Given that the data and VPA model were not considered BSIA, there was minimal interest in examining the model diagnostics in any detail. There were moderate levels of retrospective pattern for the model, especially for the Fs at age-2 and 3+, which does suggest that there may be some inconsistency in the data, and that the assumption of F at age-2 being equal to F at age-3+ may be inappropriate. Although these levels of retrospective pattern are not high relative to many other models, it is still unexpectedly high for a model with only catch-at-age data and no abundance index.

4.3.3 Uncertainty

The treatment of uncertainty in the JA assessment was poor. No uncertainties in the data and parameters were considered. The estimated population structure (N-at-age) and F-at-age in the terminal years were assumed to be known without error to estimate the stock-recruitment relationship (SRR) and projections. Given that the estimated N-at-age and F-at-age have uncertainties that were unaccounted for, the estimated probability distributions from the projections will likely be erroneous. Along with the N-at-age and F-at-age, the uncertainty in the estimated recruitment and spawning biomass also appeared to be neglected. It is understood that it is more difficult to work with uncertainty in a VPA model, especially when the data were not developed according to the assumptions made (e.g., catch-at-age is known without error). It may be beneficial to explore the use of integrated models, which can more easily include the uncertainty in the data and biological processes throughout the model (see Section 4.3.1).

4.3.4 Recommendations on estimation modeling

Based on the above findings, there are several recommendations for improving the estimation models used for the JA assessment:

1) Given that the current sampling program and data for the JA stock assessment is relatively poor and not appropriate for a VPA model (see Section 4.1.1), it is strongly recommended to consider using an integrated modelling approach (e.g., Stock Synthesis, Woods Hole Assessment Model) for future assessments instead of a VPA model.

2) It would be important to re-evaluate the overall approach of this assessment so that the modelling matches the available data and biological parameters, rather than making strong but invalid assumptions.

4.4 Projections

The outputs from the VPA model were used to estimate the stock-recruitment relationship (SRR) and used in the short- and long-term projections. Given that the data and results from the VPA model were not considered to be BSIA (Section 4.3), this review also concluded that the results of the projections were not BSIA. It should be noted that the details of the projections were based at least in part on discussions during a stakeholder meeting, and maybe outside the scope of this review. Nevertheless, this review briefly examined the methodology of the projections.

4.4.1 Stock-recruitment relationship (SRR)

The estimated recruitment and SSB from the VPA model was used to develop the appropriate SRR to use for calculating biological reference points (BRPs) and for the future projections. It was reported that the assessment authors follow guidelines from FRA to estimate the SRR in a stock assessment. These guidelines appear to be used across all FRA stock assessments and are beyond the ToRs for this review. Therefore, this section does not consider whether the SRR was BSIA. However, it was noted that the historical recruitment and SSB were assumed to be estimated without error and therefore uncertainties in the SRR and future projections were under-estimated.

4.4.2 Short-term projections

The terminal year of the estimation model is 2022 but the allowable catch for the current fishing year is based on SSB in 2024. Therefore two years of projections were required to bring forward the estimated N-at-age in 2022 to the projected spawning biomass in 2024. Given that the results from the VPA model were not considered to be BSIA, it follows that the results of these projections were also not considered to be BSIA because the projections relied on the VPA model results for inputs.

In terms of the methodology, the projection models are relatively straightforward and appropriate, given the information available. However, the problems with using short-term projections to estimate SSB in 2024 and set allowable catch should be noted. One of the problems is that the projection results are highly sensitive to the estimates of age-0 and age-1 fish in the terminal year of the VPA model but these estimates are highly uncertain due to the lack of observations, which is common for assessments without well-designed surveys of these age classes in the terminal years. It should also be noted that the uncertainties in these projections are currently under-estimated because the N-at-age in the terminal year of the estimation model is assumed to be known without error, which is considered inappropriate. It is therefore important and recommended to incorporate that uncertainty in the N-at-age in the terminal year into the short term projections. The uncertainty in the projected recruitment is also likely under-estimated.

4.4.3 Long-term projections

One of the management objectives for the JA stock was the probability of exceeding the target reference point (MSY) over the next 10 years should not exceed 50%. In addition, there are management objectives and probabilities associated with the limit and ban reference points. Long-term projections out to 2054 were used to estimate the appropriate F-multiplier (e.g., relative to F MSY) to achieve the management objectives. Similar to the short-term projections, the results of these long-term projections were not considered to be BSIA because the inputs for the projections relied on the VPA model results, which were not considered to be BSIA.

The models used for the long-term projections were similar to the short-term projections, and were relatively straightforward and appropriate. However, the uncertainty for the long-term projections are currently under-estimated because the N-at-age in the terminal year of the estimation model is assumed to be known without error, which is considered inappropriate (Section 4.3.3). It is therefore important to incorporate that uncertainty in the N-at-age in the terminal year into both the short term and long-term projections. This would be especially important for calculating the probabilities of exceeding the limit and ban reference points, which are based on the tail of the probabilities.

4.4.4 Recommendations on projections

Based on the above findings, there are several recommendations for improving the SRRs and projections used for the JA assessment:

1) It was recommended that the recruitment and SSB estimates used to develop the SRRs have uncertainty associated with them.

2) It was recommended to incorporate that uncertainty in the N-at-age in the terminal year into the short term and long-term projections.

5. Pacific Cod – Honshu Northern Sea of Japan Stock (PC-HNSJ)

The PC-HNSJ assessment used a relatively standard tuned VPA modelling approach with catch-at-age data that were based on adequate sampling and aging, and matched the basic assumptions of a VPA model. The model was tuned by fitting an abundance index from standardized catch-per-unit-effort (CPUE) of a commercial fishery. One of the interesting aspects of this stock assessment was that the primary target for this stock were the adult male fish for their milt, which resulted in relatively large and mature fish being targeted.

5.1 Data

A tuned VPA model was used to estimate the population dynamics of the PC-HNSJ stock during the historical period. A VPA model assumes that the catch-at-age is known without error. Therefore, it was critical to examine the preparation of the catch-at-age data and determine if this assumption was substantially violated. In addition, it was also important to examine the tuning indices used in the assessment. This review considered that the data used in this assessment was BSIA but there were several recommendations for potential improvements.

5.1.1 Catch-at-age

The catch-at-age data for the PC-HNSJ stock assessment was considered to be BSIA because the total catch and age compositions for the stock assessment were adequate and considered to be BSIA. However, this review also made several recommendations that may improve future assessments.

The total catch of this stock was considered to be reasonably well-known and BSIA, given the stock structure assumptions. However, there were some concerns that the stock structure and distribution assumptions for this stock may be more uncertain than is currently assumed (see Section 5.2.1). Based on these discussions, this review recommended developing alternative catch scenarios based on alternative stock structure and distribution assumptions, especially with regards to the connectivity between this stock and the fish around Hokkaido. Importantly, the distribution of the entire stock appeared to be within the Japanese EEZ and the assumption of negligible non-Japanese sources of fishing mortality appeared to be reasonable. Important for discussions on the modelling (Section 5.3), it was noted that the current assessment had a start year of 2000 but the total catch (in weight) data for this stock extended back to at least the 1960s and the pre-2000 data had substantially larger contrast (Fig 3-1 in the assessment report). It was recommended to extend the total catch time series back to the entire time series.

For each year in the assessment, the annual total catch in weight for this stock were converted into catch-at-age in numbers by: 1) sampling fish lengths from market samples of the major fisheries on this stock; 2) length compositions from the various fisheries were raised to the catch and combined into a single length composition for the entire stock; 3) the combined length composition was converted into an age composition using an age-length key (ALK) for the entire stock and the specific year; and 4) total catch in weight was converted into catch-at-age in numbers using the age composition and a fixed length-weight relationship. Overall, the above process and the resulting catch-at-age data were considered to be BSIA. However, it was recommended to examine differences in the age-length relationships for the major fisheries, and consider using fishery- and year-specific ALKs, especially if the ALK variability between fisheries is high. Season- or month-specific ALKs might be worth examining but would likely not make a substantial difference because most of the catch are taken in a single season (winter). In addition, all the catch appears to be age-3+ and seasonal differences in the age-length relationships are likely to be less apparent at those older ages. It was also recommended to examine if season-, year- and fishery-specific length-weight relationships would be substantially more representative than the single fixed length-weight relationship for all years and seasons. It was also recommended to investigate and report the estimated aging error by performing reads by multiple readers.

5.1.2 Relative abundance indices

A single tuning index for the 2000-2022 period was used in the PC-HNSJ assessment. The index was based on the logbook data from large offshore trawlers and standardized with a delta two step process made up of two generalized additive models (GAMs) with one model for the probability of zero catch, and another model for the CPUE of non-zero catch effort. Both models had explanatory variables with year, month, prefecture, fishing target, 2D splines of latitude and longitude, and several interactions with year and month. After substantial discussion on the standardization model, this review considered the tuning index to be BSIA.

Important for discussions on the modelling (Section 5.3), it was noted that the logbook data for the large offshore trawlers extended back to the 1970s and the pre-2000 data appeared to indicate larger contrasts than during the 2000-2022 assessment period. It was reported that there were preliminary attempts to extend the index to before 2000 but there were problems with standardizing the CPUE with the current model structure. After some discussion, it was considered important to extend the index to before 2000 and it was recommended to develop a standardized CPUE of the offshore trawlers using alternative spatiotemporal models (e.g., INLA, VAST), which may allow a longer standardized abundance index. Alternatively, a simpler standardization model structure may allow a longer standardized CPUE time series.

It was noted that there were two alternative sources of CPUE data: 1) gillnet fishery; and 2) bottom trawling fishery, but not used to develop abundance indices for the assessment. It was reported that previous assessments had used these two data sources but data quality problems were detected in these two data sources and not currently used. However, there were concerns that with the number of large offshore trawler vessels steadily decreasing due to aging of the vessel crews and captains, there may be a need to use these alternative CPUEs in the future. It was recommended to consider developing alternative abundance indices, especially if the current abundance index from the large offshore trawlers become more unreliable.

5.1.3 Recommendations on data

Based on the above findings, there are several recommendations for improving the data used for the PC-HNSJ assessment:

1) It was recommended to consider extending the total catch time series back to the entire time series, especially if an integrated modelling approach is taken in the future.

2) It was recommended to examine differences in the age-length relationships for the major fisheries, and consider using fishery- and year-specific ALKs, especially if the ALK variability between fisheries is high.

3) It was recommended to examine if season-, year- and fishery-specific length-weight relationships would be substantially more representative than the single fixed length-weight relationship for all years and seasons.

4) It was also recommended to investigate and report the estimated aging error by performing reads by multiple readers.

5) It was recommended to develop a standardized CPUE of the large offshore trawlers using alternative spatiotemporal models (e.g., INLA, VAST), which may allow the standardized abundance index to be extended back to the start of the logbook data.

6) It was recommended to consider developing alternative abundance indices, especially if the current abundance index from the large offshore trawlers become more unreliable.

5.2 Biology

5.2.1 Stock structure and distribution

The stock structure for this stock assessment appeared to be appropriate, given the currently available information. However, there was some uncertainty in the current stock boundaries, especially the northern boundary between this stock and the fish around Hokkaido and Mutsu Bay. Based on the presented evidence, there appeared to be some (likely low levels) movement between the fish around Hokkaido, Mutsu Bay, and the Sea of Japan. In addition, the population around Hokkaido and Mutsu Bay appeared to be increasing and was likely larger than this stock. Therefore, even low levels of migration from Hokkaido and Mutsu Bay could result in large impacts on the estimated population dynamics of this stock. There was also some uncertainty in the southern boundary of this stock but the catch and population levels in the southern boundary area appeared to be low. Therefore, the impact of the uncertainty around the southern boundary is likely to be low. After some discussion, this review recommended that work continue on the stock structure and distribution of this stock, especially with regards to the links between this stock and the fish around Hokkaido and Mutsu Bay. In addition, it was recommended that future assessments develop sensitivity models that show the impact of alternative stock structure assumptions on the estimated population dynamics and management quantities of this stock.

5.2.2 Natural mortality

The M of the PC-HNSJ assessment was assumed to be 0.28 y^{-1} for all ages, which was based on a meta-analysis by Tanaka (1960) and a maximum age of 9 years. This was similar in approach to the M schedule for the JA and PC-HNP stocks, and similarly impactful on assessment results. Therefore, the recommendations were also similar.

It was strongly recommended that the M schedule for this stock to be re-evaluated and improved. It would be recommended to perform a meta-analysis using more recent empirical relationships or incorporating the metadata from Tanaka (1960) into more recent metadata (Then et al. 2015), and developing a prior for M. It was recommended that future assessments consider estimating M within the assessment model, with a prior for M. If estimating M within the assessment model is not possible, it was recommended to perform a series of sensitivity runs that represent the estimated uncertainty of the M values (i.e., the previously mentioned M prior) and reasonable, alternative M patterns (e.g., average M for all ages; Lorenzen size-specific M).

5.2.3 Maturity

The maturity of this stock appeared to be BSIA and uncertainty in maturity would not likely have been substantially impactful on the estimated population dynamics of this stock.

5.2.4 Growth

Given that there was annual sampling and aging of otoliths, there was minimal impact of the growth model on this assessment. However, it was recommended to develop a growth model that incorporates the inter-annual variability of the age-length relationships. While doing so is not highly useful if the annual sampling program continues in the future, it may be useful if and when there are problems with the sampling and aging of otoliths. In addition, such a growth model would be very useful if future assessments have start years before 2000.

5.2.5 Recommendations on biology

Based on the above findings, there were several recommendations for improving the biological assumptions and parameters used for the PC-HNSJ assessment:

1) It was recommended that work continue on the stock structure and distribution of this stock, especially with regards to the links between this stock and the fish around Hokkaido and Mutsu Bay.

2) It was recommended that future assessments develop sensitivity models that show the impact of alternative stock structure assumptions on the estimated population dynamics and management quantities of this stock.

3) It was strongly recommended that the M schedule for this stock to be re-evaluated and improved.

4) It would be recommended to perform a meta-analysis using more recent empirical relationships or incorporating the metadata from Tanaka (1960) into more recent metadata (Then et al. 2015), and developing a prior for M.

5) It was recommended that future assessments consider estimating M within the assessment model, with a prior for M.

6) If estimating M within the assessment model is not possible, it was recommended to perform a series of sensitivity runs that represent the estimated uncertainty of the M values (i.e., the previously mentioned M prior) and reasonable, alternative M patterns (e.g., average M for all ages; Lorenzen size-specific M).

7) It was recommended to develop a growth model that incorporates the inter-annual variability of the age-length relationships.

5.3 Estimation Model

This review found that the results of the VPA model of the PC-HNSJ stock were BSIA.

5.3.1 VPA model

The estimation model was a standard VPA model with a tuning index. The catch-at-age data and abundance index were considered BSIA (Section 5.1), and did not violate the assumptions for a VPA model. The estimated parameters appeared to be reasonable and the diagnostics for the VPA model were also adequate. Therefore, this review considered the results of the VPA model to be BSIA.

As noted in Section 5.1, the current assessment model and data has a start year of 2000 but the catch and abundance index appeared to have better contrast before 2000. The authors recognized that the estimation model would likely have better estimates of the stock's productivity if the model and data included the periods with greater contrast. It was reported that the main reason for a start year of 2000 was the lack of size and/or age data prior to 2000. The authors also reported an interest in developing state-space and/or integrated models for this stock. This review noted that if an integrated model was used and no size data before 2000 was available, it would still be reasonable to start the model earlier with just catch and abundance index data by assuming that the fisheries had a similar selectivity prior to 2000. Therefore, it was recommended that work continue on developing state-space and/or integrated models for the PC_HNSJ stock. It was also recommended that in combination with developing catch time series and abundance indices covering the period of high contrast before 2000, an estimation model (likely a state-space and/or integrated model) be developed with an earlier start year that includes the periods of high contrast.

5.3.2 Model diagnostics

For the retrospective analysis, there were low levels of variable bias in the estimated values (SSB, recruitment, biomass-at-age and F-at-age) with a reduction in the time series length. There were also no clear patterns in the bias. The estimated parameters also appeared to be reasonable. The data inputs were also considered BSIA, and did not violate the assumptions for a VPA model. Based on the above evidence, this review found that the VPA model performed adequately and that the results were considered to be BSIA. However, given the importance of projections in the management of this stock, it was recommended to use hindcasting to test and report the prediction skill of the model (e,g., Kell et al. 2016) in the future.

5.3.3 Uncertainty

Similar to the JA and PC-HNP stock assessments, the treatment of uncertainty in the PC-HNSJ assessment was inadequate. There were large uncertainties throughout the assessment but minimal consideration of uncertainty in the data, modeling, and results of the assessment. For example, the estimated population structure (N-at-age) and F-at-age in the terminal years were assumed to be known without error and used as such in the projections. Therefore, the estimated probability distributions from the projections were likely to be erroneous. It is understood that it is more difficult to work with uncertainty in a VPA model. Similar to Section 5.3.1, it is recommended that work continue on developing state-space and/or integrated models, which can more easily include the uncertainty in the data and biological processes throughout the model.

5.3.4 Recommendations on estimation modeling

Based on the above findings, there were several recommendations for improving the estimation models used for the PC-HNSJ assessment:

1) It was recommended that work continue on developing state-space and/or integrated models for this stock assessment.

2) It was recommended that in combination with developing catch time series and abundance indices covering the period of high contrast before 2000, an estimation model (likely a state-space and/or integrated model) be developed with an earlier start year that includes the periods of high contrast.

5.4 Projections

The outputs from the VPA model were used to estimate the stock-recruitment relationship (SRR) and in the projections.

5.4.1 Stock-recruitment relationship (SRR)

It was reported during the meeting that the assessment scientists follow guidelines from FRA to estimate the SRR in a stock assessment. These guidelines appear to be used across all FRA stock assessments and are beyond the ToRs for this review. Therefore, this review did not consider whether the SRR was BSIA. However, it was noted that the estimated recruitments and SSBs were assumed to be known without error, which was considered inappropriate and likely resulted in the uncertainties of the SRRs being estimated too low.

5.4.2 Short-term projections

In terms of the methodology, the projection models were relatively straightforward and appropriate, given the information available. Therefore, the results from the short-term projections were considered to be BSIA. It should be noted that the uncertainties of the projections were under-estimated because the uncertainty in the SRR were likely underestimated (Section 5.4.1) and N-at-age in the terminal year of the estimation model were assumed to be known without error, which was considered inappropriate. It would be recommended to incorporate that uncertainty in the N-at-age into the short-term projections.

5.4.3 Long-term projections

Similar to the short-term projections, the projection models were relatively straightforward and appropriate, given the information available. Therefore, the results from the long-term projections were considered to be BSIA. However, the uncertainty for the long-term projections were under-estimated because the uncertainty in the SRR were likely underestimated (Section 5.4.1) and the N-at-age in the terminal year of the estimation model was assumed to be known without error, which was considered inappropriate. This would be especially important for calculating the probabilities of exceeding the limit and ban reference points, which are based on the tails of the probability distributions. It would be recommended to incorporate that uncertainty in the N-at-age into the long-term projections.

5.4.4 Recommendations on projections

Based on the above findings, there were several recommendations for improving the SRRs and projections used for the PC-HNSJ assessment:

1) It is recommended that the recruitment and SSB estimates used to develop the SRRs have uncertainty associated with them.

2) It is recommended to incorporate that uncertainty in the N-at-age in the terminal year into the short term and long-term projections.

6. Pacific Cod – Honshu Northern Pacific Stock (PC-HNP)

The stock assessment of the PC-HNP stock was similar to the PC-HNSJ assessment (Section 5). Both assessments used virtual population analysis (VPA) models (i.e., cohort analysis) with catch-at-age data that were based on adequate sampling and aging, and matched the basic assumptions of a VPA model. Therefore, similar comments were made for both assessments and references will be made to earlier PC-HNSJ sections whenever possible to reduce the report length and repetitiveness. However, the recommendations for each stock were kept separate for the sake of clarity. In contrast to the PC-HNSH fisheries, the PC-HNP fisheries did not substantially target adult males, and the age and size of the catch were generally younger and smaller.

6.1 Data

Similar to the PC-HNSJ stock assessment, a tuned VPA model was used as the estimation model, and hence, this review focused on the preparation of the catch-at-age data and relative abundance indices.

6.1.1 Catch-at-age

The approach to develop the catch-at-age data for the PC-HNP stock assessment was very similar to the PC-HNSK stock. Similarly, the catch-at-age data for the PC-HNP stock was also considered to be BSIA because the total catch and age compositions for the stock assessment were adequate and considered to be BSIA. Several similar recommendations were made that may improve future assessments.

The total catch data of this stock was considered to be reasonably well-known and BSIA, given the stock structure assumptions. However, there were some concerns that the stock structure and distribution assumptions for this stock may be more uncertain than is currently assumed (see Section 6.2.1). It was recommended that future assessments develop alternative catch scenarios based on alternative stock structure and distribution assumptions, especially with regards to the connectivity between this stock and the fish around Hokkaido. However, unlike the PC-HNSJ stock, the current assessment period of 1996-current for this stock contains periods of high contrast. Therefore, extending the model and data back further may be interesting but not critical for future assessments.

The processes to develop the age composition data of this stock were similar to the PC-HNSJ stock (see Section 5.1.1), and the age composition data of this stock were considered to be BSIA. Similar to the PC-HNSJ stock, it was recommended for the PC-HNP stock to examine differences in the age-length relationships for the major fisheries, and consider using fishery- and year-specific ALKs, especially if the ALK variability between fisheries is high. Season- or month-specific ALKs might also be worth examining because the primary age of the catch is relatively young. It was also recommended to examine if season-, year- and fishery-specific length-weight relationships would be substantially more representative than the single fixed length-weight relationship for all years and seasons. It was also recommended to investigate and report the estimated aging error by performing reads by multiple readers.

6.1.2 Relative abundance indices

Similar to the PC-HNSJ assessment, a single tuning index was used for the PC-HNP assessment. However, the PC-HNP assessment used an abundance index from an annual bottom trawl survey (1996-2022) during October-November. The survey design appears to overlap the habitat of the stock, and the fish caught are sampled appropriately (length, weight, otoliths, and gonads). The trawl data is converted to an age-specific abundance index using the area-density method, which is essentially the density of the fish (fish km^{-2}) multiplied by the area swept, combined with the age-composition data from the survey. The age-specific abundance index from the trawl survey was considered to be BSIA. It was recommended to examine the sources of uncertainty and estimate the uncertainty of the index of the abundance index from the bottom trawl survey.

In addition to the bottom trawl survey, there were several other candidate indices that were based on catch and effort data from commercial fisheries. While the indices from the commercial fisheries were not inconsistent with the trawl survey, the trawl survey was considered the better source of data for an abundance index.

6.1.3 Recommendations on data

Based on the above findings, there are several recommendations for improving the data used for the PC-HNP assessment:

1) It was recommended to examine differences in the age-length relationships for the major fisheries, and consider using fishery- and year-specific ALKs, especially if the ALK variability between fisheries is high.

2) It was recommended to examine if season-, year- and fishery-specific length-weight relationships would be substantially more representative than the single fixed length-weight relationship for all years and seasons.

3) It was also recommended to investigate and report the estimated aging error by performing reads by multiple readers.

4) It was recommended to examine the sources of uncertainty and estimate the uncertainty of the index of the abundance index from the bottom trawl survey.

6.2 Biology

6.2.1 Stock structure and distribution

Similar to the PC-HNSJ stock, the stock structure for the PC-HNP stock appeared to be appropriate, given the currently available information. However, there was some uncertainty in the current stock boundaries, especially the northern boundary between this stock and the fish around Hokkaido and Mutsu Bay. Similarly, after some discussion, this review recommended that work continue on the stock structure and distribution of this stock, especially with regards to the links between this stock and the fish around Hokkaido and Mutsu Bay. In addition, it was recommended that future assessments develop sensitivity models that show the impact of alternative stock structure assumptions on the estimated population dynamics and management quantities of this stock.

6.2.2 Natural mortality

The M of the PC-HNP assessment was assumed to be 0.357 y^{-1} for all ages, which was slightly higher than the PC-HNSJ assessment (0.28 y^{-1} for all ages). However, the approach by both assessments were the same, which was the meta-analysis by Tanaka (1960) but the PC-HNP assessment used a maximum age of 7 years (instead of 9 years for the PC-HNSJ stock). This was similar in approach to the M schedule for the JA and PC-HNSJ stocks, and similarly impactful on assessment results. Therefore, the recommendations were also similar.

It was strongly recommended that the M schedule for this stock to be re-evaluated and improved. It would be recommended to perform a meta-analysis using more recent empirical relationships or incorporating the metadata from Tanaka (1960) into more recent metadata (Then et al. 2015), and developing a prior for M. It was recommended that future assessments consider estimating M within the assessment model, with a prior for M. If estimating M within the assessment model is not possible, it was recommended to perform a series of sensitivity runs that represent the estimated uncertainty of the M values (i.e., the previously mentioned M prior) and reasonable, alternative M patterns (e.g., average M for all ages; Lorenzen size-specific M).

6.2.3 Maturity

The maturity of this stock appeared to be BSIA and uncertainty in maturity would not likely have been substantially impactful on the estimated population dynamics of this stock.

6.2.4 Growth

Similar to the PC-HNSJ stock, there was minimal impact of the growth model on this assessment. However, it was recommended to develop a growth model that incorporates the inter-annual variability of the age-length relationships. While doing so is not highly useful if the annual sampling program continues in the future, it may be useful if and when there are problems with the sampling and aging of otoliths.

6.2.5 Recommendations on biology

Based on the above findings, there were several recommendations for improving the biological assumptions and parameters used for the PC-HNP assessment:

1) It was recommended that work continue on the stock structure and distribution of this stock, especially with regards to the links between this stock and the fish around Hokkaido and Mutsu Bay.

2) It was recommended that future assessments develop sensitivity models that show the impact of alternative stock structure assumptions on the estimated population dynamics and management quantities of this stock.

3) It was strongly recommended that the M schedule for this stock to be re-evaluated and improved.

4) It would be recommended to perform a meta-analysis using more recent empirical relationships or incorporating the metadata from Tanaka (1960) into more recent metadata (Then et al. 2015), and developing a prior for M.

5) It was recommended that future assessments consider estimating M within the assessment model, with a prior for M.

6) If estimating M within the assessment model is not possible, it was recommended to perform a series of sensitivity runs that represent the estimated uncertainty of the M values (i.e., the previously mentioned M prior) and reasonable, alternative M patterns (e.g., average M for all ages; Lorenzen size-specific M).

7) It was recommended to develop a growth model that incorporates the inter-annual variability of the age-length relationships.

6.3 Estimation Model

This review found that the results of the VPA model of the PC-HNP stock were BSIA.

6.3.1 VPA model

The estimation model for the PC-HNP stock was the same as the PC-HNSJ stock. The model was a standard VPA model with a tuning index. The catch-at-age data and abundance index were considered BSIA (Section 6.1), and did not violate the assumptions for a VPA model. The estimated parameters appeared to be reasonable and the diagnostics for the VPA model were also adequate. Therefore, this review considered the results of the VPA model to be BSIA. Similar to the PC-HNSJ stock, it was recommended that the authors examine and consider state-space and/or integrated models for future PC-HNP stock assessments.

6.3.2 Model diagnostics

The model diagnostics (retrospective analysis, model fit, estimated parameters) for the PC-HNP assessment were generally similar to the PC-HNSJ assessment (see Section 5.3.2). Based on the above evidence, this review similarly found that the VPA model performed adequately and that the results were considered to be BSIA. However, given the importance of projections in the management of this stock, it was recommended to use hindcasting to test and report the prediction skill of the model (e,g., Kell et al. 2016) in the future.

6.3.3 Uncertainty

Similar to the JA and PC-HNSJ stock assessments, the treatment of uncertainty in the PC-HNP assessment was inadequate. The estimated probability distributions from the projections were likely to be erroneous. Similar to Section 5.3.1, it was recommended that the authors examine and consider state-space and/or integrated models for future PC-HNP stock assessments because these types of models can more easily include the uncertainty in the data and biological processes throughout the model.

6.3.4 Recommendations on estimation modeling

Based on the above findings, there were several recommendations for improving the estimation models used for the PC-HNP assessment:

1) It was recommended that the authors examine and consider state-space and/or integrated models for future PC-HNP stock assessments.

6.4 Projections

The outputs from the VPA model were used to estimate the stock-recruitment relationship (SRR) and in the projections.

6.4.1 Stock-recruitment relationship (SRR)

It was reported during the meeting that the assessment scientists follow guidelines from FRA to estimate the SRR in a stock assessment. These guidelines appear to be used across all FRA stock assessments and are beyond the ToRs for this review. Therefore, this review did not consider whether the SRR was BSIA. Nevertheless, it was recommended that the recruitment and SSB estimates used to develop the SRRs have uncertainty associated with them. In addition, the estimated SSB_{MSY} / SSB0 was very low, which suggested that the authors may want to re-examine the SRR and other parameters used in the calculation of management quantities.

6.4.2 Short-term projections

Similar to the PC-HNSJ assessments, the projection models were relatively straightforward and appropriate, given the information available. Therefore, the results from the short-term projections were considered to be BSIA (see Section 5.4.2). It was recommended that uncertainty in the N-at-age be incorporated into the short-term projections.

6.4.3 Long-term projections

The long-term projection models were also similar to the mPC-HNSAJ assessment, being relatively straightforward and appropriate, given the information available (see Section 5.4.3). Therefore, the results from the long-term projections were considered to be BSIA. Similar to the short-term projections, it would be recommended to incorporate that uncertainty in the N-at-age into the long-term projections.

6.4.4 Recommendations on projections

Based on the above findings, there were several recommendations for improving the SRRs and projections used for the PC-HNP assessment:

1) It is recommended that the recruitment and SSB estimates used to develop the SRRs have uncertainty associated with them.

2) It is recommended to incorporate that uncertainty in the N-at-age in the terminal year into the short term and long-term projections.

7. Pointhead Flounder – Northern Hokkaido (PF-NH)

The assessment approach of the PF-NH stock was different to the other stocks. In contrast to the other stocks, this assessment used a production model approach to assess the stock due to a lack of age or size structured data. Interestingly, it was reported that the Hokkaido Research Organization (HRO) had also conducted an independent assessment of the same stock, and that the HRO assessment was an age-structured assessment. It should be noted that results from the HRO assessment were used as inputs into this assessment but the HRO assessment was outside the scope of this review. In addition, it was reported that the data from the HRO assessment were not available for this assessment.

7.1 Data

The primary data sources for this assessment were total catch and abundance indices, which were consistent with a production model.

7.1.1 Catch

Overall, this review found that total catch data used by the PF-NH stock assessment were BSIA. Based on the available biological studies (see Section 7.2), the stock structure assumed in this assessment appeared to be appropriate, and there were no major fisheries on this stock outside of Japanese waters. The landings of Japanese fisheries catching this stock are relatively well monitored and can be considered to be known with minimal uncertainty, as was assumed in this assessment. However, it was noted that this is a bycatch species for some fisheries and there is likely some level of discard mortality for the stock but the level of discards is currently unknown. It is therefore recommended that future research be conducted on estimating the level of discards and include the discard mortality in future assessments.

7.1.2 Relative abundance indices

One relative abundance index was developed from month-aggregated logbook data of the offshore bottom trawl fishery. The logbook data consisted of the monthly effort (number of hauls) of individual vessels within each fishing area and the catches of various species from that effort. The data were first filtered by port (only Wakkanai, Esashi and Otaru ports were used) and depth (<340 m were used) to extract the relevant data for the stock. Subsequently, Generalized Additive Models (GAMs) were used to standardize the CPUE data. Importantly, the square-root transformed proportions of each species was used in a Principal Components Analysis (PCA) to develop two principal components as explanatory variables in the GAMs for changes in operational and targeting effects (Winker et al. 2014). Although Winker et al (2014) showed that this approach to account for targeting changes in multi-species fisheries resulted in improved abundance indices relative to nominal CPUE, there was some concern that the use of principal components that were based on proportions of each species (albeit square-root transformed) as explanatory variables, may lead to index hyperstability. After substantial discussion, it was considered that this index from the offshore bottom trawl fishery could be considered as BSIA. It is recommended that future research be conducted on the logbook data of the offshore bottom trawl fishery to further improve future abundance indices. It was also recommended to develop a bottom trawl survey for this and other demersal stocks in the area.

This PF-NH assessment also used the estimated female biomass from the HRO VPA model as a relative abundance index in this assessment. The female biomass was scaled with a fixed ratio (1.8) to account for the slightly smaller male fish and a 1:1 sex ratio. However, it was noted that more female fish are caught by the fishery and natural mortality is generally related to body size and therefore the assumed 1:1 sex ratio may not be appropriate. It was noted that the q of this index was assumed to have a prior of 1 (SD=0.3) and the estimated posterior was close to 1 (~0.73) and therefore approximated an absolute index. The use of this index was intended to ensure consistency with the HRO assessment. The HRO assessment model and data are beyond the scope of this review but the use of the estimated female biomass from the HRO model as an index is within the scope of this review. It is generally not appropriate to use the results from one assessment model and use it as an index in another assessment model. Therefore, this review considers the use of the estimated female biomass from the HRO assessment model to be inappropriate and cannot be considered to be BSIA. Subsequently, models that did not use estimated female biomass (see Section 7.3) showed consistent results with this assessment, albeit with larger uncertainties, and showed that the use of female biomass estimates to be largely unnecessary. Therefore, it is strongly recommended to make available the data from the HRO assessment for future assessments of the PF-NH stock. If such data continues to be unavailable, it is strongly recommended to not include the results from HRO assessment as data for this assessment.

7.1.3 Other data

Although beyond the scope of this review, the HRO assessment of the PF-NH stock utilized sex, size and/or age data to develop sex-specific catch-at-age data for the VPA model. Given the sexual dimorphism and sex-specific differences in fishing, these sex, size and/or age data, if made available, will likely be a major improvement to the current assessment. Therefore, it is strongly recommended to make available these sex, size and/or age data for future assessments of this stock.

7.1.3 Recommendations on data

Based on the above findings, there were several recommendations for improving the data used for the PF-NH assessment:

1) Most importantly, it is strongly recommended to make available the data, including sex, size and/or age data from the HRO assessment for future assessments of the PF-NH stock.

2) If data from the HRO assessment continues to be unavailable, it is strongly recommended to not include the results from HRO assessment as data for this assessment.

3) It is recommended that future research be conducted on estimating the level of discards and include the discard mortality in future assessments.

4) It is recommended that future research be conducted on the logbook data of the offshore bottom trawl fishery to further improve future abundance indices.

5) It is recommended to develop a bottom trawl survey for this and other demersal stocks in the area.

7.2 Biology

7.2.1 Stock structure and distribution

The stock structure for this stock assessment was considered to be appropriate and BSIA, given the currently available information. The growth, maturity and spawning season were reported to be different for different stocks of this species. An analysis of mtDNA showed significant differences between fish in the Sea of Japan and the Pacific Ocean around Hokkaido (Xiao et al. 2011). Substantial tagging research was reported during the meeting, with most fish being recaptured close to the release site. Although some records show migration of more than 100 km, a year after release, most fish are recaptured within identified stock distribution.

7.2.2 Biological parameters

Unlike the other stocks, the PF-NH assessment used a production model and hence used population growth rate as the primary biological parameter to define stock productivity. Two priors for population growth rate were developed using the R package called Fishlife (Thorson 2020b) with the same mean of 0.32 and SDs of 1.0 and 0.5, respectively. Similarly, two priors for the shape parameter were developed, with the same mean of 2.0 and SDs of 1.0 and 0.5, respectively. The shape parameter (n) is especially important for the calculation of reference points and a n of 2 is equivalent to a Schaefer model, with an MSY of 0.5*K. The posteriors appeared to be derived largely by the data rather than the priors. Nevertheless, it would have been useful to look into models with alternative priors. It was also unclear about the implications of $n < 1$ has on the shape of the production curve and reference points.

Overall, the population growth rate priors used for this assessment appeared to be appropriate and were considered to be BSIA. Although the priors were considered appropriate and BSIA, it was recommended to conduct future research on the biology of this stock and to develop stock-specific biological parameters and population growth rate priors, especially if future assessments used age-structured models.

7.2.5 Recommendations on biology

Based on the above findings, there were several recommendations for improving the biological assumptions and parameters used for the PF-NH assessment:

1) It was recommended to conduct future research on the biology of the PF-NH stock.

2) It was recommended to develop stock-specific biological parameters and population growth rate priors, especially if future assessments used age-structured models.

7.3 Estimation Model

This review found that the results of the state-space surplus production models for this assessment were not BSIA because the model fitted to the estimated female biomass from the HRO VPA-based model, which was considered to be inappropriate (see Section 7.1). However, it is important to note that a model that was not fitted to the estimated female biomass time series showed similar results, albeit with larger uncertainties.

7.3.1 State-space surplus production model

Stochastic state-space surplus Pella-Tomlinson production models (Pedersen and Berg 2017) were used to estimate the population dynamics of the PF-NH stock in this assessment. The models were fitted to the total catch and two abundance indices, one index based on the CPUE of the offshore bottom trawl fishery and the other from the estimated female biomass of the HRO VPA-based assessment model (see Section 7.1). The models and priors were discussed during the meeting and found to be appropriate and BSIA.

However, there was substantial concern that the female biomass time series estimated from the HRO assessment model were being fitted as data. Doing so was found to be inappropriate and not BSIA. In general, the results from one assessment model should not be used as input data for another assessment model of the same stock. It would have been reasonable if the first model used independent data (e.g., tagging data) to develop biomass posteriors, which were in turn used as priors for this assessment. However, it was clear that similar data sources, like total catch, were used in both the HRO VPA assessment and this assessment. The relatively low sigma estimated for the index (~0.06) could have been due to independent but consistent data but given that it was a VPA model, it was more likely due to similar data sources. Fitting the HRO VPA assessment results as data also likely resulted in underestimated uncertainty of the assessment results. The authors of this assessment were asked to present the results of alternative models that fitted to only the bottom trawl CPUE or VPA-estimated female biomass. An examination of these alternative models showed that fitting to only the bottom trawl CPUE resulted in similar results albeit with larger uncertainty. Therefore, the conclusion of this review was that the results of the base case models were not BSIA because fitting to the female biomass estimated by the HRO assessment was not appropriate and not necessary. However, it was noted that the results without using the HRO assessment VPA-derived female biomass, would not have been substantially different.

The base case results were based on an ensemble of 2 models with slightly different priors (SDs of 0.5 and 1.0 for n and r). Given the small difference in priors, it was not surprising to find that the results were highly similar. It was recommended that future assessments examine the full range of uncertainty, especially if an ensemble approach is being taken.

Given the sexual dimorphism and sex-specific differences in fishing, this review considered that future assessments would likely be improved if future assessments use integrated models to incorporate sex, size, and/or age structure. However, doing so is largely dependent on having available sex, size and/or age data for future assessments.

7.3.2 Model diagnostics

The authors presented a suite of model diagnostics, including convergence, residual, retrospective, factor, and jitter analyses. All these diagnostics indicated that both models performed well. An examination of the prior and posterior distributions showed that there was some influence of the assumed priors on the posteriors but not substantial enough to be a concern. It was recommended that future assessments test the prediction skill of the models using hindcasting because of the importance of 1 and 2 year predictions in the management of this stock. Given that the factor analysis showed the population dynamics of the stock were largely explained by surplus production and fishing mortality, the prediction skill is expected to be good.

7.3.3 Uncertainty

The treatment of uncertainty in the PF-NH stock assessment is generally good and considered to be BSIA. However, the fitting of the estimated female biomass likely led to an underestimation of the uncertainty of the assessment results (see Section 7.3.1). Otherwise, the uncertainty of the data and biological parameters were propagated appropriately through the model into the results.

7.3.4 Recommendations on estimation modeling

Based on the above findings, there are several recommendations for improving the estimation models used for the PF-NH assessment:

1) As recommended in Section 7.1, it is strongly recommended to not use the results from the HRO assessment as data for this assessment.

2) It is recommended that future assessments use integrated models to incorporate sex, size, and/or age structure but this is largely dependent on having available sex, size and/or age data.

3) It is recommended that future assessments examine the full range of uncertainty, especially if an ensemble approach is being taken.

7.4 Projections

A simple management strategy evaluation (MSE) was developed from the estimation models (Section 7.3.1). Based on the results of the MSE, management procedures and reference points were proposed. However, it was noted that the MSE, management procedures and reference points are beyond the scope of this review and are not evaluated here.

Overall, the projection models were based on the estimation models and were found to be appropriate and BSIA. However, given that the results of the estimation models were not considered BSIA (Section 7.3.1) and the uncertainty of the assessment results were likely underestimated (Section 7.3.3), the uncertainty of the projections were also likely to have been under-estimated. Therefore, the results of the projections would also not be considered BSIA, and should be interpreted with caution.

7.4.1 Recommendations on projections

Based on the above findings, there were no recommendations for improving the projections used for the PF-NH assessment, other than improving the estimation models and the associated uncertainties.

8. References

Cope, J. M., and O. S. Hamel. 2022. Upgrading from M version 0.2: An application-based method for practical estimation, evaluation and uncertainty characterization of natural mortality. Fisheries Research 256:106493.

Kell, L. T., A. Kimoto, and T. Kitakado. 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research 183:119–127.

Methot, R. D., and C. R. Wetzel. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142:86–99.

Pedersen, M. W., and C. W. Berg. 2017. A stochastic surplus production model in continuous time. Fish and Fisheries 18(2):226–243.

Stock, B. C., and T. J. Miller. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research 240:105967.

Tanaka, S. 1960. Suisan seibutsu no Population Dynamics to gyogyo shigen kanri (Population

Dynamics of fishery organisms and fishery resource management). Bulletin of Tokai Regional Fisheries Research Laboratory 28:1–200.
Then, A. Y., J. M. Hoenig, N. G. Hall, and D. A. Hewitt. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science 72(1):82–92.
Thorson, J. T. 2020a. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries 21(2):237–251.
Thorson, J. T. 2020b. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries 21(2):237–251.
Winker, H., S. E. Kerwath, and C. G. Attwood. 2014. Proof of concept for a novel procedure to standardize multispecies catch and effort data. Fisheries Research 155:149–159.
Xiao, Y., Y. Zhang, T. Yanagimoto, J. Li, Z. Xiao, T. Gao, S. Xu, and D. Ma. 2011. Population genetic structure of the point-head flounder, Cleisthenes herzensteini, in the Northwestern Pacific. Genetica 139(2):187–198.

9. Appendix 1 – Terms of Reference

Independent peer review for Japanese Fish Stock Assessment will be conducted by external experts based on the Stock Assessment Reports provided by Japan Fisheries Research and Education Agency (hereafter, FRA). External experts review the appropriateness of the assessment as mainly for points A to G below and provide recommendation and suggestions for future improvements. The FRA either reflects the recommendations to future stock assessment process or provides valid explanation if the suggested proposal is not applicable.

Points of review
A) Determine whether the data used for stock assessment are adequate to understand the stock dynamics of the target species and represent the best scientific information available.

B) Discuss whether the biological parameters used for stock assessment are appropriate.

C) Discuss whether the basic biological information such as distribution, migration pattern, and population are appropriate.

D) Evaluate whether the stock assessment methodology is based on the most appropriate available study and performed analytically.

E) Evaluate whether the data are treated statistically correctly.

F) Evaluate whether the stock assessment result obtained from the input data and methodology used is appropriate.

G) Evaluate the validity of methodology and result used for the future projection.

Review process
At first, the FRA will provide external experts with Stock Assessment Reports and supplemental materials as the subject under the review. After receiving pre-questions from external experts and confirming them by holding specific meeting, the external experts shall prepare and provide a final review report. In response to this, FRA will consider future responses and make a written document. These results shall be made public through the FRA web site.