



**SCIENTIFIC COMMITTEE
TWENTY-SECOND REGULAR SESSION**

11-19 August 2026
Apia, Samoa (Hybrid)

**NATURAL MORTALITY ESTIMATION FOR WCPO BIGEYE TUNA: A JOINT COHORT ANALYSIS OF
CORAL SEA AND REGION 4 TAGGING DATA**

WCPFC-SC22-2026-SA-IP14

11 July 2026

Submitted by

N. Ducharme-Barth¹, T. Peatman², J. Scutt Phillips³, C. Minte-Vera⁴

¹ NOAA Fisheries, Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Bldg. 176, Honolulu HI 96818

² Shearwater Analytics, Frome, England, BA11 6QD

³ Oceanic Fisheries Programme, Pacific Community, Noumea, New Caledonia

⁴ Inter-American Tropical Tuna Commission, La Jolla, California, USA

Table of contents

Executive summary	3
1 Introduction	4
2 Methods	4
2.1 Description of tagging data	4
2.2 Conversion of length to age	5
2.3 Cohort construction	5
2.4 Sex-ratio data	6
2.5 Cohort analysis model	7
2.5.1 Population dynamics	7
2.5.2 Natural mortality parametrization	7
2.5.3 Description of random effects	8
2.5.4 Description of penalties on reporting rates	8
2.5.5 Likelihood	9
2.5.6 Implementation	9
2.6 Sensitivities	10
2.6.1 Estimating the Lorenzen exponent	10
2.6.2 Fitting to sex-ratio data	10
2.6.3 Growth	10
2.7 Diagnostics and model selection	11
2.7.1 Estimability and the Hessian	11
2.7.2 Convergence and bounds	11
2.7.3 Fit to the observed releases	11
2.7.4 Reporting rates relative to the penalty	11
2.7.5 Model selection	11
2.8 Analytic approach	12
3 Results	13
3.1 Model progression and selection	13
3.2 Natural mortality at age	13
3.3 Release-event random effects	13
3.4 Fit to the observed releases	14
3.5 Estimated reporting rates	14
3.6 Sensitivities	14
3.6.1 Lorenzen exponent and sex structure	14
3.6.2 Growth	14
4 Discussion	15
4.1 Future work	17
5 Acknowledgements	18
6 Declaration of generative AI use	18
7 References	18
8 Tables	20
9 Figures	28

Executive summary

Natural mortality (M) is an influential parameter in tuna stock assessments and difficult to estimate directly. For bigeye tuna (*Thunnus obesus*) in the western and central Pacific Ocean (WCPO), the natural mortality schedules used in assessment have largely been derived from life-history relationships, meta-analytic priors, or values transferred from other stocks. Direct empirical estimates of M from WCPO tagging data can provide an independent basis for evaluating these assumptions. This paper estimates M at age using a reverse-time cohort analysis of tag release and recovery data.

The analysis pools three tagging programs archival tags from the Coral Sea Tagging Program (CSTP), and conventional tags from the Regional Tuna Tagging Programme (RTTP) and the Pacific Tuna Tagging Programme (PTTP, Region 4) within a common tag-cohort structure. The model is implemented in RTMB, with a shared Lorenzen mortality schedule, program-specific reporting rates and tag-loss inputs, and release-event random effects integrated out by the Laplace approximation. Models were built up in stages, from independent per-program fits through to a joint model with random effects and penalised estimation of reporting rates, with model selection guided by conditional AIC.

The joint random-effects model with penalised PTTP and RTTP reporting rates was selected as the final model based on conditional AIC. It yields a declining M-at-age schedule with baseline mortality at the reference length of $M_0 = 0.062 \pm 0.008$ per quarter, with M_a falling from 0.14 per quarter at age one year to approximately 0.06 per quarter by age ten. This estimate falls in the lower tail of the Hamel & Cope (2022) longevity-based prior and modestly below the 2023 diagnostic-case value, though its upper confidence bound approaches the latter. This broadly corroborates the previously assumed order of magnitude and provides an indication that the tagging data in isolation support somewhat lower M than previously assumed.

The key caveat of the analysis revolves around the confounding between reporting rate and M . Freeing the PTTP reporting rate under penalty moved estimated M_0 from 0.158 to 0.068 per quarter, and the freed rates settled well below their penalty centres. The release-recovery data alone cannot distinguish whether this reflects genuine information about reporting rate or the absorption of other unmodelled attrition (e.g., spatially variable reporting, emigration from the fished area, misspecified tag loss, or violation of the terminal condition). Comparable confounding exists within the Multifan-CL integrated assessment and the present analysis does not resolve the confounding but isolates and quantifies it.

Structural extensions estimating the Lorenzen exponent or adding sex structure were examined as sensitivities, along with models that used alternative growth curves to define the age of tagged individuals. Priorities for further work are finer stratification of reporting rates, tighter reporting-rate priors from continued tag-seeding, simulation testing against a known M schedule, and a Bayesian treatment of the same model.

1 Introduction

Natural mortality (M) is an influential parameter in tuna stock assessments and difficult to estimate directly. For bigeye tuna (*Thunnus obesus*) in the western and central Pacific Ocean (WCPO), the natural mortality schedules used in assessment have largely been derived from life-history relationships, meta-analytic priors, or values transferred from other stocks. Direct empirical estimates of M from WCPO tagging data can provide an independent basis for evaluating these assumptions.

This paper estimates natural mortality at age for WCPO bigeye tuna using a cohort analysis of tag release and recovery data. The cohort analysis treats a group of fish tagged and released together as a tag-cohort whose abundance is tracked through subsequent recoveries. Since removals by fishing are observed as tag recoveries, the residual attrition of the cohort is attributable to natural mortality and to processes that act like it (tag shedding, tag-induced mortality, non-reporting). The approach follows the framework of Hearn et al. (1987), Maunder et al. (2009) and the implementation of Minte-Vera et al. (2025), developed for yellowfin tuna in the eastern Pacific Ocean and adapted here for WCPO bigeye tuna.

The analysis is deliberately external to the integrated stock assessment. Estimating M inside the assessment model confounds it with the simultaneous estimation of recruitment, selectivity, movement, and growth; isolating the tagging data in a dedicated cohort analysis avoids that confounding and makes the information content of the tagging data, and its limitations, explicit. The intent is to produce an estimate of M at age, with a defensible characterisation of uncertainty, that can inform the natural mortality assumptions of the next WCPO bigeye tuna assessment.

Three tagging datasets are used: archival (electronic) tags from the Coral Sea Tagging Program (CSTP; Evans et al. (2008)), conventional tags from the Regional Tuna Tagging Programme in the Coral Sea (RTTP; Hampton & Gunn (1998)), and conventional tags from the Pacific Tuna Tagging Programme in WCPFC Region 4 (PTTP). The analysis proceeds from fits to each program in isolation, through a joint model that pools all three, to a joint model with release-event random effects and penalised estimation of program-specific reporting rates. Model selection is guided by conditional AIC (Zheng et al., 2024), and an analysis of posterior predictive checks. Sensitivity to the Lorenzen exponent, to the inclusion of sex-ratio data, and to the assumed growth curve used to convert lengths to age is examined.

2 Methods

2.1 Description of tagging data

Three tag release-recovery datasets contribute to the analysis. Each is a separate tagging program with its own gear composition, reporting environment, and tag technology, and each is processed into a common tag-cohort structure (Section 2.3) before being combined.

The Coral Sea Tagging Program (CSTP) comprises archival-tagged bigeye tuna released in the western Coral Sea (Evans et al., 2008). Archival tags are recovered through the fishery and returned by fishers; the recovery process is treated as a single gear category. The released count is taken directly without a usability correction as tags did not lack the required information: release length, release date, or recovery date.

The Regional Tuna Tagging Programme (RTTP) comprises conventional dart-tagged bigeye tuna released in

the Coral Sea (Hampton & Gunn, 1998). Recoveries are disaggregated by recovery gear into longline (L), pole-and-line (P), purse-seine (S), and handline (H). The released count is corrected for the fraction of tags judged not usable for the analysis (e.g., those lacking a usable release length, a usable release date, or a recovery date of sufficient reliability) following the data quality criteria of Berger et al. (2014), and Vincent et al. (2019).

The Pacific Tuna Tagging Programme (PTTP) comprises conventional dart-tagged bigeye tuna released in WCPFC Region 4. PTTP recoveries used here are purse-seine (S) recoveries. As with RTTP, the released count is corrected for usability: recoveries are retained only when the recovery date carries a reliability classification of day, week, fortnight, or month, following the data quality criteria of Berger et al. (2014) and Vincent et al. (2019).

For all three programs the analysis uses only tags with a usable release length, a usable release date, and a usable recovery date if the tag was recovered, since each of these is required to assign the tag to a release-age cohort and to compute time at liberty. The distributions of release length and the resulting release age under each growth scenario are shown in Figure 1. The empirical recovery-age distribution across programs and gears is shown in Figure 2. Tag release and recapture counts by program, including the numbers excluded at each processing step, are summarised in Table 1.

2.2 Conversion of length to age

Age is the time axis of the cohort analysis, but the tagging data record fork length at release, not age. Following Minte-Vera et al. (2025), age at release is obtained by inverting an external growth curve at each tagged fish’s measured release length, and age at recovery is age at release plus time at liberty (recovery date minus release date, in quarters).

The growth curve is a Schnute-Richards parameterisation for WCPO bigeye tuna, with age in quarters:

$$L(a) = L_1 + (L_2 - L_1) \frac{1 - e^{-k(a-a_1)}}{1 - e^{-k(a_2-a_1)}},$$

with reference ages $a_1 = 1$ quarter and $a_2 = 40$ quarters. The baseline analysis uses the growth parameters estimated internally from the 2023 assessment diagnostic case ($k = 0.1128$, $L_1 = 23.33$ cm, $L_2 = 148.75$ cm). Two alternative parameterisations, the 2020 assessment diagnostic case and the curve of Eveson et al. (2020), are carried through as a sensitivity (Section 2.6.3).

To assign age at release, the growth function is numerically inverted for each measured fork length and the resulting continuous age is rounded to the nearest integer quarter. Lengths below L_1 are assigned the minimum reference age and lengths above L_2 the maximum. Inside the model, length at age is required for the allometric mortality scaling (Section 2.5.2); it is evaluated at the midpoint of each quarterly age class, $L(a + 0.5)$, and the reference length for the scaling is $L_{\text{ref}} = L(40.5)$.

2.3 Cohort construction

A tag-cohort is the unit of the analysis. Tags from each program are grouped by the combination of release date and integer age at release: every distinct (release date $\times$ age-at-release) pair defines one cohort. Defining cohorts at this granularity, rather than at the level of a release time-step, keeps a single tagging event that

released fish of several ages as several cohorts that nonetheless share that event, which is the structure required to attach a release-event random effect (Section 2.5.3).

For each cohort the data record the number released and a recovery array indexed by recovery age and recovery gear. Recoveries are kept disaggregated by gear so that a gear-specific reporting rate can be applied to each recovery pathway independently. When the three programs are combined, gear labels are made globally unique by prefixing them with the program name, producing the gear union { `CSTP_archival` , `PTTP_S` , `RTTP_L` , `RTTP_P` , `RTTP_S` , `RTTP_H` }. Each program’s recoveries are routed to its own columns of the stacked recovery array, so a single shared mortality schedule can be estimated while every program-gear pathway retains its own reporting rate.

Two further per-cohort indices are constructed. The first maps each cohort to a release-event group: release-event identifiers are prefixed by program so that release dates coinciding across programs remain distinct, and cohorts sharing a prefixed identifier share a group. The second maps each cohort to a tag type (CSTP archival, RTTP conventional, PTTP conventional), which selects the continuous tag-loss rate η and the initial tag-loss probability τ applied to that cohort. Both loss parameters are treated as known inputs, pre-combined across loss processes in the data-preparation step, and supplied to the model by tag type.

The initial tag-loss probability combines shedding and post-release mortality as $\tau = 1 - (1 - s)(1 - m)$, where s is the initial shedding fraction and m is the post-release tag-induced mortality probability. For CSTP archival tags, both continuous shedding and tag-induced mortality are assumed negligible, giving $\eta = 0$ and $\tau = 0$ (Minte-Vera et al., 2025). For RTTP conventional tags, the continuous shedding rate is $\eta = 0.00694$ per quarter (the monthly pooled rate of Hampton (1997) multiplied by three for the quarterly time step), and the initial tag-loss probability is $\tau = 1 - (1 - 0.0586)(1 - 0.07) \approx 0.125$, where the initial shedding fraction ($s = 0.0586$) is from Hampton (1997) and the post-release mortality ($m = 0.07$) is from Berger et al. (2014). For PTTP conventional tags, continuous shedding is negligible ($\eta = 0$) and the initial tag-loss probability is $\tau = 1 - (1 - 0.0697)(1 - 0.07) \approx 0.135$, where the initial shedding fraction ($s = 0.0697$) is from Berger et al. (2014), Vincent et al. (2019), and T. Peatman (pers. comm.), and the post-release mortality ($m = 0.07$) is from Berger et al. (2014) and Vincent et al. (2019).

2.4 Sex-ratio data

The tagging data are sex-combined as the sex of a tagged fish is generally not known. The natural mortality used in the cohort dynamics is therefore a numbers-weighted average of male and female mortality at age (Section 2.5.2). To inform any difference between male and female mortality, the model can additionally be fitted to data on the proportion of females at age.

The sex-ratio observations are derived from WCPO longline observer sampling. Following Minte-Vera et al. (2025), the observations supplied to the model are smoothed proportions. A four-parameter logistic curve is fitted to the sex-at-length data and propagated through an age-length key to give a proportion-female-at-age series, rather than raw empirical proportions. The smoothed series is treated as a best estimate of the true proportion female at age, consistent with the normal, constant-variance likelihood used for this component (Section 2.5.5). Before the sex-ratio series is joined to the model it is filtered to the age range covered by the tagging cohorts, so that every sex-ratio observation maps to a modeled age class.

Whether the sex-ratio component is included at all is treated as a structural choice and examined as a

sensitivity (Section 2.6.2). The baseline model is sex-combined and does not use the sex-ratio data.

2.5 Cohort analysis model

The model adapts the cohort analysis of Minte-Vera et al. (2025). It reconstructs, for each tag-cohort, the abundance of tagged fish that is consistent with the observed recoveries, and estimates the natural mortality schedule that best reconciles the reconstructed releases with the observed releases across all cohorts. It is implemented in R using RTMB (Kristensen, 2023), with fixed effects estimated by `nlminb` and release-event random effects integrated out by the Laplace approximation.

2.5.1 Population dynamics

The abundance of tagged fish in a cohort is projected in reverse-time, from the oldest modeled age class back to the age at release. Reverse-time projection is used because, with recoveries treated as known removals, it keeps reconstructed abundance positive without constraint. Recoveries are assumed to be removed at the midpoint of each quarterly age class and releases to occur at the start of the release age class. For a cohort at age class a ,

$$N_a = \frac{\frac{N_{a+1}}{e^{-0.5M_a - 0.5\eta}} + \sum_g \frac{C_{g,a}}{\lambda_g}}{e^{-0.5M_a - 0.5\eta}},$$

where N_{a+1} is the reconstructed abundance at the next age class (zero beyond the last modeled age), $C_{g,a}$ is the number of recoveries by gear g at age a , λ_g is the reporting rate for gear g , M_a is natural mortality at age a , and η is the continuous tag-loss rate for the cohort's tag type. The reporting rate enters on the prediction side of the recursion, scaling observed recoveries up to the recoveries implied by the tagged population; this placement follows Equation 6.2 of Minte-Vera et al. (2025).

The predicted number of releases for the cohort is the reconstructed abundance at the release age class, corrected for the initial tag-loss probability τ of the cohort's tag type:

$$\hat{R} = \frac{N_{a_{\text{rel}}}}{1 - \tau}.$$

The continuous and initial tag-loss parameters η and τ are supplied by tag type as known inputs.

2.5.2 Natural mortality parametrization

Natural mortality at age is parameterised with a Lorenzen (Lorenzen, 2022) form, in which M scales allometrically with length at age:

$$M_{a,\text{males}} = M_0 \left(\frac{L_a}{L_{\text{ref}}} \right)^{\omega_L},$$

where M_0 is the mortality rate at the reference length $L_{\text{ref}} = L(40.5)$, $L_a = L(a + 0.5)$ is mid-quarter length at age, and ω_L is the Lorenzen exponent. In the baseline model ω_L is fixed at -1 . Estimating ω_L as a free parameter (constrained negative), as in Minte-Vera et al. (2025), is examined as a sensitivity (Section 2.6.1). M_0 is estimated on the log scale.

Because the tagging data are sex-combined, the mortality used in the cohort dynamics is a numbers-weighted

average of male and female mortality. When the sex-structured configuration is active, female mortality carries a multiplicative logistic offset relative to male mortality,

$$M_{a,\text{females}} = M_{a,\text{males}} \left[\frac{\Delta_{\max}}{1 + e^{-\ln(19) (L_a - L_{50}) / (L_{95} - L_{50})}} + 1 \right],$$

with $\Delta_{\max}$ the maximum female offset and L_{50} , L_{95} the lengths at which 50% and 95% of the offset is realised. Starting from a 1:1 sex ratio at the youngest modeled age class and applying sex-specific survival, the model projects the proportion female at age, $p_{\text{female},a}$, and forms the mixed-sex schedule used in the dynamics as

$$M_a = \frac{N_{\text{males},a} M_{a,\text{males}} + N_{\text{females},a} M_{a,\text{females}}}{N_{\text{males},a} + N_{\text{females},a}}.$$

In the baseline sex-combined model the female offset is inactive, male and female mortality are identical, and M_a reduces to the allometric schedule. The mortality schedule is additionally evaluated over a fixed full reporting age range (quarters 1-40) for presentation, independent of the per-cohort age range used in the likelihood.

2.5.3 Description of random effects

Tagging events differ in handling conditions, post-release predation environment, and other factors that affect the realised mortality experienced by tagged fish in ways not captured by a single shared schedule. To absorb this between-event variation, the model can attach a multiplicative random effect to the mortality experienced by each release-event group. For release-event group r , the effective mortality at age is $M_a \exp(\varepsilon_r)$, with

$$\varepsilon_r \sim \text{Normal}\left(-\frac{1}{2}\sigma_\varepsilon^2, \sigma_\varepsilon^2\right).$$

The mean of $-\frac{1}{2}\sigma_\varepsilon^2$ is a bias correction: it makes $E[\exp(\varepsilon_r)] = 1$, so that the population-mean schedule M_a retains its interpretation as the mean over release events rather than being rescaled by the random-effect variance. The deviates ε_r are estimated as random effects and integrated out by the Laplace approximation; their standard deviation σ_ε is estimated as a free parameter on the log scale. In the fixed-effects configuration the random-effect term is deactivated and every $\exp(\varepsilon_r)$ is held at 1.

2.5.4 Description of penalties on reporting rates

Each gear's reporting rate λ_g enters the dynamics on the logit scale as $\text{logit}(\lambda_g)$. A reporting rate may be (i) held fixed at a value taken from program documentation or tag-seeding studies, or (ii) estimated subject to a penalty. The penalty is a normal density on the logit scale,

$$-\ln L_\lambda = - \sum_{g \in \mathcal{P}} \ln \text{Normal}(\text{logit}(\lambda_g) \mid \mu_g, \sigma_g),$$

summed over the set $\mathcal{P}$ of gears with a penalty active; gears without a penalty (those held fixed) do not contribute. The penalty mean μ_g and standard deviation σ_g are derived from an external prior specified on the natural scale (e.g., a central reporting rate and a 95% interval) transformed to the logit scale. This

penalised formulation is what allows a reporting rate to be estimated while still being anchored to external evidence; the penalty parameters μ_g and σ_g are supplied as data.

The prior central values by program are as follows. For the CSTP archival gear, the reporting rate is fixed at $\lambda = 0.95$ and is not estimated (Minte-Vera et al., 2025). For RTTP gears, the prior is centred on $\lambda = 0.58$ for the purse-seine (S) and handline (H) gears and $\lambda = 0.75$ for the longline (L) and pole-and-line (P) gears (Hampton, 1997; Hampton & Gunn, 1998); all four gears are estimated subject to a penalty in the joint models, with the purse-seine and handline rates tied to a single shared value. For the PTTP purse-seine gear, the prior is centred on $\lambda = 0.53$ (T. Peatman, pers. comm.) and is estimated subject to a penalty in the joint model.

2.5.5 Likelihood

Parameters are estimated by minimising a total negative log-likelihood with up to five additive components.

The release component is a lognormal likelihood comparing observed and predicted releases across cohorts. With σ_R the release standard deviation (estimated on the log scale), and a bias-corrected mean so that $\hat{R}_i$ is the arithmetic-scale mean,

$$-\ln L_R = -\sum_{i=1}^n \ln \text{Normal}(\ln R_i \mid \ln \hat{R}_i - \frac{1}{2}\sigma_R^2, \sigma_R).$$

The sex-ratio component, active only in the sex-structured configuration, is a normal likelihood on the proportion-female-at-age scale with constant standard deviation σ_p (estimated on the log scale),

$$-\ln L_p = -\sum_a \ln \text{Normal}(p_a \mid \hat{p}_a, \sigma_p).$$

The random-effect component is the process density of the release-event deviates, $-\sum_r \ln \text{Normal}(\varepsilon_r \mid -\frac{1}{2}\sigma_\varepsilon^2, \sigma_\varepsilon^2)$, active only when the random-effect branch is on. The reporting-rate component is the logit-scale penalty $-\ln L_\lambda$ of Section 2.5.4, active only when one or more reporting rates are penalised. A fifth component is available for optional normal priors on the fixed effects.

The total objective is the sum of the active components:

$$-\ln L = -\ln L_R - \ln L_p - \ln L_\varepsilon - \ln L_\lambda - \ln L_{\text{priors}}.$$

For the fixed-effects models this reduces to the release component alone; for the joint random-effects models with penalised reporting rates it is the release, random-effect, and reporting-rate components.

2.5.6 Implementation

The model is implemented in RTMB (Kristensen, 2023). The objective function is built once and reused across configurations: structural switches in the data list activate or deactivate the sex-ratio, random-effect, and reporting-rate components, and the `map` argument of `MakeADFun` fixes parameters or ties them together (e.g., tying the RTTP purse-seine and hand-line reporting rates to a single estimated value). All rate and scale

parameters are estimated on transformed scales (e.g., log for mortality, standard deviations, and tag-loss-free quantities; logit for reporting rates) so the optimiser works in an unconstrained space, with box constraints retained as a safeguard.

Fixed effects are estimated with `nlminb`, run twice from the first solution to confirm the optimum is stable. Release-event random effects are declared through the `random` argument and integrated out by the Laplace approximation. Standard errors for derived quantities, including natural mortality at age, are obtained by the delta method via `sdreport`. After `sdreport`, the automatic-differentiation object’s cached evaluation point is re-anchored at the optimum before any further gradient or report evaluation, because the Hessian computation inside `sdreport` leaves it perturbed. The codebase is modular (e.g., separate files for the model function, likelihood components, growth and utility functions, and diagnostics) with the fitting workflow for each model orchestrated by a dedicated Quarto vignette. The implementation departs from the reference implementation of Mente-Vera et al. (2025) only in using log- and logit-scale parameterisation with `nlminb` and in its modular file structure.

2.6 Sensitivities

Three structural choices in the cohort analysis were examined by re-fitting the model under alternatives and comparing the resulting mortality estimates, parameter estimability, and model-selection scores. All three sensitivities were run from the joint RE+PTTP+RTTP configuration, which is the basis for reported inference (Section 2.8).

2.6.1 Estimating the Lorenzen exponent

The baseline model fixes the Lorenzen exponent ω_L at -1 . The reference implementation of Mente-Vera et al. (2025) estimates ω_L as a free parameter, constrained negative. Two model configurations relax the fixed value (one estimating ω_L with the baseline sex-combined structure, one estimating ω_L alongside the sex-structured extension) and are compared with their fixed- ω_L counterparts on conditional AIC, estimability diagnostics, and posterior predictive checks.

2.6.2 Fitting to sex-ratio data

The baseline model is sex-combined and does not use the sex-ratio observations. The sex-structured configuration (Section 2.5.2) activates the female mortality offset and the sex-ratio likelihood component, adding the four parameters σ_p , $\Delta_{\max}$, L_{50} , and L_{95} . Two configurations were fitted (sex structure with ω_L fixed, and sex structure with ω_L estimated) and compared with the corresponding sex-combined models. The comparison asks both whether the additional structure is supported by conditional AIC and whether the additional parameters are estimable.

2.6.3 Growth

Length-to-age conversion enters the analysis at two points: in data preparation, where the inverted growth curve assigns each tagged fish to a release-age cohort, and inside the model, where mid-quarter length at age scales the mortality schedule. The growth sensitivity holds these two uses consistent within each scenario and re-fits the joint RE+PTTP+RTTP model under three Schnute-Richards parameterisations: the 2023 assessment diagnostic case (the baseline), the 2020 assessment diagnostic case, and the curve of Eveson et

al. (2020). Reference ages are $a_1 = 1$ and $a_2 = 40$ quarters in every scenario. Because the inverted growth curve maps the same observed lengths to different age bins, the per-scenario cohort age range differs; the full reporting age range used for presentation is held fixed at quarters 1-40.

2.7 Diagnostics and model selection

Each fitted model was evaluated with a common set of diagnostics, and competing models were compared using conditional AIC.

2.7.1 Estimability and the Hessian

Parameter estimability was checked by eigendecomposition of the Hessian of the fixed effects: a Hessian eigenvalue at or below the square root of machine precision flags a likelihood ridge along which some linear combination of parameters is not separately identified. The check reports which parameters load onto any near-singular eigenvector. A model whose Hessian is not positive-definite has at least one non-estimable direction, and its delta-method standard errors are not reliable. Alongside the estimability check, the fixed-effect correlation matrix (obtained by inverting the Hessian) was inspected for parameter pairs with absolute correlation above 0.95, since strong pairwise correlation, even short of a singular Hessian, indicates parameters the data cannot separate well.

2.7.2 Convergence and bounds

For each model the maximum absolute gradient component at the reported solution was recorded as a convergence diagnostic, and each estimated parameter was checked against its box constraints. A parameter resting at a bound indicates that the constraint, rather than the data, is determining its value and that the corresponding estimate should be interpreted cautiously.

2.7.3 Fit to the observed releases

Goodness of fit was assessed against the release observations. For each model, datasets were simulated from the fitted model and used to form, per cohort, a probability-integral-transform (PIT) value. A well-calibrated model produces PIT values that are uniform on the unit interval. PIT values were examined through the empirical CDF of the PIT values grouped by tagging program, which exposes program-specific over- or under-prediction, and the predicted releases were plotted against the observed releases with simulation envelopes.

2.7.4 Reporting rates relative to the penalty

For models that estimate one or more reporting rates, the estimated rate was compared against the penalty that anchors it, on both the logit and natural scales, to show whether the data move the rate away from its externally specified centre and by how much relative to the penalty width.

2.7.5 Model selection

Model selection was guided by conditional AIC (cAIC; Zheng et al., 2024) and scrutiny of PIT empirical CDFs plots. For a mixed-effects model, marginal AIC scores the model only based on the fixed effects while conditional AIC takes the random effects into account. Conditional AIC is the appropriate criterion

for comparisons between models using random effects. cAIC is computed from the conditional negative log-likelihood and the effective degrees of freedom (EDF) contributed by the random-effect block; EDF substantially below the number of release-event groups indicates that the data support fewer effective degrees of release-event variation than the nominal random-effect count. Conditional AIC is not defined for fixed-effects-only models, which have no random-effect block; those models are reported with cAIC absent and compared on marginal AIC.

2.8 Analytic approach

The analysis was built up in stages, with each stage motivated by the diagnostics of the previous one.

The first stage fitted the cohort analysis to each tagging program in isolation CSTP, then RTTP, then PTTP each in a fixed-effects and a release-event random-effects configuration. Random-effects were trialled to see if they improved fit by absorbing spatiotemporal variability in processes impacting the observed releases (e.g., reporting, emmigration, etc.). The independent fits establish what each program’s data can support on their own and provide a baseline against which the pooled models can be judged.

The second stage pooled all three programs into a single joint model with a shared mortality schedule, program-specific reporting rates, and program-specific tag-loss parameters, fitted first as a fixed-effects model and then with release-event random effects. Conditional AIC was used to judge whether the random-effect structure was supported.

The diagnostics of the joint random-effects model identified a specific problem with the PTTP component. Given the assumed PTTP reporting rate and the observed ratio of PTTP recoveries to PTTP releases, natural mortality for the PTTP cohorts was estimated to be very large, and the PIT empirical CDF for the PTTP program indicated a poor fit. This motivated the third stage: estimating the PTTP reporting rate subject to a penalty anchored to external evidence, while holding the CSTP and RTTP reporting rates fixed.

The fourth stage extended penalised estimation to the RTTP reporting rates as well, estimating the RTTP longline and pole-and-line rates under a broad penalty and the tied RTTP purse-seine and hand-line rate under an informative penalty, giving the joint RE+PTTP+RTTP model. Conditional AIC across the four joint models (fixed effects; random effects; random effects with the PTTP reporting rate estimated; random effects with the PTTP and RTTP reporting rates estimated) guided the selection of the model used for reported inference.

A final set of sensitivities (Section 2.6) examined whether the joint RE+PTTP+RTTP model could be extended to estimate the Lorenzen exponent or to incorporate sex structure, and how sensitive the mortality estimate is to the assumed growth curve.

3 Results

3.1 Model progression and selection

The four joint models (fixed effects, release-event random effects, random effects with the PTTP reporting rate estimated, and random effects with the PTTP and RTTP reporting rates estimated) were compared on conditional AIC (Table 2), with the independent per-program fits and the structural extensions providing context. The fit and likelihood summaries, including parameter counts, negative log-likelihood by component, marginal and conditional AIC, and effective degrees of freedom, are given in Table 3. The diagnostics and estimability summaries, including the maximum gradient component, the positive-definite-Hessian status, the count of non-estimable parameters, the count of strongly correlated parameter pairs, and the bounds status, are given in Table 4.

Among the joint models for which conditional AIC is defined, the lowest cAIC is the **Joint RE+PTTP+RTTP** model. The release-event random-effect block in that model carries an effective degrees of freedom of 25.3 against 240 release-event groups, indicating that the data support substantially fewer effective degrees of release-event variation than the nominal random-effect count. The Joint RE+PTTP+RTTP model is taken as the basis for reported inference; the natural mortality estimates and uncertainty that follow are from that model. The estimated M_0 and Lorenzen exponent ω_L for all fitted models, with delta-method standard errors, are given in Table 5.

The progression from the fixed-effects joint model through the random-effects and penalised-reporting-rate models is reflected in the per-component negative log-likelihoods and the information criteria of Table 3. The fixed-effects joint model has only the release likelihood component active; activating the release-event random effects adds the random-effect process density; estimating the PTTP and then the RTTP reporting rates adds the reporting-rate penalty component and changes the release component as the reconstructed releases respond to the revised reporting rates.

3.2 Natural mortality at age

The natural mortality schedules estimated from the three programs in isolation are shown in Figure 3. The per-program estimates reflect the information each dataset carries independently before pooling into the joint model. The natural mortality schedule estimated by the selected joint RE+PTTP+RTTP model is shown in Figure 4, extrapolated over the full reporting age range, and the schedules from all four joint models are compared in Figure 5. The Lorenzen form gives a mortality rate that declines with age as length increases. The comparison across the four joint models shows the effect of the penalised reporting-rate estimation on the inferred schedule: revising the PTTP and RTTP reporting rates downward, where the data and penalties support it, reduces the apparent attrition of those programs' cohorts and therefore the inferred mortality. The full schedule from the selected model at annual ages is tabulated with delta-method uncertainty in Table 6. The estimated M_0 is shown relative to the Hamel & Cope (2022) longevity-based prior and the corresponding 2023 assessment value in Figure 6.

3.3 Release-event random effects

The release-event mortality deviates from the selected model are shown in Figure 7. They quantify the between-event variation in realised mortality that the shared schedule does not capture; the spread of the

deviates and the effective degrees of freedom reported in Table 3 together indicate how much release-event-level variation the data actually support.

3.4 Fit to the observed releases

The fit to the observed releases for the selected model is summarised by the program-grouped PIT empirical CDF (Figure 8) and by the predicted-against-observed release plot (Figure 9). The evolution of the program-grouped PIT empirical CDF across the four joint models is shown in Figure 10 through Figure 13. The program-grouped PIT curves are the diagnostic that motivated the penalised estimation of the PTPP, and then the RTPP, reporting rates: under fixed reporting rates the PTPP curve departed markedly from the diagonal, and estimating the reporting rate under a penalty was the response to that departure. Estimating the PTPP reporting rate improved the PTPP PIT curve, and all programs appear to be satisfactorily fitted at this stage.

3.5 Estimated reporting rates

The gear-specific reporting rates of the selected model are given in Table 7, and the estimated reporting rates are shown against their penalties in Figure 14. The comparison shows, for each estimated rate, whether the tagging data move it away from the externally specified penalty centre and by how much relative to the penalty width.

3.6 Sensitivities

3.6.1 Lorenzen exponent and sex structure

Neither estimating the Lorenzen exponent nor adding sex structure simultaneously improved on the baseline joint RE+PTTP+RTPP model while also satisfying estimability criteria and producing satisfactory PIT empirical CDF diagnostics. The baseline extension model had a conditional AIC of 1758.2 with the +omega_L (cAIC 1753.9) configuration scoring lower. Although the +omega_L model had a lower cAIC, the grouped PIT empirical CDF shown in Figure 15 indicated a higher degree of misfit than the baseline model. Both the +Sex, and +Sex+omega_L configurations failed the estimability checks. The configurations that add sex structure carry non-estimable parameters (Table 8), indicating the tagging and sex-ratio data cannot separately identify the additional parameters. The baseline joint RE+PTTP+RTPP model is therefore retained for reported inference, with the Lorenzen exponent fixed at -1 and the sex-ratio data not used. For reference, the estimated mortality schedules from these sensitivity models are shown in Figure 16.

3.6.2 Growth

The sensitivity of the estimated mortality schedule to the assumed growth curve is shown in Figure 17. Re-fitting the joint RE+PTTP+RTPP model under the 2020 diagnostic-case growth curve and the curve of Eveson et al. (2020), in place of the 2023 diagnostic-case baseline, did not appear to be very influential: the estimated schedules under the three growth assumptions are broadly similar over the modeled age range.

4 Discussion

This analysis estimated natural mortality at age for WCPO bigeye tuna from a reverse-time cohort analysis that pooled three tagging programs, built up in stages from independent per-program fits through to a joint model with release-event random effects and penalised estimation of program-specific reporting rates. The primary motivation for conducting the estimation outside the integrated stock assessment was to isolate the information that the tagging data carry about M , and to make the limitations of that information explicit, rather than have them absorbed into the simultaneous estimation of recruitment, selectivity, movement, and growth within an integrated assessment model.

In the current analysis, model-selection was driven by the diagnostics at each stage rather. The release-event random effects were retained because conditional AIC favoured them over the fixed-effects joint model. The penalised estimation of the PTTP reporting rate was not part of the original design; rather, it emerged as the response to a specific and diagnosable failure of the fixed-reporting-rate joint model, in which natural mortality for the PTTP cohorts was estimated to be implausibly high given the assumed reporting rate and the observed ratio of recoveries to releases, and the PTTP program-grouped PIT curve departed markedly from the diagonal (Figure 10 and Figure 11). Extending the same penalised treatment to the RTTP reporting rates followed naturally for the same reason.

The aforementioned reporting rate sensitivity, and the ability for it to ‘rescue’ the fit is a direct consequence of how reporting rates enter the model and form the basis for a key caveat of the current analysis. Given that reconstructed removals scale as C/λ , a lower reporting rate implies that more fish were removed by fishing for each observed recovery, so that a greater share of a cohort’s attrition is attributed to fishing and correspondingly less to natural mortality. Reporting rate and M are negatively confounded, and the magnitude of this confounding in the present analysis is substantial. Freeing the single PTTP reporting rate under a penalty moved the estimated baseline mortality M_0 from 0.158 to 0.068 per quarter (Table 5), and the reporting rates for the two recovery groups settled well below their penalty centres¹. While this shift could be interpreted as evidence that the release data are informative about reporting rate, it could also indicate that the reporting rate is absorbing other, unmodelled attrition (e.g. spatially variable reporting, emigration, mis-specified tag loss, or violation of the terminal condition that no tagged fish are alive after the last recovery) that would otherwise be expressed as inflated M . Individually, the release–recovery data alone cannot distinguish between these interpretations, because each of them acts on M through the same $1/\lambda$ process. However, pooling the data across the three programs², including the CSTP archival program which assumes a fixed high reporting rate, may help constrain (along with the reporting rate penalties and the fixed Lorenzen exponent) the M estimate sufficiently to simultaneously estimate M as well as the PTTP and RTTP reporting rates. Indeed, the estimability diagnostics confirm that these parameters are jointly identified in the fitted model, which has a positive-definite Hessian, no non-estimable parameters, and no fixed-effect pair with absolute correlation above 0.95 (Table 4).

This confounding between reporting rate and M is not an artefact of estimating M externally. Comparable

¹Under the selected model the estimated PTTP purse-seine reporting rate was 0.23 against a penalty centre of 0.53, and the RTTP longline rate 0.35 against a centre of 0.75 (Table 7; Figure 14). The RTTP pole-and-line and the tied purse-seine/handline rates, by contrast, remained closer to their penalty centres.

²Pooling the data across multiple programs assumes that the natural mortality schedule experienced by fish tagged in each program is the same. Given that the programs operated in different areas and at different times, this assumption is not guaranteed to hold since natural mortality is likely to be non-stationary in space and time as a function of underlying environmental and oceanographic conditions.

behaviour exists within the Multifan-CL integrated assessment, which typically requires allowing reporting rates considerable freedom and applying event-level adjustments to the effective number of releases in order to achieve plausible fits to the tagging data. The current analysis does not resolve this confounding but helps to highlight it. A mutually consistent M -at-age schedule across the three programs is obtained only once the reporting rates are freed under penalty and release-event variation is admitted.

Relative to the natural mortality estimated in the 2023 assessment, the estimate presented here serves as a useful reality check. The estimated M_0 at the reference length falls in the lower tail of the Hamel & Cope (2022) longevity-based prior and modestly below the 2023 diagnostic-case value at the reference age, though its upper 90% confidence bound approaches that value (Figure 6). This is interpreted as broad corroboration of the order of magnitude previously estimated (Day et al., 2023), and an indication that the tagging data in isolation support somewhat lower natural mortality than previously assumed.

One of the challenges associated with the Multifan-CL integrated assessment is how to appropriately deal with tag mixing, particularly when fishing mortality is spatially and temporally varying. One of the advantages of the current analysis is that it is conditioned on the observed removals rather than having to predict them from a fishing-mortality parameter. In doing so, this recapture conditioning implicitly integrates over any heterogeneity in fishing mortality that a cohort may have encountered, and side-steps the tag mixing issue. However, this does not mean that the analysis is immune to contamination from spatiotemporal heterogeneity, and estimates of M may still be confounded in several ways. As mentioned previously, the reporting rate assumption is critical. The reconstruction of released tags is correct only insofar as λ accurately converts recoveries into removals for a cohort's actual catch composition. In the current analysis λ is only stratified by gear and program. However, the true probability that a recovered tag is reported may vary by space and time as a function of the capture vessel flag, sub-fleet, or post-capture pathway (e.g., transhipped or landed locally). A fish that is caught but not reported is neither recovered nor reconstructed as a removal, and is attributed to natural mortality. Emigration can also contaminate M estimates, particularly if tags released in certain areas are more prone to leave the fished area. The model contains no term for permanent movement out of the effectively fished and reported area, so within the model a tagged fish that leaves is indistinguishable from one that dies due to natural mortality. In this way M may more accurately be described as the total non-fishing attrition of tagged fish than purely as natural mortality. Lastly, and related to emigration, the reverse-time cohort analysis assumes that no fish are alive after the last recovery. However, some fish that move into low-effort or low-reporting areas could survive longer than the observed tagged fish. In this way, spatially varying fishing mortality, combined with a finite recapture window, may violate that assumption and bias estimates of M .

Several further limitations also exist with this analysis. The continuous and initial tag-loss parameters are supplied as known inputs rather than estimated. The reporting-rate penalties, though they permit the data to inform the rates, nonetheless depend on externally specified centres and widths; the centres adopted here are provisional, and should be revisited against published tag-seeding values as these become available. Additionally, the analysis depends on the growth curve used to convert length to age, though in this case it did not appear overly sensitive. Future work should look to formally account for the probability that an individual of a certain length is of a given age by integrating over the uncertainty in the growth curve and the probability that an individual of a given age exists within the population.

Lastly, the structural extensions we examined were not strongly supported by the data. Estimating the

Lorenzen exponent, as the reference implementation of Mente-Vera et al. (2025) does, and adding sex structure informed by the longline sex-ratio data, were trialled and either showed issues with estimability (Table 8) or model fit evaluated using the PIT empirical CDF. The sex-structured configurations carried non-estimable parameters. We take this to indicate that the tagging data assembled here, while sufficient to estimate a baseline level of M with a fixed Lorenzen exponent, do not contain enough information to separately identify a sex-specific mortality offset. Estimation of the Lorenzen exponent alone did appear technically possible though introduced some degradation to the fit according to the PIT empirical CDF. Freely estimating this exponent also drove it to approximately -2 which is anomalously lower than commonly reported values in the literature (Lorenzen et al., 2022; Lorenzen, 2022) resulting in a further halving of M_0 (Table 5) to a value well in the tail of the age-based mortality prior (Hamel & Cope, 2022). This may be a consequence of the potential trade-off between the Lorenzen exponent and M_0 , primarily fitting to younger cohorts which may be more sensitive to the Lorenzen exponent and a lack of older cohorts to stabilize the M_0 estimate. Alternatively, it may be possible that the two parameter Lorenzen mortality-at-age model is not sufficiently flexible to capture the true mortality schedule resulting in estimates that don't appear to be biologically plausible. As mentioned above, the estimated schedule was apparently robust to the choice among the three growth curves considered (Figure 17) though these growth curves likely did not capture the full range of growth variation.

4.1 Future work

Given its central role in the analysis, reducing the confounding between reporting rate and M is the most natural focus of future work, and can be pursued along two complementary pathways. The first is to resolve the reporting rates more finely. Given that an external analysis is not bound by the fishery definitions of the integrated assessment, there is scope to resolve reporting rates well beyond the customary stratification by gear, fishing mode, and region (e.g., by flag or sub-fleet operating in the areas where recoveries occur, and by post-capture pathway). A finer stratification of this sort could allow the reconstruction to absorb spatial variation in reporting that is otherwise loaded onto M . The second path is to constrain the reporting rates externally through continued tag seeding work. The recent expansion of tag-seeding effort should in theory yield tighter and better-resolved reporting-rate priors. Either development would reduce the flexibility for the reporting rate to absorb misspecification.

A direct test of the method's ability to recover a known mortality schedule under realistic conditions is also a priority. Applying the cohort analysis to tagging data simulated from an individual-based model such as IKAMOANA (Scutt Phillips et al., 2018), with prescribed spatial structure in fishing and natural mortality, tag-induced mortality, and spatially variable reporting, would allow the recovery of a known M schedule to be assessed directly. Such an exercise would quantify the bias introduced by exactly the mechanisms identified above, and could in turn guide the design of future tagging programmes and reporting-rate monitoring so as to maximise the information content of the data for the estimation of natural mortality.

Finally, a Bayesian treatment of the same joint model is a natural next step. Posterior sampling would replace the delta-method intervals with a full characterisation of posterior uncertainty given available prior information on natural mortality or reporting rates. A Bayesian workflow would additionally open up diagnostic checks that are not available in the maximum-likelihood setting. Prior predictive checks would allow the prior pushforward through the cohort dynamics to be inspected for biological plausibility before any data are fit, and the distinction between the specified prior and the prior actually admitted by the model

structure, the realised prior, to be made explicit. Simulation-based calibration would validate the prior and sampling infrastructure, while approximate leave-one-out cross-validation would provide a predictive-accuracy comparison among models to complement conditional AIC. A comparison of the Bayesian posterior with the maximum-likelihood estimates reported here would also serve as a check on the accuracy of the Laplace approximation for this model.

5 Acknowledgements

We thank the Pacific Community (SPC) for access to the Coral Sea and Regional Tuna Tagging Programme and Pacific Tuna Tagging Programme datasets, and the staff and observers responsible for the tagging operations and for the longline observer sampling that underlies the sex-ratio data. We thank colleagues at the Pacific Islands Fisheries Science Center and the Inter-American Tropical Tuna Commission for discussion of the cohort analysis methodology.

6 Declaration of generative AI use

Generative AI tools were used to assist with code development, refactoring, and documentation, and with drafting and editing text in this document. All analytical choices, model specifications, diagnostic interpretations, and conclusions are those of the authors, who reviewed and verified all AI-assisted output. The cohort analysis methodology, the model implementation, and the scientific findings are the responsibility of the authors.

7 References

- Berger, A., McKechnie, S., Abascal, F., Kumasi, B., Usu, T., & Nicol, S. (2014). *Analysis of tagging data for the 2014 tropical tuna assessments: Data quality rules, tagger effects, and reporting rates* (WCPFC-SC10-2014/SA-IP-06).
- Day, J., Magnusson, A., Teears, T., Hampton, J., Davies, N., Castillo-Jordan, C., Peatman, T., Scott, R., Scutt Philips, J., McKechnie, S., Scott, F., Yao, N., Natadra, R., Pilling, G., Williams, P., & Hamer, P. (2023). *Stock assessment of bigeye tuna in the western and central Pacific Ocean: 2023* (WCPFC-SC19-2023/SA-WP-05 (Rev. 2)). <https://meetings.wcpfc.int/node/19353>
- Evans, K., Langley, A., Clear, N. P., Williams, P., Patterson, T., Sibert, J., Hampton, J., & Gunn, J. S. (2008). Behaviour and habitat preferences of bigeye tuna (*Thunnus obesus*) and their influence on longline fishery catches in the western Coral Sea. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(11), 2427–2443. <https://doi.org/10.1139/F08-148>
- Eveson, P., Vincent, M., Farley, J., Krusic-Golub, K., & Hampton, J. (2020). *Integrated growth models from otolith and tagging data for yellowfin and bigeye tuna in the western and central Pacific Ocean* (WCPFC-SC16-2020/SA-IP-03).
- Hamel, O. S., & Cope, J. M. (2022). Development and considerations for application of a longevity-based prior for the natural mortality rate. *Fisheries Research*, 256, 106477. <https://doi.org/10.1016/j.fishres.2022.106477>
- Hampton, J. (1997). Estimates of tag-reporting and tag-shedding rates in a large-scale tuna tagging experiment in the western tropical Pacific Ocean. *Fishery Bulletin*, 95(1), 68–79.

- Hampton, J., & Gunn, J. (1998). Exploitation and movements of yellowfin tuna (*Thunnus albacares*) and bigeye tuna (*T. Obesus*) tagged in the north-western Coral Sea. *Marine and Freshwater Research*, 49(6), 475–489. <https://doi.org/10.1071/MF97210>
- Hearn, W. S., Sandland, R. L., & Hampton, J. (1987). Robust estimation of the natural mortality rate in a completed tagging experiment with variable fishing intensity. *ICES Journal of Marine Science*, 43(2), 107–117. <https://doi.org/10.1093/icesjms/43.2.107>
- Kristensen, K. (2023). *RTMB: 'R' Bindings for 'TMB'*. <https://doi.org/10.32614/CRAN.package.RTMB>
- Lorenzen, K. (2022). Size- and age-dependent natural mortality in fish populations: Biology, models, implications, and a generalized length-inverse mortality paradigm. *Fisheries Research*, 255, 106454. <https://doi.org/10.1016/j.fishres.2022.106454>
- Lorenzen, K., Camp, E. V., & Garlock, T. M. (2022). Natural mortality and body size in fish populations. *Fisheries Research*, 252, 106327. <https://doi.org/10.1016/j.fishres.2022.106327>
- Maunder, M., Aires-da-Silva, A., Deriso, R., Schaefer, K., & Fuller, D. (2009). *Preliminary estimation of age- and sex-specific natural mortality of bigeye tuna in the eastern pacific ocean by applying a cohort analysis with auxiliary information to tagging data* (Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 10; pp. 253–278). <https://www.iattc.org/GetAttachment/1f468f0a-ec78-489e-be69-229988f13752/Status%20of%20the%20tuna%20and%20billfish%20stocks%20in%202008>
- Minte-Vera, C., Maunder, M. N., Xu, H., Bi, R., & Fuller, D. (2025). *Auxiliary analyses for the 2025 benchmark assessment of yellowfin tuna in the EPO* (SAC-16-INF-F). IATTC.
- Scutt Phillips, J., Sen Gupta, A., Senina, I., Van Sebille, E., Lange, M., Lehodey, P., Hampton, J., & Nicol, S. (2018). An individual-based model of skipjack tuna (*Katsuwonus pelamis*) movement in the tropical Pacific ocean. *Progress in Oceanography*, 164, 63–74. <https://doi.org/10.1016/j.pocean.2018.04.007>
- Vincent, M., Aoki, Y., Kiyofuji, H., Hampton, J., & Pilling, G. (2019). *Background analyses for the 2019 stock assessment of skipjack tuna* (WCPFC-SC15-2019/SA-IP-04).
- Zheng, N., Cadigan, N., & Thorson, J. T. (2024). *A note on numerical evaluation of conditional Akaike information for nonlinear mixed-effects models*. arXiv. <https://doi.org/10.48550/ARXIV.2411.14185>

8 Tables

Table 1: Tag release and recapture counts for the three tagging programs used in the analysis, under the 2023 baseline growth scenario. — indicates the criterion is not applicable for that program. Failure reasons for recaptures are not mutually exclusive: a single record can fail more than one check. CSTP: archival (electronic) tags, Coral Sea Tagging Program. RTTP: conventional dart tags, Regional Tuna Tagging Programme (Coral Sea). PTTP: conventional dart tags, Pacific Tuna Tagging Programme (Region 4).

	CSTP	RTTP	PTTP
Raw tags	220	4257	50894
Excl.: duplicate records	169	—	—
Excl.: missing/invalid release data	—	157	0
Tags retained for analysis	51	4100	50894
Total recaptures	18	307	10098
Excl.: missing catch date or gear	—	0	1291
Excl.: negative days at liberty	—	0	7
Excl.: low/missing date reliability	—	—	1043
Usable recaptures	18	307	8381
Cohorts defined	11	65	1221
Cohorts excl. (zero recaptures)	3	26	540
Cohorts retained	8	39	681

Table 2: Information criteria for the four joint models, in development order. mAIC is marginal AIC; cAIC is conditional AIC (Zheng et al. 2024); EDF is the effective degrees of freedom of the release-event random-effect block. cAIC and EDF are not defined for the fixed-effects model, which has no random-effect block. Max grad. is the maximum absolute gradient component at the reported solution.

	Model	n par (fixed)	n par (random)	Total NLL	mAIC	cAIC	EDF	Max grad.
7	Joint (FE)	2	0	1101.44	2206.9	NA	NA	3.473e-13
8	Joint (RE)	3	240	1046.51	2099.0	2039.1	87.84	2.249e-06
9	Joint RE+PTTP	4	240	908.79	1825.6	1764.4	30.75	9.798e-05
10	Joint RE+PTTP+RTTP	7	240	907.51	1829.0	1758.2	25.34	1.982e-04

Table 3: Fit and likelihood summaries for every model fitted in the analysis. Stage groups the models into the independent per-program fits, the joint models, the structural extensions, and the growth sensitivity. n par is the total parameter count, split into fixed and random. The four NLL columns are the release, sex-ratio, random-effect, and reporting-rate-penalty components of the negative log-likelihood; components a model does not use are reported as 0. NLL total is the objective at the reported solution. mAIC is marginal AIC; cAIC is conditional AIC (Zheng et al. 2024); EDF is the effective degrees of freedom of the random-effect block. cAIC and EDF are not defined for fixed-effects-only models and are shown as an em dash.

Model	Stage	n par	fixed	random	NLL rel.	NLL sex	NLL RE	NLL pen.	NLL total	mAIC	cAIC	EDF
CSTP (FE)	Independent	2	2	0	7.29	0.00	0.00	0.00	7.29	18.6	—	—
CSTP (RE)	Independent	7	3	4	6.15	0.00	-2.17	0.00	7.21	20.4	20.2	0.9
RTTP (FE)	Independent	2	2	0	58.54	0.00	0.00	0.00	58.54	121.1	—	—
RTTP (RE)	Independent	10	3	7	58.54	0.00	-41.92	0.00	58.54	123.1	123.1	0.0
PTTP (FE)	Independent	2	2	0	1000.42	0.00	0.00	0.00	1000.42	2004.8	—	—
PTTP (RE)	Independent	232	3	229	876.73	0.00	-21.25	0.00	969.58	1945.2	1907.1	73.8
Joint (FE)	Joint	2	2	0	1101.44	0.00	0.00	0.00	1101.44	2206.9	—	—
Joint (RE)	Joint	243	3	240	928.72	0.00	13.49	0.00	1046.51	2099.0	2039.1	87.8
Joint RE+PTTP	Joint	244	4	240	847.43	0.00	7.75	22.98	908.79	1825.6	1764.4	30.8
Joint RE+PTTP+RTTP	Joint	247	7	240	846.78	0.00	0.91	29.95	907.51	1829.0	1758.2	25.3
Extension: Base	Extension	247	7	240	846.78	0.00	0.91	29.95	907.51	1829.0	1758.2	25.3
Extension: +Sex	Extension	251	11	240	846.78	-48.75	0.91	29.95	858.76	1739.5	1668.7	25.3
Extension: +omega_L	Extension	248	8	240	841.75	0.00	-26.36	28.88	902.19	1820.4	1753.9	27.2
Extension: +Sex+omega_L	Extension	252	12	240	841.75	-48.75	-26.36	28.88	853.44	1730.9	1664.4	27.2
Growth: 2020	Growth sensitivity	247	7	240	802.52	0.00	129.53	29.75	886.16	1786.3	-190790.4	-96204.7
Growth: Farley 2020	Growth sensitivity	247	7	240	927.13	0.00	-6.91	27.01	988.74	1991.5	1924.9	28.3

Table 4: Diagnostics and estimability summaries for every model fitted in the analysis. Max grad. is the maximum absolute gradient component at the reported solution. PDH is the positive-definite-Hessian status. n non-est. is the count of non-estimable parameters from the Hessian eigendecomposition. n corr. pairs is the count of fixed-effect parameter pairs with absolute correlation above 0.95. Bounds is OK when no estimated parameter rests at a box constraint, otherwise the count of parameters at a bound.

Model	Stage	Max grad.	PDH	n non-est.	n corr. pairs	Bounds
CSTP (FE)	Independent	4.835e-08	yes	0	0	OK
CSTP (RE)	Independent	1.419e-06	yes	0	0	OK
RTTP (FE)	Independent	2.569e-15	yes	0	0	OK
RTTP (RE)	Independent	5.083e-05	yes	0	0	1
PTTP (FE)	Independent	9.799e-14	yes	0	0	OK
PTTP (RE)	Independent	1.726e-04	yes	0	0	OK
Joint (FE)	Joint	3.473e-13	yes	0	0	OK
Joint (RE)	Joint	2.249e-06	yes	0	0	OK
Joint RE+PTTP	Joint	9.798e-05	yes	0	0	OK
Joint RE+PTTP+RTTP	Joint	1.982e-04	yes	0	0	OK
Extension: Base	Extension	1.982e-04	yes	0	0	OK
Extension: +Sex	Extension	3.590e-05	no	3	19	OK
Extension: +omega_L	Extension	4.211e-04	yes	0	0	OK
Extension: +Sex+omega_L	Extension	2.459e-04	no	3	21	OK
Growth: 2020	Growth sensitivity	6.461e+07	—	—	2	OK
Growth: Farley 2020	Growth sensitivity	8.947e-05	yes	0	0	OK

Table 5: Estimated baseline natural mortality M0 (mortality at reference length L_ref) and Lorenzen allometric exponent omega_L for all 16 fitted models. Values are formatted as estimate $\pm$ delta-method standard error. Parameters held constant at their initial value are marked as ‘(fixed)’ and carry no standard error.

Model	Group	M0	omega_L
CSTP (FE)	Independent	0.094 $\pm$ 0.018	-1.000 (fixed)
CSTP (RE)	Independent	0.091 $\pm$ 0.021	-1.000 (fixed)
RTTP (FE)	Independent	0.093 $\pm$ 0.007	-1.000 (fixed)
RTTP (RE)	Independent	0.093 $\pm$ 0.007	-1.000 (fixed)
PTTP (FE)	Independent	0.146 $\pm$ 0.004	-1.000 (fixed)
PTTP (RE)	Independent	0.163 $\pm$ 0.006	-1.000 (fixed)
Joint (FE)	Joint	0.128 $\pm$ 0.003	-1.000 (fixed)
Joint (RE)	Joint	0.158 $\pm$ 0.006	-1.000 (fixed)
Joint RE+PTTP	Joint	0.068 $\pm$ 0.007	-1.000 (fixed)
Joint RE+PTTP+RTTP	Joint	0.062 $\pm$ 0.008	-1.000 (fixed)
Extension: Base	Extension	0.062 $\pm$ 0.008	-1.000 (fixed)
Extension: +Sex	Extension	0.062 $\pm$ 0.008	-1.000 (fixed)
Extension: +omega_L	Extension	0.040 $\pm$ 0.009	-1.793 $\pm$ 0.246
Extension: +Sex+omega_L	Extension	0.040 $\pm$ 0.009	-1.793 $\pm$ 0.246
Growth: 2020	Growth sensitivity	0.059 $\pm$ 0.002	-1.000 (fixed)
Growth: Farley 2020	Growth sensitivity	0.062 $\pm$ 0.007	-1.000 (fixed)

Table 6: Natural mortality at annual ages from the selected joint RE+PTTP+RTTP model. Only the 10 annual ages are shown here (quarterly ages 4, 8, $\dots$, 40; corresponding to years 1–10); the complete schedule over all 40 quarterly ages is shown in the Figures section. Standard errors and 90% confidence intervals are delta-method approximations for the M_full ADREPORT quantity.

Age (yr)	Age (quarters)	M	SE	Lower 90%	Upper 90%
1	4	0.1435	0.0174	0.1149	0.1722
2	8	0.0970	0.0118	0.0777	0.1163
3	12	0.0804	0.0097	0.0644	0.0964
4	16	0.0725	0.0088	0.0580	0.0870
5	20	0.0682	0.0083	0.0546	0.0818
6	24	0.0658	0.0080	0.0526	0.0789
7	28	0.0643	0.0078	0.0515	0.0771
8	32	0.0634	0.0077	0.0507	0.0760
9	36	0.0628	0.0076	0.0503	0.0753
10	40	0.0624	0.0076	0.0500	0.0749

Table 7: Gear-specific reporting rates for the selected joint RE+PTTP+RTTP model, on the natural scale. CSTEP_archival is held fixed. PTTP_S, RTTP_L, and RTTP_P are estimated under logit-normal penalties; RTTP_S and RTTP_H are estimated under a single tied value. The penalty centre is the natural-scale mean of the logit-normal penalty anchoring each estimated rate.

Gear	Reporting rate	Status	Penalty centre (natural)
CSTEP_archival	0.950	fixed	—
PTTP_S	0.231	estimated	0.530
RTTP_L	0.349	estimated	0.750
RTTP_P	0.583	estimated	0.750
RTTP_S	0.583	estimated	0.586
RTTP_H	0.583	estimated (tied to RTTP_S)	—

Table 8: Structural extensions of the joint RE+PTTP+RTTP model: estimating the Lorenzen exponent ω_L and adding sex structure. PDH is the positive-definite-Hessian status; n non-est. is the count of non-estimable parameters from the Hessian eigen check; n corr. pairs is the count of fixed-effect parameter pairs with absolute correlation above 0.95.

	Model	n par (fixed)	n par (random)	Total NLL	cAIC	EDF	Max grad.	PDH	n non- est.	n corr. pairs
11	Extension: Base	7	240	907.51	1758.2	25.3	1.982e-04	TRUE	0	0
12	Extension: +Sex	11	240	858.76	1668.7	25.3	3.590e-05	FALSE	3	19
13	Extension: + ω_L	8	240	902.19	1753.9	27.2	4.211e-04	TRUE	0	0
14	Extension: +Sex+ ω_L	12	240	853.44	1664.4	27.2	2.459e-04	FALSE	3	21

9 Figures

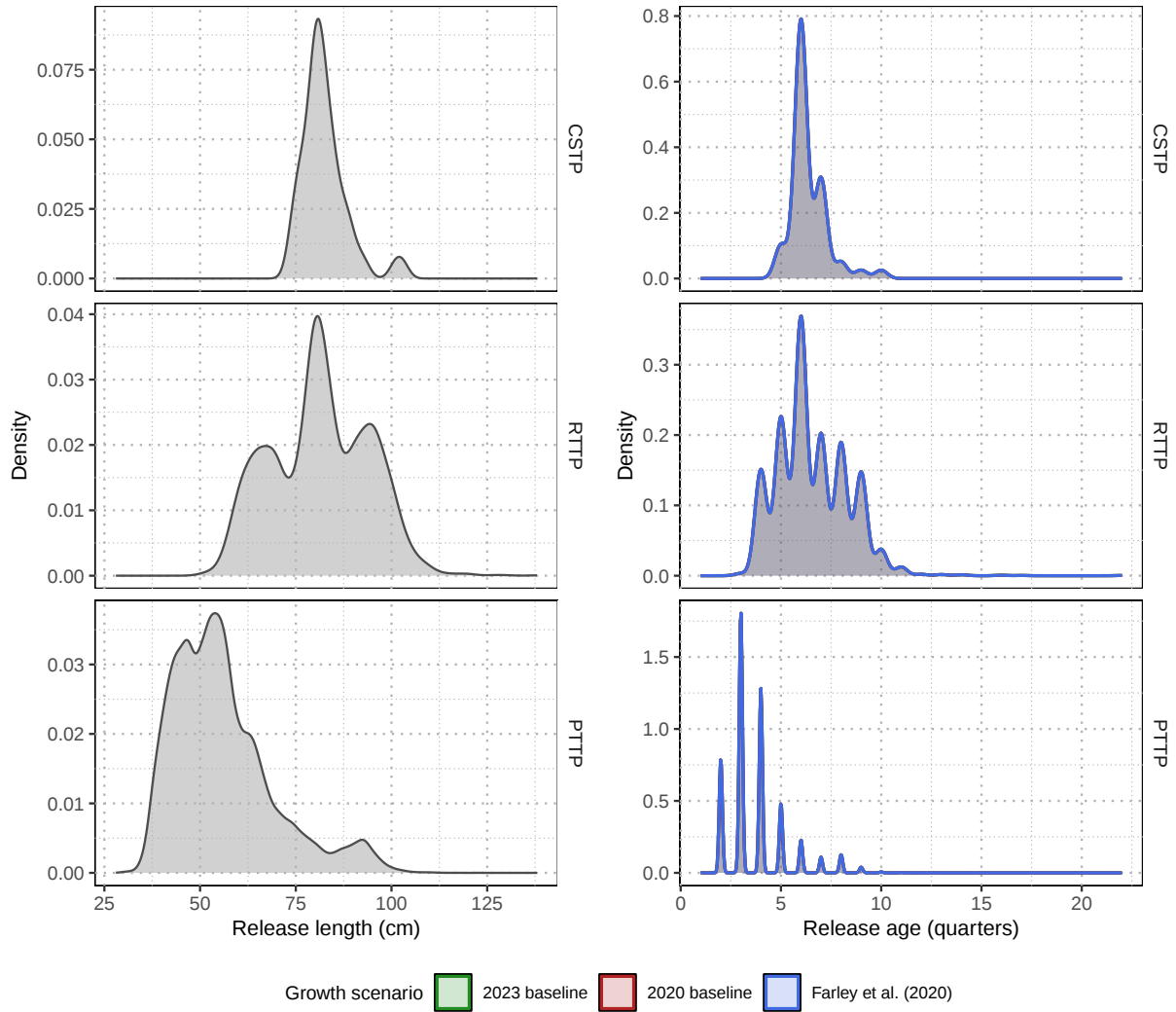


Figure 1: Distributions of release lengths (left column) and release ages in quarters (right column) by tagging program. Release lengths are identical across growth scenarios; the 2023 baseline is shown. Release ages are shown for each of the three growth-curve scenarios used in the analysis.

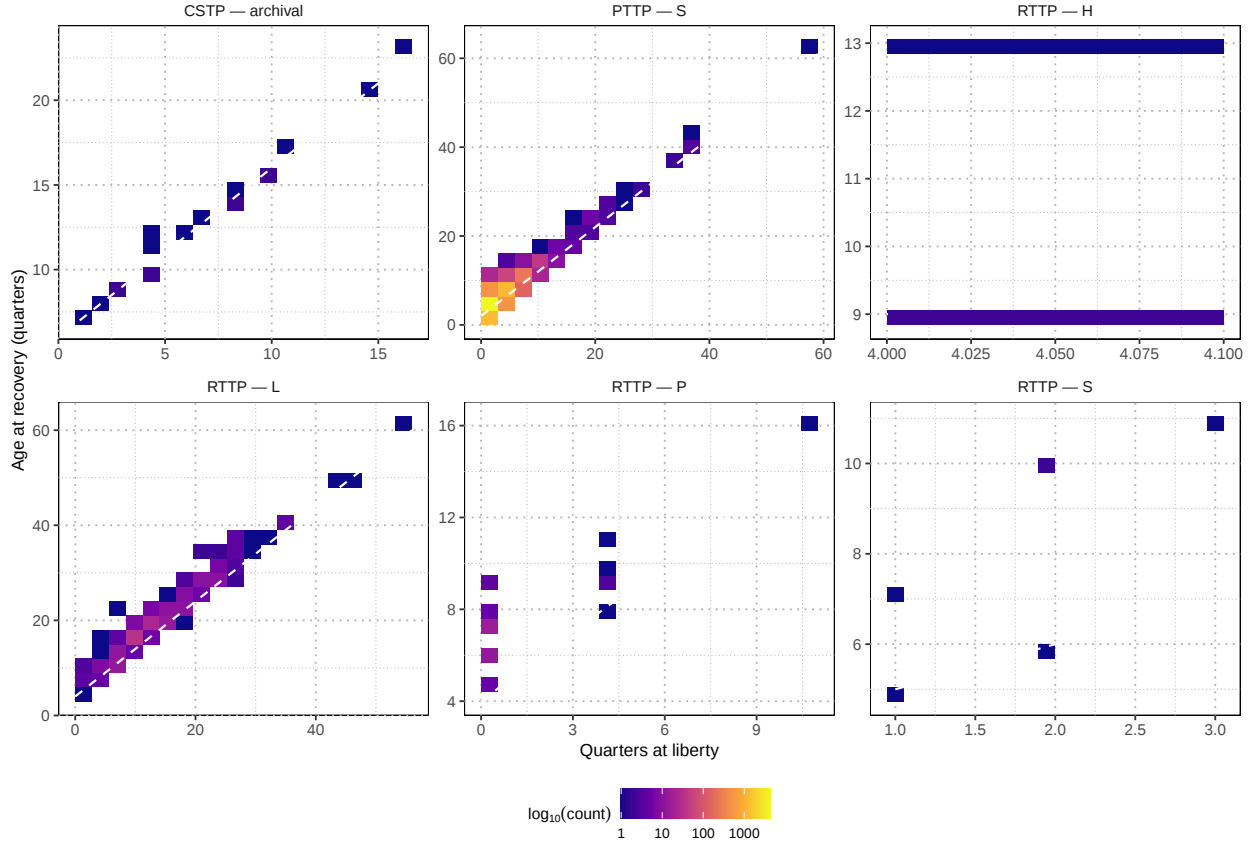


Figure 2: Empirical recovery-age distribution by program and gear, conditional on time at liberty. Each panel is one program by gear combination; bin counts are shown on a log10 fill scale. The dashed diagonal is the 1:1 reference, age at recovery = age at release + quarters at liberty. Systematic departure from the diagonal would indicate age-assignment or growth-curve misspecification.

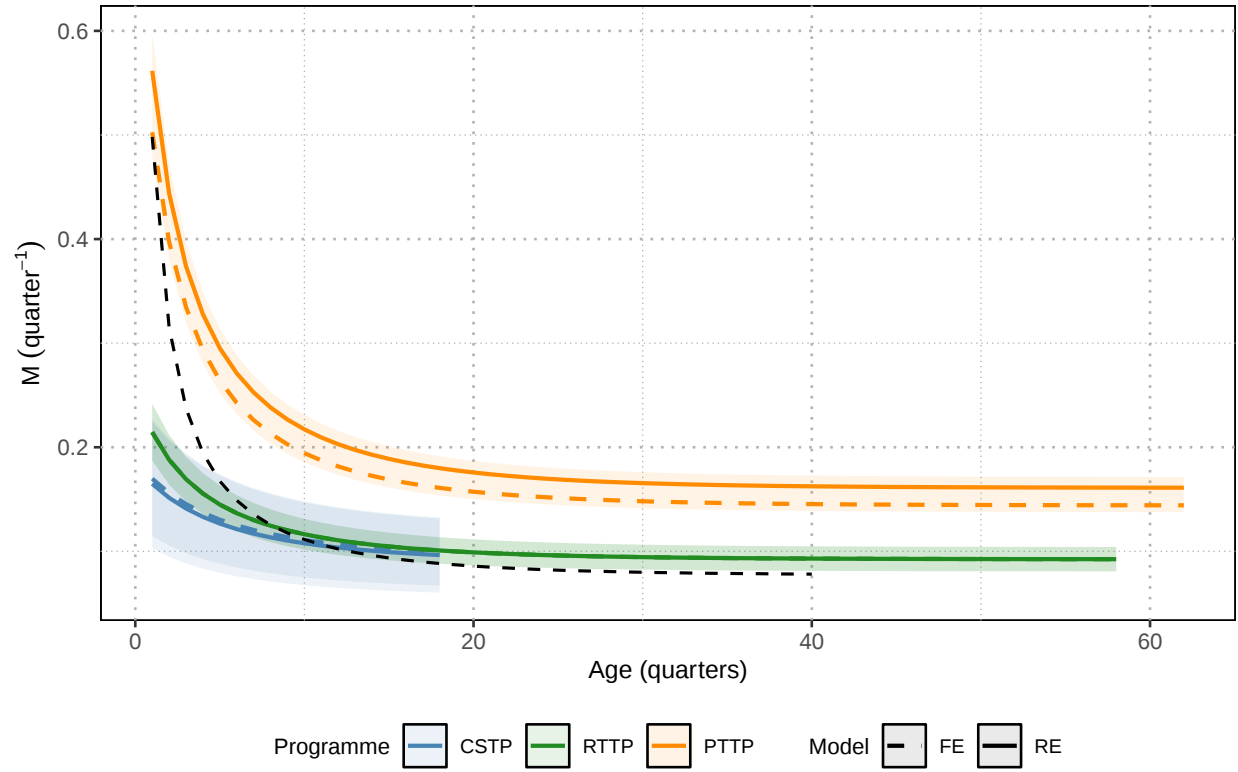


Figure 3: Natural mortality schedules estimated from the independent per-program fits, for CSTP (steelblue), RTTP (forestgreen), and PTTP (darkorange), under fixed-effects (dashed) and release-event random-effects (solid) configurations. Ribbons show 90% delta-method confidence intervals. The dashed black line is the WCPO 2023 diagnostic-case reference schedule.

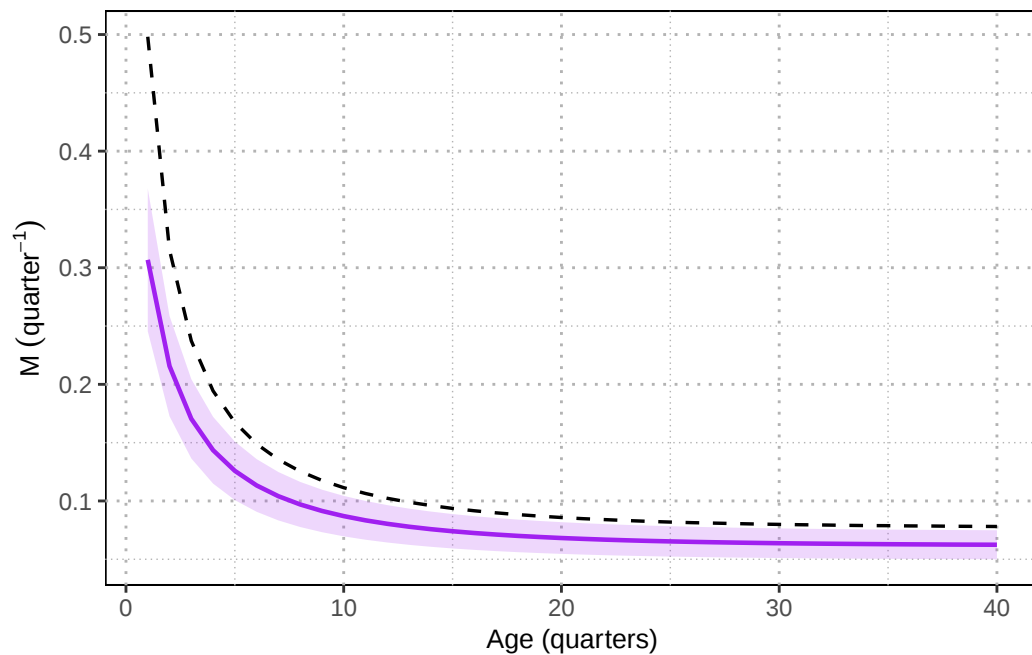


Figure 4: Estimated natural mortality at age from the selected joint RE+PTTP+RTTP model (purple line) with a 90% confidence interval (shaded), extrapolated over the full reporting age range of quarters 1-40. The dashed black line is the WCPO 2023 diagnostic-case M schedule, shown for reference.

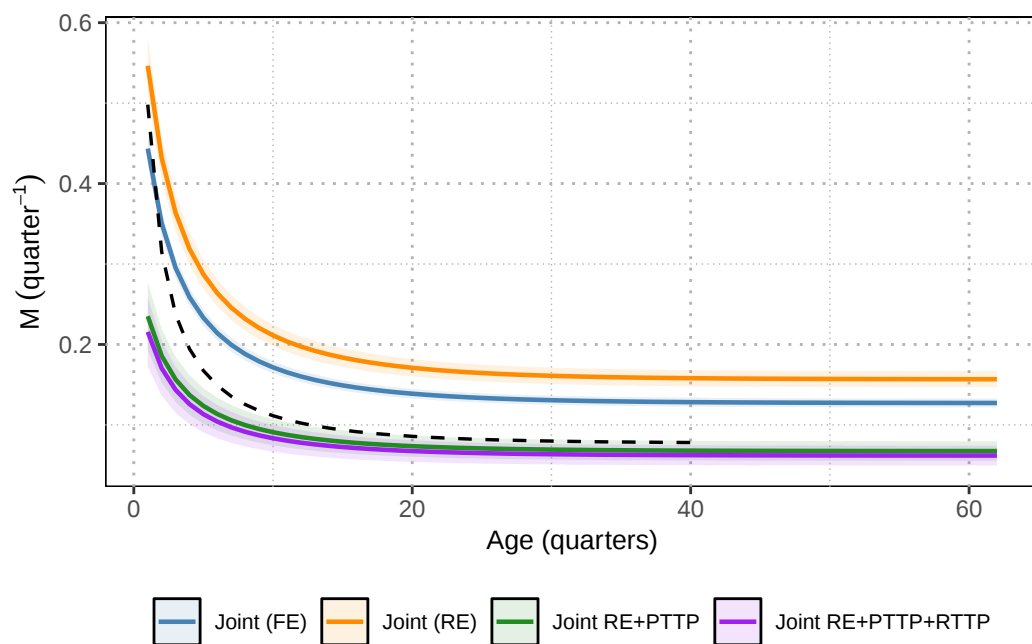


Figure 5: Estimated natural mortality at age across the four joint models, each with a 90% confidence interval ribbon, over the per-cohort modeled age range. The dashed black line is the WCPO 2023 diagnostic-case M schedule. The fixed-effects and random-effects models share fixed reporting rates; the RE+PTTP and RE+PTTP+RTTP models estimate one and four reporting rates respectively under penalties.

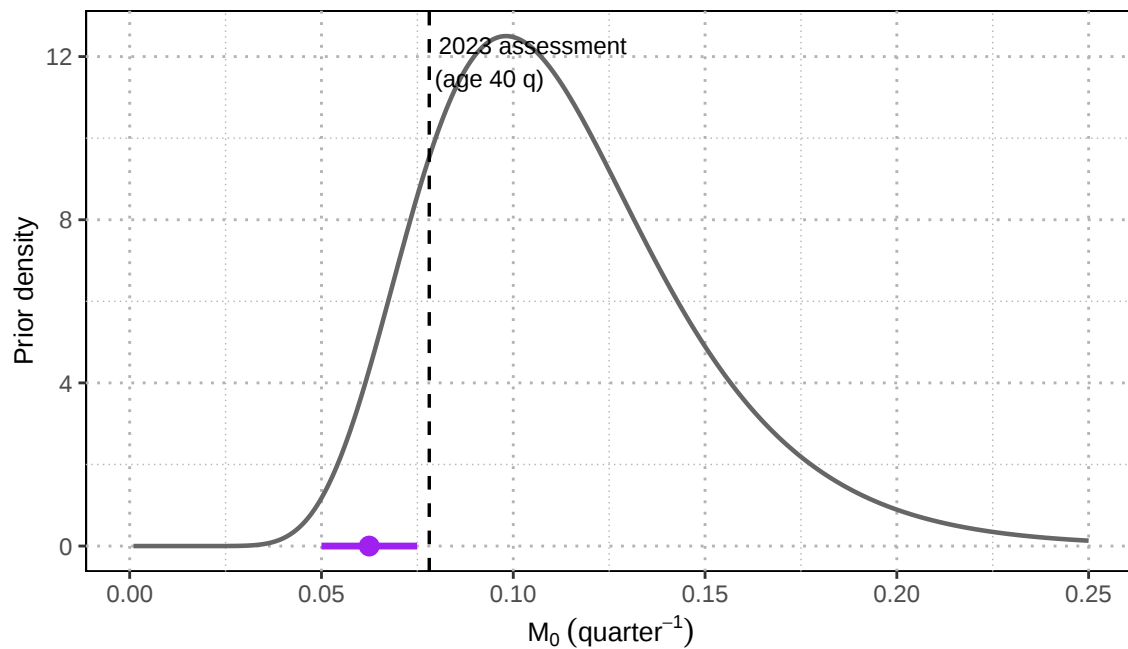


Figure 6: Estimated M_0 (mortality at reference length $L_{\text{ref}} = L(40.5 \text{ quarters})$) from the selected joint RE+PTTP+RTTP model (filled point; horizontal segment spans the 90% delta-method confidence interval) relative to the Hamel & Cope (2022) lognormal prior (curve) and the WCPO 2023 diagnostic-case M at age 40 quarters (vertical dashed line). All values are on the quarterly scale (quarter^{-1}).

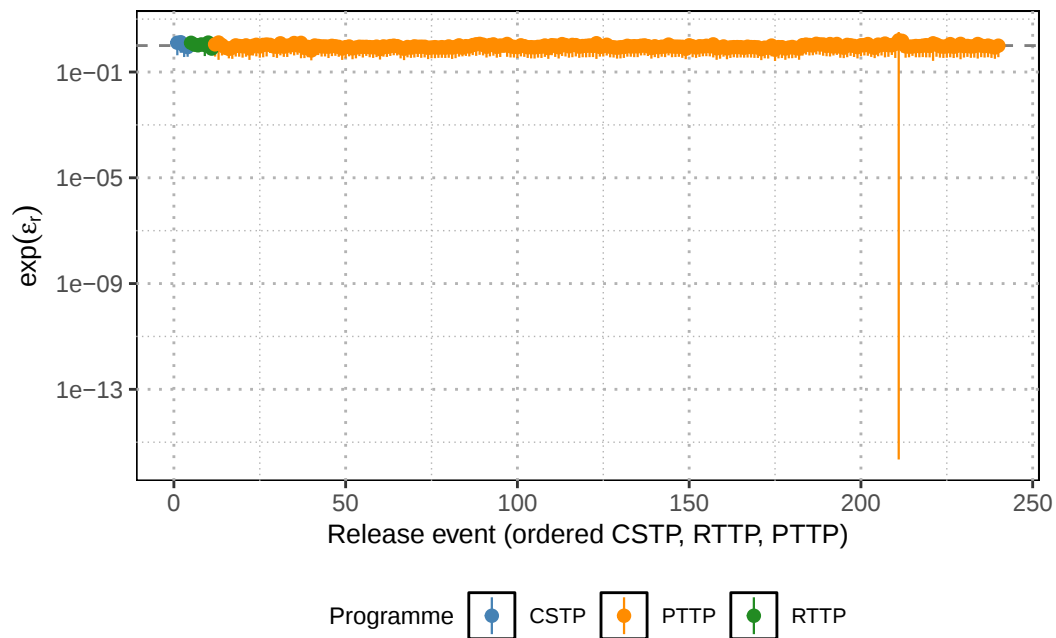


Figure 7: Empirical-Bayes estimates of the multiplicative release-event mortality deviates $\exp(\epsilon_r)$ from the selected joint RE+PTTP+RTTP model, coloured by tagging program. Points are conditional modes; vertical bars are approximate 90% confidence intervals. The dashed line at 1 is the no-deviation reference. Release events are ordered by their first appearance in the stacked cohort list.

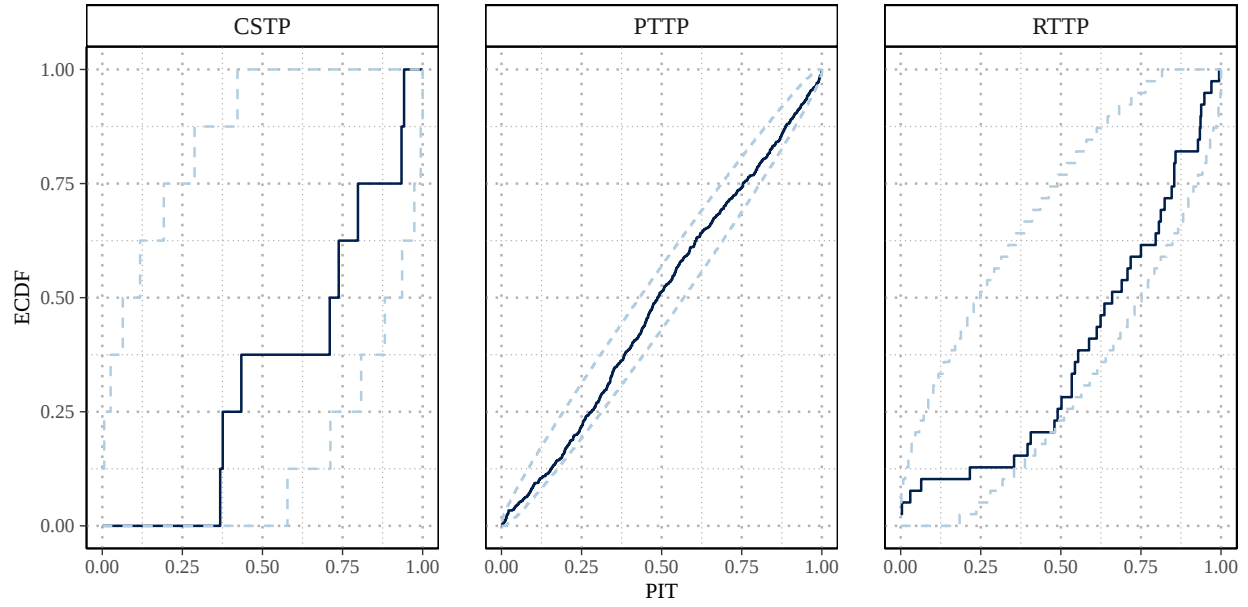


Figure 8: PIT empirical-CDF curves by tagging program for the selected joint RE+PTTP+RTTP model. A well-calibrated model produces curves close to the diagonal; departure within a program indicates systematic over- or under-prediction of that program's releases.

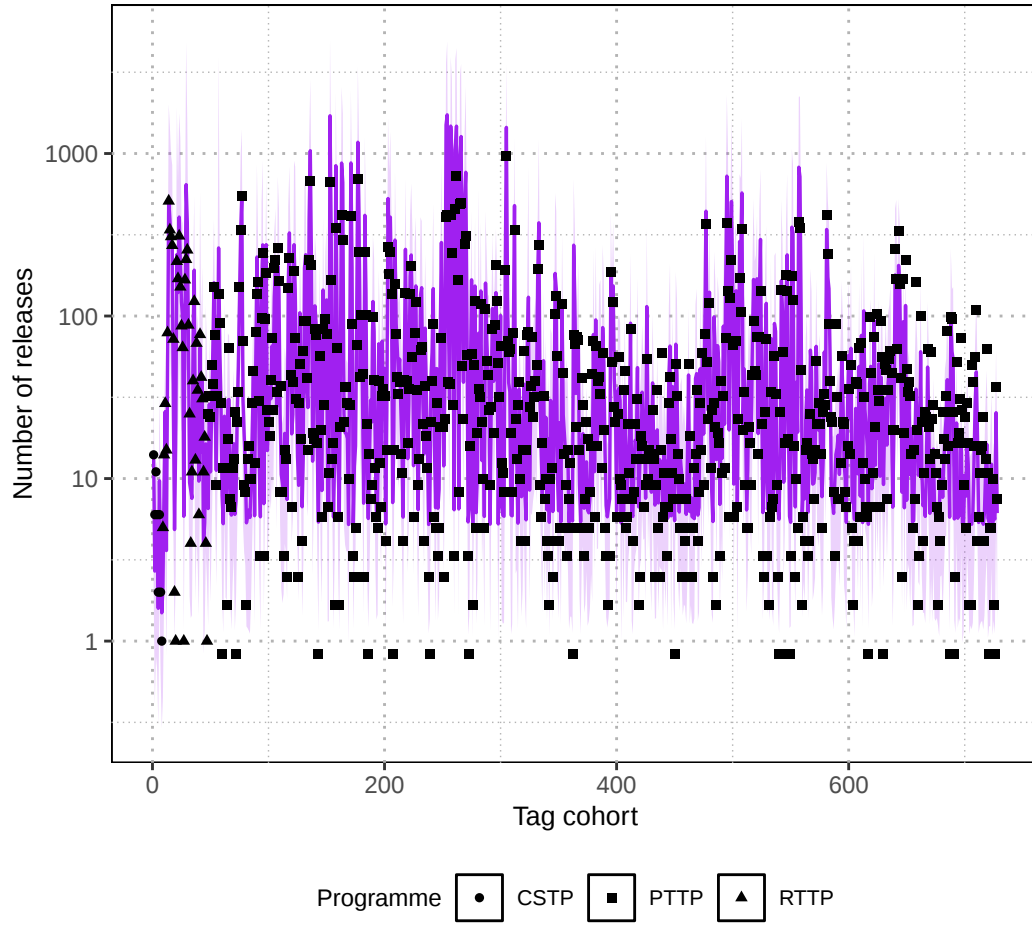


Figure 9: Observed releases (points, by program) and model-predicted releases (purple line) per tag cohort for the selected joint RE+PTTP+RTTP model. The shaded band spans the 5th-95th percentile of the simulated predictive distribution. The vertical scale is on a log10 scale.

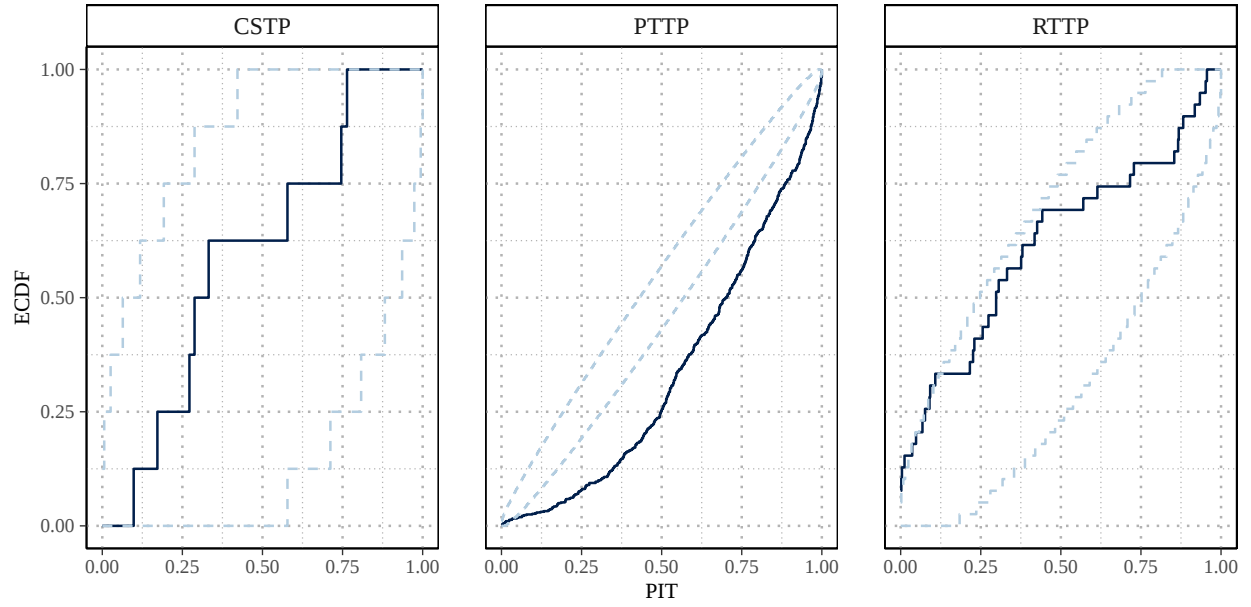


Figure 10: PIT empirical-CDF curves by tagging program for the Joint (FE) model. A well-calibrated model produces curves close to the diagonal; departure within a program indicates systematic over- or under-prediction.

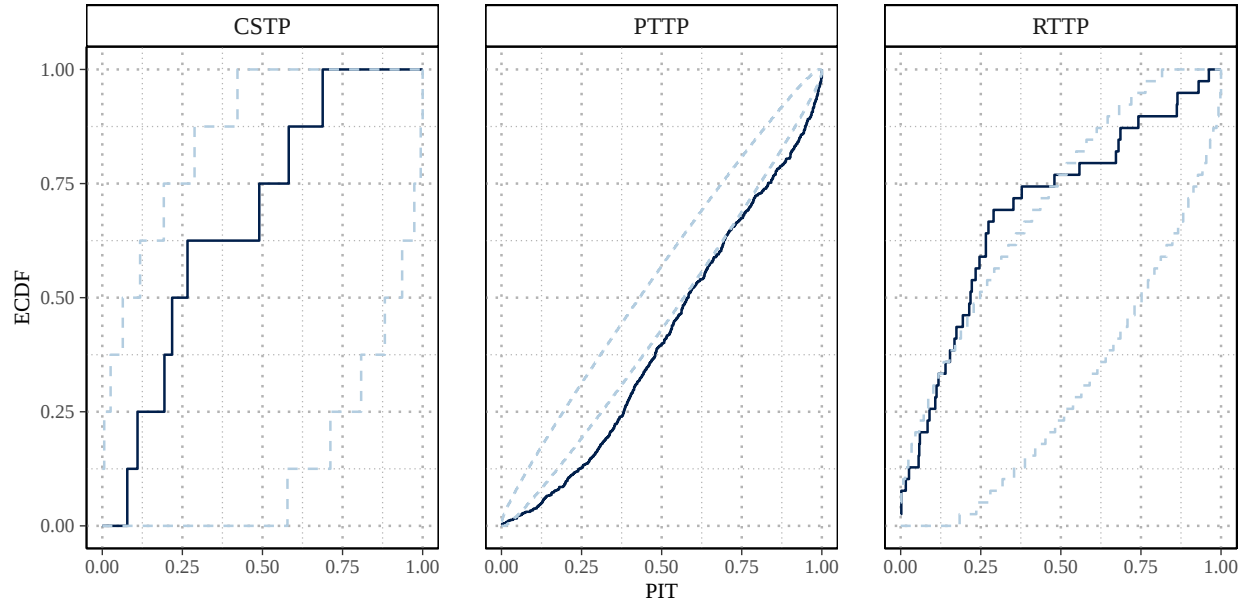


Figure 11: PIT empirical-CDF curves by tagging program for the Joint (RE) model.

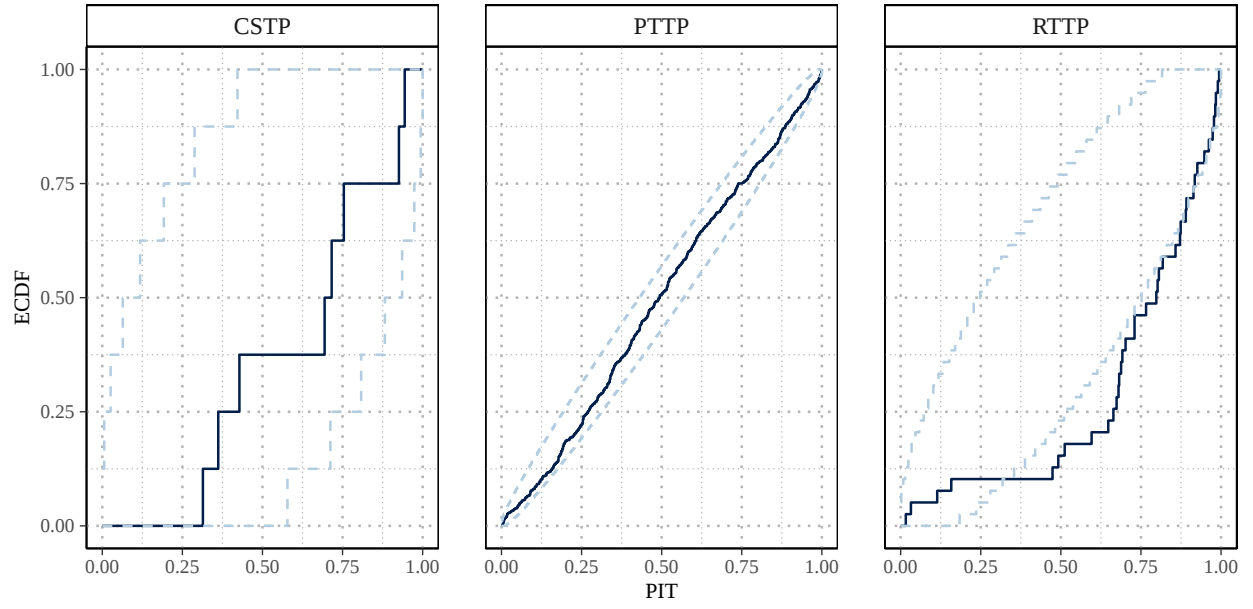


Figure 12: PIT empirical-CDF curves by tagging program for the Joint RE+PTTP model.

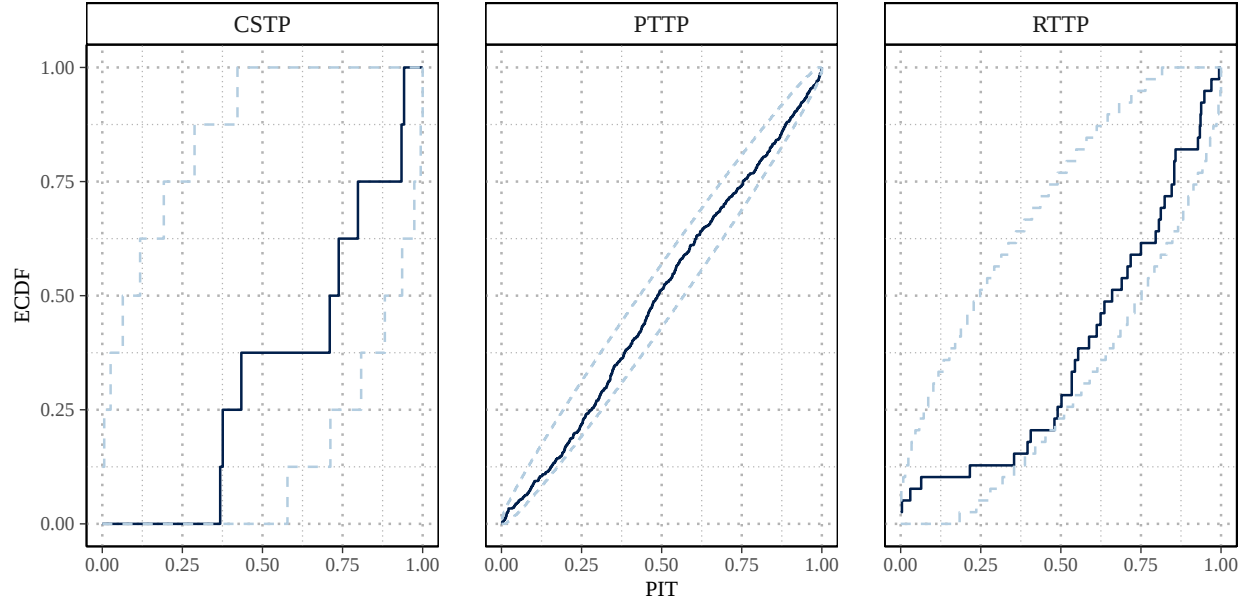


Figure 13: PIT empirical-CDF curves by tagging program for the Joint RE+PTTP+RTTP model (equivalent to Figure 8).

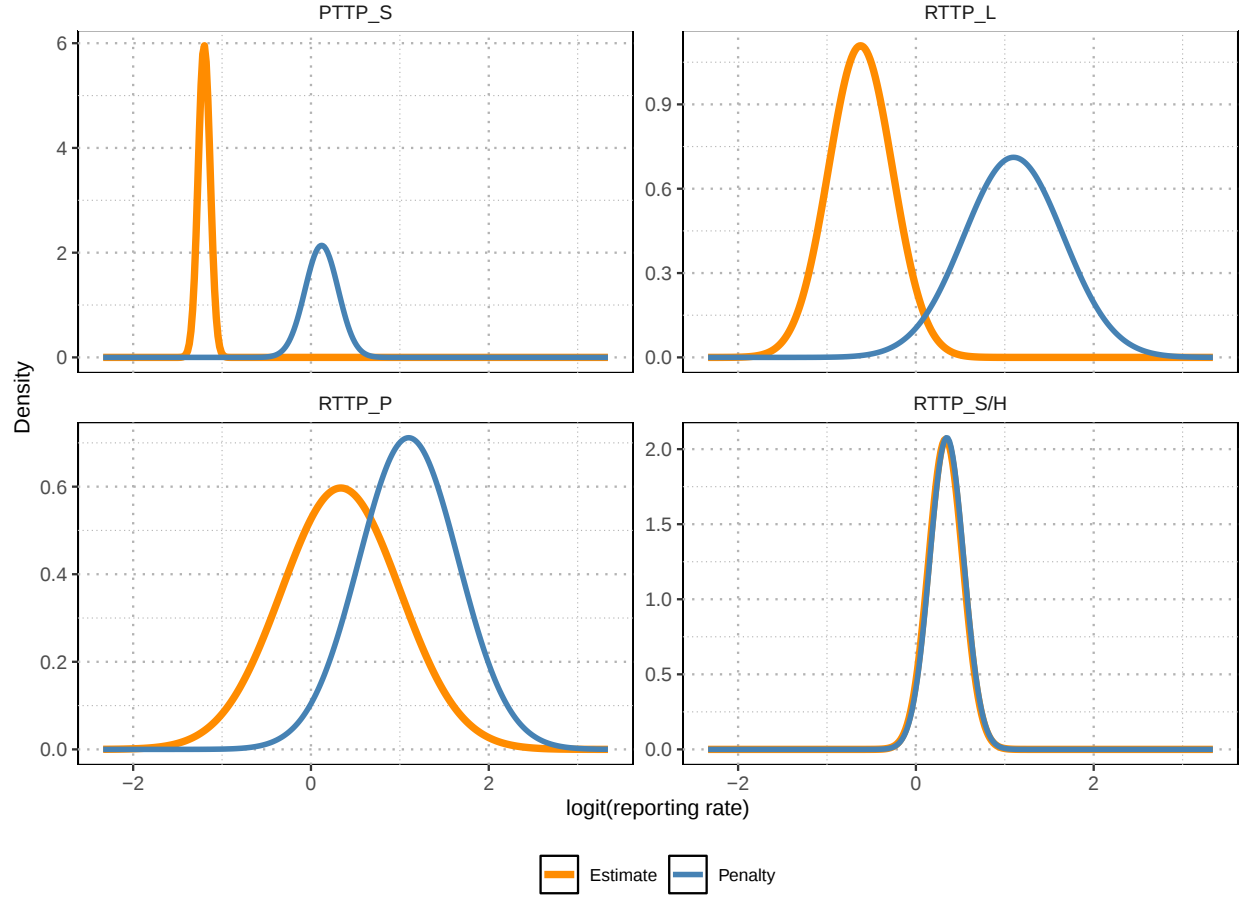


Figure 14: Penalty (blue) and estimated (orange) distributions on the logit scale for the four estimated reporting rates of the selected joint RE+PTTP+RTTP model. The penalty is the logit-normal density supplied as data; the estimate is the logit-scale point estimate with its delta-method standard error.

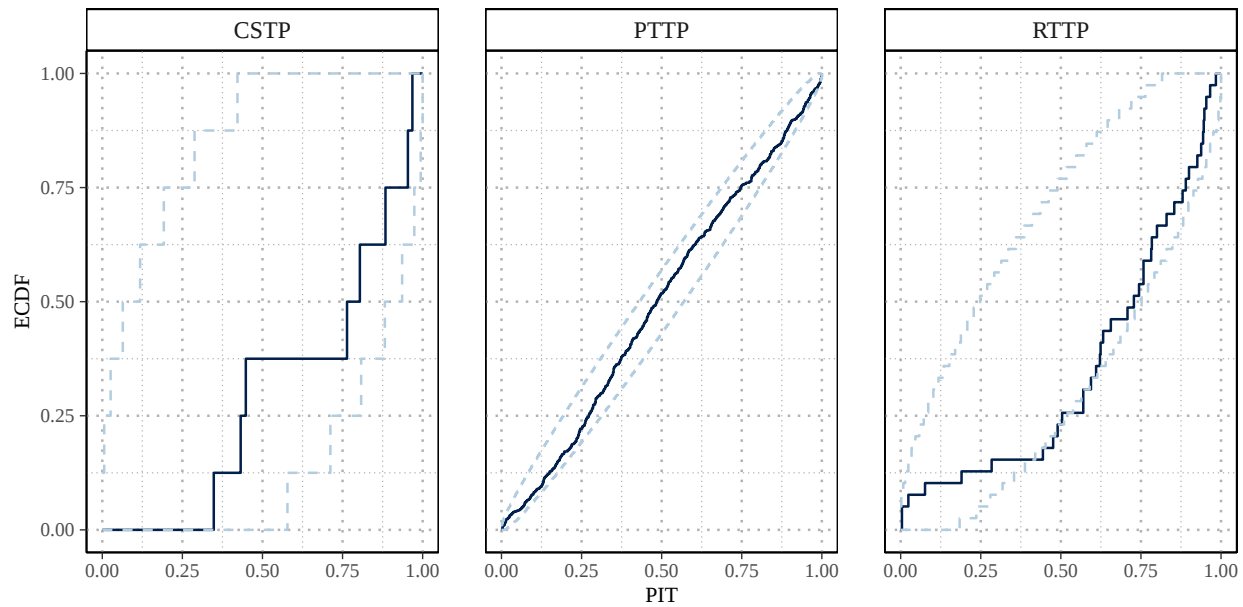


Figure 15: PIT empirical-CDF curves by tagging program for the $+\omega_L$ extension model. A well-calibrated model produces curves close to the diagonal; departure within a program indicates systematic over- or under-prediction of that program's releases.

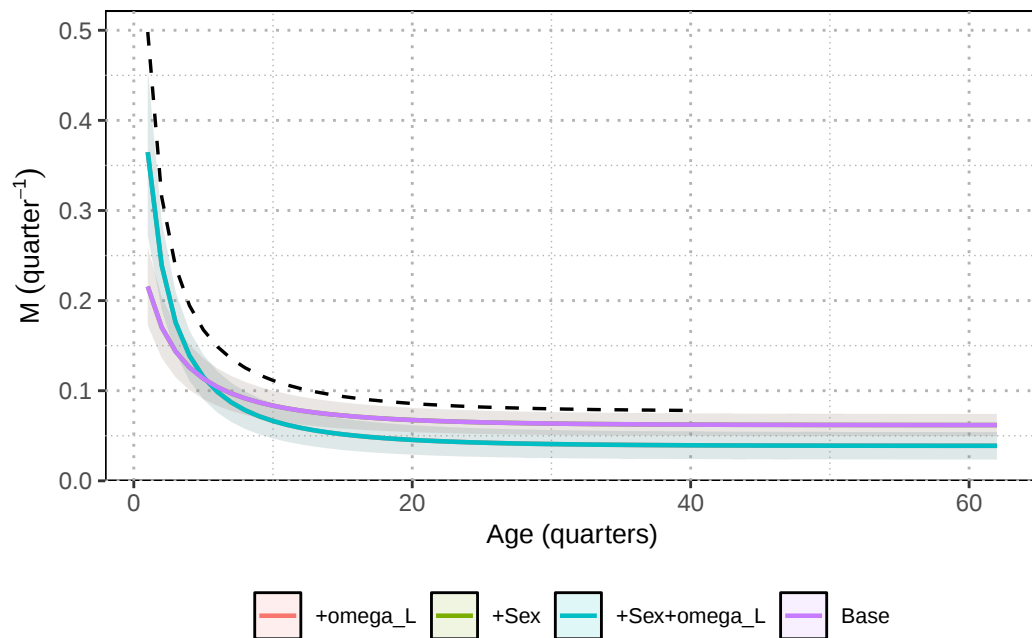


Figure 16: Estimated natural mortality at age across the four extension models, each with 90% confidence interval ribbons. The dashed black line shows the WCPO 2023 diagnostic-case M schedule for reference. Note that the *Base* model covers the *+Sex* model and the *+Sex+omega_L* model covers the *+omega_L* model in the figure.

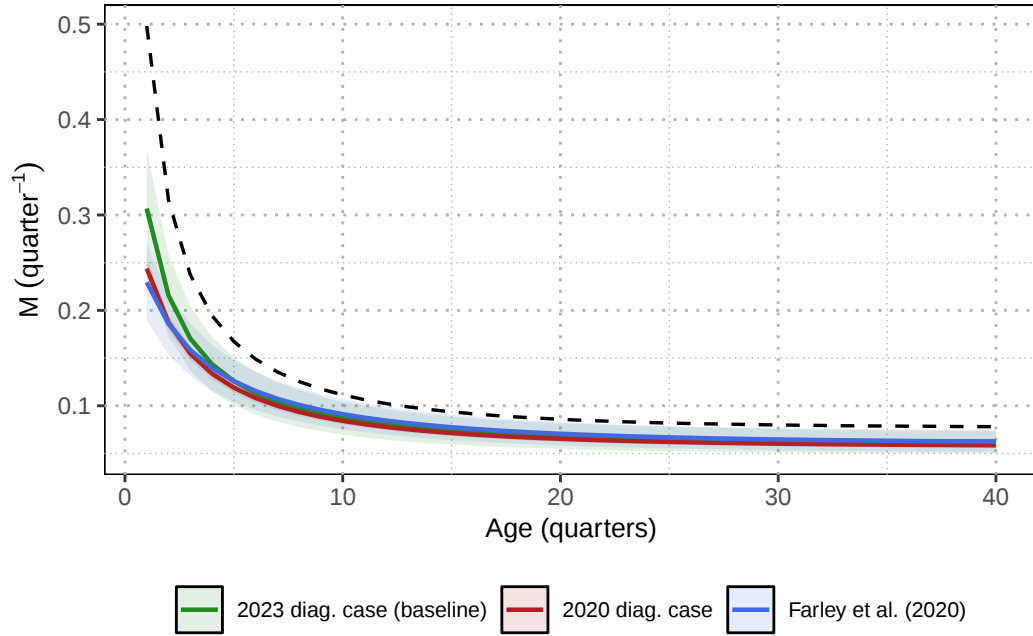


Figure 17: Sensitivity of the estimated natural mortality schedule to the assumed growth curve. The joint RE+PTTP+RTTP model is re-fitted under three Schnute-Richards parameterisations: the 2023 assessment diagnostic case (the baseline), the 2020 assessment diagnostic case, and the curve of Farley et al. (2020). Each line is the M_{full} schedule over quarters 1-40 with a 90% confidence interval ribbon; the dashed black line is the WCPO 2023 diagnostic-case M schedule.