Western and
Central Pacific
Fisheries
Commission

SCIENTIFIC COMMITTEE
TWENTIETH REGULAR SESSION

Manila, Philippines
14 – 21 August 2024

Synthesizing a network of evidence on a seabird bycatch mitigation measure

WCPFC-SC20-2024/EB-IP-08

Eric Gilman, Milani Chaloupka, Igor Debski, Mi Ae Kim, Eric Kingma, Daisuke Ochi

DRAFT - June 2024 - Under Review

Information Paper to the WCPFC Ecosystem and Bycatch Mitigation Theme of SC20

Synthesizing a network of evidence on a seabird bycatch mitigation measure

Eric Gilman[1,*], Milani Chaloupka[2], Igor Debski[3], Mi Ae Kim[4], Eric Kingma[5], Daisuke Ochi[6]

[1] Fisheries Synthesis Science Program, The Safina Center, USA
[2] Ecological Modelling Services Pty Ltd & Marine Spatial Ecology Lab, School of the Environment, University of Queensland, Australia
[3] New Zealand Department of Conservation, New Zealand
[4] Office of International Affairs, Trade and Commerce, NOAA Fisheries, USA
[5] Hawaii Longline Association, USA
[6] Fisheries Resources Institute, Japan Fisheries Research and Education Agency
* Corresponding author: EGilman@utas.edu.au; 3661 Loulu Street, Honolulu, Hawaii 96822 USA

ABSTRACT

Robust estimates of the relative efficacies of alternative management interventions are essential for developing evidence-informed fisheries bycatch policy. Bycatch is a major threat to the conservation of albatrosses and other pelagic seabirds. Branchline weighting is one approach prescribed by regional fisheries management organizations and the Agreement on the Conservation of Albatrosses and Petrels to reduce seabird bycatch in pelagic longline fisheries. We used a Bayesian multilevel network meta-regression modelling approach to conduct the first synthesis of available evidence to assess the relative efficacies at mitigating seabird bycatch of alternative pelagic longline weighting designs. Unlike conventional pairwise meta-analysis, network meta-analysis enables the simultaneous comparison of multiple interventions within a coherent modelling framework. There was a >97% probability that all weighting designs significantly reduced seabird bycatch compared to a reference design with no weight within 5m of the hook. Nonetheless, some weighting designs were significantly more effective at reducing seabird bycatch than others — for instance, the 2 designs with weights >60g and >1m from the hook performed the best with >93% probability that those 2 designs performed significantly better than 2 more commonly used designs with less weight but attached closer to the hook. These two best performing designs reduced seabird bycatch by *ca.* 89% relative to the reference design. These relative efficacies and rankings, when combined with other performance criteria such as costs to commercial viability and crew safety, support robust evaluations of alternative bycatch management strategies.

Keywords: Bayesian network meta-analysis; Bycatch mitigation; Evidence synthesis; Fishing gear weights; RFMO; Seabirds

1. INTRODUCTION

Incidental capture in fisheries is a major threat to the conservation of pelagic seabirds and reduces fishing efficiency (Brothers et al., 1999; Anderson et al. 2011; Phillips et al., 2016). This mortality source is particularly problematic for albatrosses and petrels, which are two of the three most threatened groups of seabirds (Phillips et al. 2016; Dias et al. 2019). Of the suite of anthropogenic threats that albatrosses and petrels are exposed to, fishing-associated mortality causes the largest population impacts (Heppell et al., 2000; Bakker et al., 2018). For depleted populations, even low fishing-associated mortality levels can have profound effects and risk extirpation (Heppell et al. 2000).

Pelagic longlining is a main fishing gear type supplying canned, frozen and fresh products from tuna and tuna like species (Scombroidei), billfishes (Xiphioidei) and other large pelagic and neritic apex- and meso-predators (Miyake et al., 2010; Williams and Ruaia, 2022). Primarily at higher latitudes, pelagic longline fisheries are a major hazard to exposed pelagic seabirds (Anderson et al. 2011; Dias et al. 2019). Several methods can substantially reduce seabird bycatch rates in pelagic longline fisheries that cause minimal or no costs to economic viability, practicality and crew safety (Hall et al., 2017; ACAP 2023). However, there has been mixed progress in the adoption of seabird bycatch mitigation measures by fisheries management authorities and compliance is uncertain due to deficits in fisheries monitoring, surveillance and enforcement (Juan-Jorda et al., 2018).

Branchline weighting is one option for seabird bycatch mitigation in pelagic longline fisheries, which is typically prescribed to be used in combination with other seabird bycatch mitigation methods. Some pelagic longline vessels conventionally incorporate lead-centered weighted swivels into branchlines of various mass and locations to: increase the gear sink rate; maintain the gear at a desired fishing depth; affect the action (movement) of the terminal baited hooks; and reduce line tangles (Brothers et al., 1999; Beverly et al., 2003). The mass of pelagic longline branchline weights and distance between the weight and hook significantly affect seabird catch risk during setting (Robertson et al., 2013; Melvin et al., 2013, 2014; Santos et al., 2019), as well as during the gear soak and haul (Gilman et al., 2014b; Phillips and Wood, 2020). During setting, these two variables affect the sink rate of baited hooks and their availability to seabirds. The initial baited hook sink rate affects seabird catch risk near the surface. The sink rate of a baited hook begins to be affected by a branchline weight once the weight sinks to a depth when the leader (the section of the branchline between the hook and weight) becomes taut (Robertson et al., 2010). The closer a weight is to the hook and the larger the mass, the less time it takes for the weight to affect the hook's sink rate near the surface and the less time baited hooks are available to seabirds to the extent of their diving depth (Robertson et al., 2013; Jimenez et al., 2019). Similarly, during the gear soak, in fisheries where hooks fish near the surface, in deeper-set fisheries when catch bring the gear to the surface, as well as when the gear shoals from currents and wind, and during the haul when hooks are brought back to the surface, the closer a weight is to the hook, the more likely the weight will reduce the availability of baited hooks to surface-foraging seabirds (Gilman et al., 2014b).

In regions where secondary seabird interactions occur or where primary interactions occasionally occur at depth, the hook sink rate to depths attainable by these deep-diving species is an additional important predictor of seabird catch risk (Jimenez et al., 2012; Melvin et al., 2014; Rollinson et al., 2016; Guilford et al., 2022). Secondary interactions occur when deep-diving seabirds access baited hooks at depth and return them to the sea surface, providing larger primarily surface-foraging seabird species with a second opportunity to access baited hooks (Jimenez et al., 2012; Melvin et al., 2014). For fisheries in these regions, the effect of the branchline weighting design on the baited hook sink rate both near the surface where most or all primary interactions occur and to as deep as *ca*. 25m for secondary interactions by proficient divers such as species of shearwaters and *Procellaria* petrels (Bentley et al., 2021; Rollinson et al., 2016; Frankish et al., 2021) will affect seabird catch risk. In some regions, such as the central north Pacific Ocean, secondary interactions are rare events and seabirds susceptible to capture typically only make body thrusts to reach prey within ca. 1 m of the surface (Prince et al., 1994; Kazama et al., 2019; Bentley et al., 2021, Guilford et al., 2022). For fisheries in these regions, the effect of the branchline weighting design on the sink rate very near the surface has a greater effect on seabird bycatch risk than the sink rate to depth.

Unlike many other seabird bycatch mitigation methods, robust compliance monitoring of a required branchline weighting design when conventional lead-centered swivels are used, which are crimped in place, is feasible by dockside inspections for vessels making relatively short trips, where it would be impractical for crew to move weights during a fishing trip (ACAP, 2023). This is an important consideration given the extremely limited onboard monitoring that occurs in global pelagic longline fisheries (Gilman et al., 2014a). Surveillance of line weighting designs of vessels making long trips would, however, require onboard observers (and possibly by electronic monitoring systems, Gilman et al., 2019). The location of attachment of sliding weights and weighted hook shielding devices can be easily changed by crew during both short and long trips, making dockside inspections an ineffective compliance monitoring approach. This could potentially be addressed by requiring unweighted swivels to be attached at the maximum permitted distance that weights can be attached from hooks.

Five regional fisheries management organizations (RFMOs) have adopted binding measures that include branchline weighting as one option to meet seabird bycatch mitigation requirements by pelagic longline fisheries in designated areas, where each RFMO defines three or four options for weighting designs (IATTC, 2011; ICCAT, 2011; WCPFC, 2018; SIOFA, 2022; IOTC, 2023). The Agreement on the Conservation of Albatrosses and Petrels (ACAP) recommends three weighting design options (ACAP, 2023). ACAP's (2023) recommendations are based on mechanistic studies that measured sink rates of alterative branchline weighting designs when considering whether the estimated sink rates are predicted to afford adequate protection when a vessel is using a bird-scaring *tori* line and considering seabird diving depths. The ACAP (2023) recommendations are also based on findings from three selected studies on the effects of alternative weighting designs on seabird interaction rates (Gianuca et al. 2011; Barrington et al., 2016; Jiménez et al. 2019a; Santos et al. 2019).

The controlled experiments have a higher strength of evidence compared to sink rate mechanistic studies, while findings from meta-analytic synthesis studies have the highest strength of evidence, producing the most robust and generalizable findings that are optimal for informing global and regional seabird bycatch policy (Gilman and Chaloupka, 2023). A synthesis of accumulated evidence of seabird bycatch rate responses to alternative branchline weighting designs, including those prescribed by RFMOs and ACAP, has not previously been conducted. Meta-analytic syntheses, which synthesize estimates sourced from a mixture of independent and context-specific studies, have the lowest risk of error and bias and the highest strength of evidence, producing the most robust and generalizable findings that are optimal for evaluating alternative regional bycatch management strategies (Dicks et al. 2014; Nichols et al. 2019).

This study provides the strongest evidence to guide regional policies on branchline weighting to mitigate seabird bycatch. The study objective was to determine whether certain pelagic longline branchline weighting designs, including those required by RFMOs and recommended by ACAP, have similar or otherwise significantly different seabird catch rates. We also explored informative predictors of seabird bycatch risk other than branchline weighting. We conducted a systematic literature review to compile relevant publications and assembled a dataset suitable for meta-analytic evaluation. We used a Bayesian multilevel network meta-regression modelling approach to synthesize available evidence to assess the relative efficacies of branchline weighting designs that are used in pelagic longline fisheries to mitigate seabird bycatch. A key outcome of the network meta-synthesis is a league table or rankings of those relative efficacies that enable identifying the least to most efficacious designs. Findings can be used by fisheries management authorities to define evidence-informed weighting designs that meet objectives for managing seabird bycatch in pelagic longline fisheries.

2. METHODS

2.1. Data Compilation

A two-tiered systematic and unstructured literature search was employed to compile relevant peer-reviewed published and grey literature, including government observer program data, and to assemble a dataset suitable for meta-analytic evaluation. The methods for the systematic literature review were adapted from the Reporting Standards for Systematic Evidence Syntheses (Haddaway et al., 2018), Collaboration for Environmental Evidence (Pullin et al., 2020), and extended PRISMA framework applicable for network meta-analyses (NMA) (Hutton et al 2015). Details on the literature review methods, including searched databases, search strings, study eligibility criteria and screening process for retention, and extracted data fields are in Supplemental Material Section S1. There were 12 retained publications obtained from the systematic review (Supplemental Material Fig. S1, Table S2). One additional retained publication (Table S2) and five retained observer program datasets from New Zealand, Taiwan, the U.S. Atlantic Ocean, the Hawaii-based U.S. Pacific Ocean shallow-set swordfish fishery, and the Hawaii-based U.S. Pacific Ocean deep-set tuna fishery (Supplemental Material Section S4) were obtained from the unstructured literature review.

2.2. Branchline Weighting Design Treatments

Supplemental Material Table S1a summarizes the branchline weighting designs prescribed by five RFMOs with binding seabird measures and recommended by ACAP. To produce a manageable number of weighting design treatments and with adequate sample sizes we used only the ≥ designs (e.g., the prescribed design ≥45g weights within 1m of the hook was retained and eliminated >45g within 1m). As none of the compiled records had >1 weight within half a meter of the hook, we combined the design prescribed by one of the RFMOs that specifies this limit with an ACAP recommended design of ≥40g within 0.5m. We adapted the RFMO/ACAP designs to provide more detailed definitions so that a single branchline weighting design can fall under only one unique category. An RFMO-prescribed designs of ≥98g attached between >3.5m and ≤4m of the hook was excluded because there was only a single record in this category. We included two categories not prescribed by RFMOs or ACAP. This resulted in the following 7-category treatment variable for branchline weighting design for the modeling, where A is referred to as the reference design and is not prescribed by RFMOs or ACAP, designs B through F are prescribed RFMO or ACAP options (sample sizes are number of records), and design G is not prescribed by RFMOs or ACAP:

A. 0g attached ≤5m from the hook (N=86)
B. ≥40g attached ≤0.5m from the hook (N=42)
C. ≥45g to <60g between >0.5m and ≤1m from the hook (N=37)
D. ≥60g between >0.5m and ≤1m from the hook (N=27)
E. ≥60g to <80g between >1m and ≤3.5m or ≥80g between >2m and ≤3.5m from the hook (N=18)
F. ≥80g between >1m and ≤2m from the hook (N=9)
G. ≥8g to ≤39g between ≤1m or ≥45g to ≤80g between ≥1.3m and ≤4m from the hook (N=52)

Table S1b identifies which of the combined RFMO and ACAP design options are met by each of the above 7 branchline weighting design treatments included in the NMA. Table S3 summarizes the records covered by each of the 7 treatments, including the number of fisheries and regions from which they were derived, and the range and median weight mass and leader lengths of the records within each treatment.

2.3. Statistical Modelling Approach

2.3.1. Visualising the network geometry

We summarise the underlying NMA structure for our study in Figure 1, which comprises the 7 pelagic longline branchline weighting designs (A-G) with 21 weighting design combinations or interactions — this plot is a mandatory component of any network modelling workflow (Liu et al 2023). Recently, Freeman et al (2023) also proposed the use of upset plots (Conway et al 2017) for further exploring and visualising the NMA dataset characteristics, which we implemented using the `ComplexUpset` R package (Krassowski 2020) for the weighting design combinations shown in the network geometry plot (Figure 1). The upset plot display of the network structure is shown in Figure S2.

2.3.2. Statistical modelling framework

NMA is a well-established statistical modelling approach used in the healthcare evaluation, pharmacometrics, psychology and clinical epidemiology domains for comparing the relative efficacies of three or more interventions or treatments by synthesising aggregate and/or individual participant evidence sourced from multiple trials (Ades et al 2024). An accessible introduction to Bayesian NMA can be found in Liu et al (2023), the comparative benefits of Bayesian inference-based meta-analysis can be found in (Pappalardo et al 2020), while an authoritative primer on Bayesian statistical modelling in general can be found in van de Schoot et al (2021).

Conventional pairwise meta-analysis is based on synthesising direct evidence comparing only two treatments from multiple trials that can only generate so-called unrelated treatment effects. For instance, say there are three possible treatments A-C but only AB and AC comparisons are known from direct head-to-head clinical trials — no BC trials have ever occurred and so any inference about such an effect remains unknown in conventional pairwise meta-analysis. NMA leverages both direct and indirect evidence for three or more treatment comparisons, so the BC effect can be inferred indirectly as part of a connected network of trials based on direct evidence from the AB and AC trials — a property known as *coherence* (Ades et al 2024). The relative rankings of those NMA-based treatment effects can then be used to identify the "best" treatment, given that specific set of treatments, and to support evidence-informed cost-effectiveness treatment recommendations given those rankings.

Specifically, we used a Bayesian multilevel (or random-effects) network meta-regression modelling approach (Phillippo et al 2020) to synthesis all available studies comparing longline gear weighting designs on seabird bycatch rates (Figure 1). Importantly, not all seven weighting designs need to be in each study so long as each study comprises a subset of at least two or more of the designs. We could only source the study-specific summarised or aggregate data and not the pelagic longline set-specific data for each study, although a mix of both data sources, if available, can be combined in a single integrated Bayesian multilevel network meta-regression model (Phillippo et al 2020, Singh et al 2022).

We also sourced a range of study-specific marginal covariate summaries that we could then use as potentially informative seabird bycatch predictors in our network meta-regression modelling framework. See Supplemental Material Sections S1 and S2 for a description of potentially informative predictors that were extracted from retained publications and explored for potential inclusion in the models. The explored predictors included latitude, bait type, hooks between floats, oceanic region and five binary predictors for the following additional seabird bycatch mitigation measures: tori line, night setting, side setting, hook shielding device, and blue-dyed bait (Supplemental Material Table S2).

We fitted our multilevel network meta-regression models using the Stan computation back-end (Carpenter et al 2017) via the `multinma` interface for R (Phillippo 2024) with Poisson likelihood (log link) appropriate for study-specific aggregate seabird catch data and with the study-specific aggregate number of hooks (fishing effort) as the exposure metric. The random-effects structure (intercepts only) comprised the study-specific identities. These models are therefore Poisson likelihood GLMMs with the response or effect size being the study-specific log rate ratio (Dias et al 2011, Hooper et al 2000) – also known as the log response ratio

(Lajeunesse 2011). We also used QR-decomposition and numeric regression term centering in our network meta-regression models, fitted using `multinma` (Phillippo 2024), to improve computational efficiency by guarding against potential multicollinearity among covariates that might affect the model posterior geometry (Betancourt 2017).

All models were fit with 4 Markov chains with 5,000 iterations per chain after a warm-up of 1000 iterations. Model convergence was assessed using parameter-specific diagnostics such as multiple chain rank plots, bulk and tail effective sample size metrics and a rank-based *Rhat* statistic (Vehtari et al 2021). All inference was based on the 20,000 network meta-regression model posterior draws, which were accessed as multidimensional arrays to support a range of post-processing functions provided by the `bayesplot` (Gabry et al 2019) and `tidybayes` R packages (Kay 2023). For instance, those posterior draws (samples) were used to derive the uncertainty intervals (HDIs or highest posterior density intervals: Kruschke and Liddell 2018; and see also Pappalardo et al 2020). The `ggplot2` (Wickham 2016), `ggdist` (Kay 2024) and `colorspace` (Zeileis et al 2020) packages for R were used for all the summary graphics while the patchwork R package (Pedersen 2020) was used for all multi-panel arrangements.

2.3.3. Relative ranking of weighting designs

Using a Bayesian inference framework for our network meta-synthesis of the 7 branchline weighting designs (or treatments) enables us to derive the full posterior probability distribution of all the relative weighting design effects and hence robust summaries of uncertainty of those effects. Additionally, this framework supports the robust ranking of those relative treatment effects in the network, which is especially helpful for supporting evidence-informed bycatch mitigation policy. In a Bayesian modelling framework, the ranks are derived from the effect-specific posterior distributions using two approaches based on: (1) the mean or median of the posterior distribution of the ranks with the 95% credible interval (Phillippo et al 2020), or (2) the probability of a specific treatment ranking the best determined as the Surface Under the Cumulative RAnking curve (SUCRA: Veroniki et al 2018, Daly et al 2019). The SUCRA probabilities are appealing in terms of the ease of its interpretation and Liu et al (2023) propose that rank probabilities > 0.90 reflect an efficacious treatment effect. On the other hand, the posterior ranking approach is more robust than SUCRA to underlying model assumptions. We used both approaches here (posterior median ranking, SUCRA).

2.3.4. Model selection using cross-validation

Model selection procedures were based on leave-one-out cross-validation (LOOcv) metrics to estimate any comparative difference in expected predictive accuracy between the various models fitted (Vehtari et al 2017) — such as whether the model included covariates or not. The weight of evidence in favour of one model over any other candidate models was also assessed using Bayesian stacking, which is the Bayesian analogue of model averaging (Yao et al 2018).

2.3.5. Assessing network model assumptions

According to Seide et al (2020), there are three fundamental assumptions in NMA: (1) transitivity, (2) network consistency and (3) network heterogeneity (see also Ades et al 2024 and the concept of exchangeability). Not all our studies comprised all the 7 weighting designs in head-to-head comparisons (see upset plot, Figure S2), so we assume that any missing designs within a study are missing at random — there are no consensus statistical tests of this specific assumption, but it is conceptually related to the following consistency assumption (Seide et al 2020).

The all-important consistency assumption concerns whether the direct (head-to-head) estimates and the indirect estimates (made by inference) within each contrast in the connected network are sufficiently similar. This assumption can be readily evaluated using either so-called global assessment tests (unrelated mean effects or UME model: Spineli 2022) and local assessment tests (node-splitting procedure: Veroniki et al 2013) — we used both forms of consistency assessment tests applied to our best-fit Bayesian network meta-regression model used for inference. The global assessment was done by comparing the expected predictive accuracy between the best-fit model and the unrelated mean effects model using leave-one-out cross-validation (see `*Model selection*` section above). The UME model is based only on direct evidence and estimates each treatment comparison as a separate and unrelated parameter (Spineli 2022).

The local consistency assumption was specifically assessed using visual review of node-splitting plots (Currier et al 2023) derived using the `multinma` package for R (Phillippo 2024). The heterogeneity assumption is especially important for small-sample studies (Ades et al 2024) and can be readily explore using prior sensitivity analysis, which we implement in our modelling workflow by fitting meta-regression models with weakly informative priors (Lemoine et al 2019) and then using more informative priors recommended for various forms of the between-study heterogeneity parameter (Turner et al 2015, Seide et al 2020).

2.3.6. Weighting design-specific relative risk reduction in seabird bycatch

We back-transformed the predicted seabird bycatch log relative risk contrasts derived for the 6 branchline weighting design treatments (B-G) relative to reference design A (no weight ≤5m from the hook) using the transform-then-summarize posterior samples approach since such summary metrics are not transform invariant (Wang et al 2018). We then used these back-transformed posterior samples to calculate the mean percent relative reduction in seabird bycatch risk (Gilman et al 2020) for the 6 branchline weighting design treatments relative to the reference design. As in Gilman et al (2020), we also used the highest posterior density interval (HDI) as our measure of uncertainty (Kruschke and Liddell 2018). The mean and 95% HDIs were summarized from the posterior samples for each design relative to the reference design using the `tidybayes` package for R (Kay 2023).

3. RESULTS

3.1. Model Evaluation Diagnostics

Leave-one-out cross-validation (LOOcv) and Bayesian stacking metrics suggested that the best-fit model fitted to the pelagic longline weighting design network geometry (Figure 1) was a meta-regression model that included predictors for oceanic region and the 5 seabird bycatch mitigation measures (looic difference = -105.9, se = 75.9; *ca.* 98% weight of evidence in favour of model with covariates) — there was little support for a NMA model without the covariates. Convergence diagnostics such as multiple chain rank plots, and the effective posterior sample size (ESS) metrics coupled with the rank-based diagnostic statistic *Rhat* < 1.01 (Vehtari et al 2021), all reflected convergence of the Bayesian network meta-regression model with Poisson likelihood that was used here for statistical inference. Nonetheless, residual deviance for all models fitted here was greater than the number of data points, which can suggest that there might be grounds for further improvement in model fit (Phillippo et al 2020) — see `Covariate effects` section below.

3.2. Prior Sensitivity

We assessed comparative predictive accuracy of three between-study heterogeneity parameter priors for the meta-regression model: (1) a weakly informative *half-normal*(scale=5^2) prior on the heterogeneity parameter standard deviation or τ (Phillippo 2024), (2) a more informative *half-normal*(scale=0.5) prior on the heterogeneity variance (τ^2) proposed by Seide et al (2020) on theoretic grounds, and (3) a more informative empirically-derived l*ognormal*(-3.93, 1.51^2) prior on τ^2 suggested by Turner et al (2015). LOOcv and Bayesian stacking metrics suggested there was no meaningful difference in predictive accuracy between a network meta-regression model with the weakly informative heterogeneity prior and the same meta-regression model but now with a more informative *lognormal* heterogeneity prior (looic difference = -1.6, se = 1.9) — there was no support for the informative prior suggested by Seide et al (2020). There is a strong empirical basis for the *lognormal* heterogeneity prior (Turner et al 2015), so we proceed with inference based on our network meta-regression model with that prior on the heterogeneity parameter variance.

3.3. Weighting Design Specific Effects

The predicted design-specific conditional effects relative to the reference design (A) based on the best-fit Bayesian network meta-regression model fitted to the weighting designs network structure (Figure 1) are summarised in Figure 2. This summary display shows both the full posterior distribution of each parameter coupled with key summary metrics (median, highest posterior density intervals) that helps to support a more informative means of communicating parameter uncertainty (van der Bles et al 2019). The six branchline weighting designs B-G are all significantly more effective at reducing expected seabird bycatch than weighting design A — the predicted effects are all significantly < 0 on the log response rate scale. Even for design G we can be > 97% sure that the predicted effect is < 0 (in other words, > 97% of the posterior density shown in Figure 2 is < 0). We can be near 100% sure of this reduced seabird bycatch effect for design E and *ca.* 99% sure for design F. The largest relative effect is for design F but there is greater uncertainty associated with that effect due to the smaller sample size for design-specific comparisons that included F (see Figure 1, Figure S2). Given this uncertainty, it is apparent that weighting designs E and F appear to have similar efficacy for reducing seabird bycatch in pelagic longline fisheries.

3.4. Covariate Effects

Figure S3 summarises the conditional effects for the covariates or predictors of seabird bycatch rate included in our best-fit network meta-regression model. The most prominent conditional effects were for (1) the apparent decreased seabird bycatch risk for pelagic longline fisheries operating in the South Pacific region compared to fisheries in the North Atlantic region, and (2) the apparent decreased seabird bycatch risk given the use of night-setting or side-setting as seabird mitigation measures. Note that all covariates were study-specific marginal covariate summaries of aggregate data and not based on set-specific predictors.

Despite inclusion of these covariates as main-effects, there remained moderate between-study heterogeneity, by comparison with expected treatment effects (Figure 2, see also Phillippo et al 2020), that was unaccounted for by our best-fit network meta-regression model (posterior mean $\tau = 2.02$; 95% quantile-based CI: 1.67-2.5). High between-study heterogeneity is an expected and common finding for nonexperimental study-based meta-analyses (Turner et al 2015). The between-study heterogeneity apparent in our study might be considerably reduced if we had included covariate interaction terms with say the weighting designs in our models but such interactions were precluded due to unbalanced covariate or effect modifier combinations (Phillippo et al. 2020) — for instance, not all 7 designs were deployed across all 5 ocean regions and so precluded an ocean*design interaction term.

3.5. Checking for (In)Consistency

The global consistency test compared the predictive accuracy of an unrelated mean effects or UME model (Spinelli 2022) with our best-fit seabird bycatch meta-regression or consistency reference model (see Figure 2). The LOOcv and Bayesian stacking metrics suggested that both models had similar predictive accuracy (`looic` difference = -0.06, se = 2.6; ca 51% weight of evidence in favour of the best-fit model compared to the UME model). This implies that there was no evidence of global inconsistency and hence our network meta-regression model was fit for inference. Local consistency was evaluated here using the network node-splitting approach and compared the direct and indirect evidence of all pairwise contrasts (node-splits) with sufficient data (15 of the 21 possible contrasts). The direct and indirect posterior density distributions for those 15 node-splits are summarised in Figure S4, which shows no evidence of inconsistency. There was no evidence for network model inconsistency based on either the global or local tests.

3.6. Efficacy Ranking

The median posterior rank for each weighting design based on the posterior samples sourced from our best-fit meta-regression model is summarised in Figure 3. Importantly, there was > 97% probability that all weighting designs significantly reduced seabird bycatch compared to reference design A with no weight attached within 5m of the hook. Nonetheless, some weighting designs were significantly more effective at reducing seabird bycatch risk. Designs E and F rank the “best”, although there is considerable uncertainty associated with design F due to small sample size, which is a concern for robust treatment rank scoring (Daly et al 208, Ades et al 2024). An alternative ranking approach uses the model-based cumulative rank order probabilities based on the SUCRA curve shown in Figure S5. The posterior ranking plot (Figure 3) and the SUCRA-based ranking plot (Figure S5) show very similar expected weighting design rankings — the only difference being a re-order between designs C and B that are in close proximity in rankings and supports the conclusion that the overall rank ordering is robust (Daly et al 2018). Overall, the two best performing designs (E and F) were found to reduce expected seabird bycatch by *ca.* 89% relative to the reference design: E v A ~ 88.8% (95% HDI: 72-99) and F v A ~ 88.9% (95% HDI: 63-99). The next best performing design D reduced expected seabird bycatch by *ca.* 76% relative to the reference design: D v A ~ 76.1% (95% HDI: 43-97).

3.7. Relative Effect Contrasts

The relative design-specific effects for all 21 pairwise contrasts between all 7 branchline weighting designs based on the best-fit Bayesian network meta-regression model are summarised in Figure 4 — the first six contrasts repeated the main effects summarised in Figure 2. There was a > 98% probability that all 5 RFMO and ACAP prescribed weighting designs (B-F) had lower expected seabird bycatch rates relative to the reference design A. Even design G (not a prescribed RFMOs or ACAP design) was estimated to reduce the expected seabird bycatch rate by ca 60% (95% HDI: 7-96) compared to the reference design. The contrast estimates in Figure 4 are especially useful for considering design effects that might be more cost-effective or more practical in terms of crew safety when deploying pelagic longline fishing gear. Using the E vs B contrast as an example, there is a > 97% probability that design E is more effective at reducing seabird bycatch than design B. Or there is a > 95% probability that design E is more effective at reducing seabird bycatch than either designs B or C. On the other hand, there is a > 97% probability that design G is more likely to increase seabird bycatch compared to design E or a > 94% probability that G is more likely to increase seabird bycatch compared to design F.

4. DISCUSSION

4.1. Relative Efficacies of Alternative Weighting Designs

Unlike conventional pairwise meta-analysis, NMA enables the simultaneous comparison of multiple treatments or interventions within a coherent modelling framework (Ades et al 2024). NMA approaches achieve this by leveraging direct within-study evidence and indirect across-studies evidence for all treatments linked in a connected network structure — even if some treatments have not been tested in direct or head-to-head comparisons. The foremost product of any NMA is a league table with rank-ordering of all the treatments to help identify for a specific outcome the relative efficacies of treatments within that network.

Our Bayesian multilevel NMA modelling approach synthesized all available evidence to assess the relative efficacies of seven branchline weighting designs for mitigating seabird bycatch in pelagic longline fisheries. The resulting league table identifies the most effective designs. Six treatments had significantly lower expected seabird bycatch rates than the reference design with no weight within 5m of the hook, consistent with the expectation that attaching weights far from the hook results in relatively slow baited hook sink rates, which increases seabird bycatch risk (Brothers et al., 1999; ACAP, 2023). The five RFMO and ACAP prescribed designs had lower expected seabird bycatch rates relative to the reference design, ranging from *ca.* 66% (design B, 95% HDI:21-95) to *ca.* 89% (design F, 95% HDI: 63-99) lower expected seabird bycatch rates. There was a > 98% probability that all 5 RFMO and ACAP prescribed weighting designs had lower expected seabird bycatch rates relative to the reference design. The one assessed design not prescribed by RFMOs or ACAP was estimated to reduce the expected seabird bycatch rate by *ca.* 60% (95% HDI: 7-96) compared to the reference design, and we can be > 97% sure that design G has a lower seabird bycatch rate than the reference design.

RFMOs and ACAP prescribe multiple options for branchline weighting designs based on the hypothesis that the alternative designs produce similar seabird catch rates. This is an implicit hypothesis of RFMO seabird measures (IATTC, 2011; ICCAT, 2011; WCPFC, 2018; SIOFA, 2022; IOTC, 2023) and explicit in the ACAP (2023) recommendations. Some of the line weighting designs included in RFMO seabird measures may have been derived from member delegations who ensured that contemporary practices of their fisheries were included as options (e.g., WCPFC, 2006). The ACAP (2023) recommended line weighting design options were defined based on mechanistic studies that identified weighting designs that produce hook sink rates that were predicted to adequately protect seabirds when used in combination with a tori line, and based on findings from three experiments on branchline weighting design effects on seabird interaction rates (ACAP, 2023). Our study findings do not support this hypothesis, as we found that some designs prescribed by RFMOs and ACAP have significantly different seabird catch risks. Of 10 possible pairwise contrasts of the 5 prescribed weighting designs, 4 pairs (E vs. B, E vs. C, F vs. B, and F vs. C) had >92% probabilities of having significantly different seabird bycatch rates. The non-prescribed weighting design included in the NMA models had similar efficacy as two of the prescribed designs. Following the reference treatment, these 3 designs had lowest posterior ranks, and had significantly higher seabird catch rates relative to the two designs with lowest seabird catch rates.

One weighting design prescribed by all 5 RFMOs with binding seabird management measures (Table S1) of ≥98g attached between >3.5m and ≤4m of the hook had too small a sample size for inclusion in the meta-analysis and has an unknown relative efficacy. Discussed in more detail below, the study was limited due to small sample sizes for some prescribed weighting designs by region and number of fisheries. With more certain estimates of the efficacies of alternative prescribed weighting designs, those with relatively low predicted efficacies could be eliminated to augment seabird conservation. Or, for two RFMOs with seabird measures with two lists of methods hypothesized to be relatively more and less effective (IATTC, 2011; WCPFC, 2018), weighting designs could be split into these two lists based on their predicted relative efficacies.

Figure 3 (see also Figure S4) shows the estimated relative efficacies of the 7 designs based on an outcome criterion measured as seabird bycatch. Other key performance criteria to holistically evaluate alternative strategies' simulated outcomes across the multiple populations and stocks of exposed commercial and at-risk bycatch species are described in Supplemental

Material Section S8. The design-specific contrasts shown in Figure 4 can be evaluated in association with these other management-focused performance criteria. Moreover, the confidence in any guidelines or recommendations based on such branchline weighting design-specific contrasts can then be assessed quantitatively within a Bayesian NMA framework using techniques such as threshold analysis (Phillippo et al 2019). Threshold analysis quantifies how much the evidence would need to change before a management recommendation would be revised, and an assessment could then be made to predict whether the evidence could plausibly change to a degree that exceeds this threshold (Phillippo et al 2019).

4.2. Effects of Other Seabird Bycatch Mitigation Methods and Ocean Region

The apparent differences in relative seabird bycatch rates of some of the five bycatch mitigation methods included as model covariates might be partly explained by the effect of crew behavior on the efficacy of some methods. Most (90%) of records in the assembled dataset fit to the NMA models were derived from observer program datasets. Efficacies of tori lines, blue-dyed bait and side setting can be strongly affected by crew behavior. As a result, they might be less effective in practice, during real-world, commercial fishing operations, as determined through analyses of observer and electronic monitoring data, than findings from experiments, despite the latter having a relatively lower risk of error and bias (Gilman and Chaloupka, 2023). Unlike in experimental studies, observational studies do not experimentally manipulate specific variables and control for others, and unlike in observational studies that apply quasi-experimental statistical modelling approaches, observational studies with nominal estimates do not standardize effort and do not explicitly account for simultaneous variability in potentially informative predictors of a response (Cox et al. 2007; Luján and Todt 2021). Contrary to ACAP (2023) criteria for identifying ‘best practice’ seabird bycatch mitigation methods, this is why it is important to validate that the efficacy of an intervention when used under controlled conditions is of similar effectiveness when employed in real-world conditions through observational and ‘pragmatic’ studies (Khorsan and Crawford 2014; Pullin et al. 2021). Variability in tori line design components that affect efficacy, including whether single or paired tori lines were employed (Sato et al., 2013), may also have contributed to the relatively low efficacy of this mitigation method. The very small sample size for HookPods (9 records, see Figure S3) precludes any meaningful assessment of relative efficacy. Furthermore, there may have been confounding effects of other operational as well as environmental predictors of seabird bycatch risk that effected the apparent relative efficacies of these 5 seabird bycatch mitigation methods – for example, all HookPod records also had simultaneous use of night setting. Or, for example, crew might employ a particular seabird bycatch mitigation method, such as a tori line or blue-dyed bait, in areas and periods where high seabird bycatch rates are expected to occur, or in response to observing high seabird interactions or high seabird density.

Due to the relationship between seabird density and bycatch risk, the apparent regional differences in relative seabird bycatch risk might be due to regional differences in the local abundance of exposed species of seabirds that are susceptible to pelagic longline capture, and thus regional differences in the magnitude of seabirds scavenging from longline vessels (Gilman et al., 2005). The density of seabirds attending vessels might also affect their scavenging behavior, where up to some threshold, the larger the local seabird abundance, the more intense competitive scavenging behavior and risk of capture will be (Gilman et al., 2005). Regional variability in the composition of seabird species complexes attending vessels might also partly explain this finding. The effect of hierarchical competitiveness between seabird species and between individuals within species, the diving behavior and capabilities of individual species in a complex attending vessels, as well as the occurrence and frequency of secondary interactions, can affect seabird bycatch risk (Jimenez et al., 2012; Melvin et al., 2014). There were also likely regional differences in a wide range of explanatory predictors of seabird bycatch risk that were not able to be accounted for in our NMA, including operational characteristics of fishing methods and gear designs (including regional differences in the use of other seabird bycatch mitigation methods), environmental variables, seabird density and species composition of the assemblage attending vessels, and the spatiotemporal distribution of fishing effort.

4.3. Study Limitations

NMA is a powerful statistical procedure to assess multiple treatment effects within a coherent modelling framework (Ades et al 2024). Nonetheless it is important to acknowledge the following limitations when interpreting our network meta-regression model findings. We used the arm-based modelling approach as opposed to using treatment differences in a contrast-based approach (Chu et al 2024) — the arm-based approach is far more flexible and it also supports estimation of pairwise contrast effects (Ades et al 2024, Chu et al 2024). However, one ongoing conceptual concern with arm-based approaches to NMA is that randomisation integrity within trials or studies might not be preserved in the network and that this might bias estimated relative treatment effects — known as the principle of concurrent control (Chu et al 2024) and related to the pattern of missing treatment combinations within a study or trial since not all the trials include all treatments as shown for example in Figure S2. This is an area of ongoing statistical research.

Importantly, our study is based on study-specific aggregate or summarised data and not set-specific data from each study. Consequently, all the study-specific covariates were also marginal covariate summaries. This can lead to covariate imbalance across the 7 weight designs and potentially increase the risk of bias in our model-based inference. This is also an area of ongoing statistical research.

Moreover, while comprehensive, our study is nonetheless based on a small sample of studies from a small number of fisheries. Sample size limitations prevented us from including various interaction terms in our network model such as ocean region-specific weighting design effects, which might well be important. Furthermore, our NMA model could be mis-specified if the branchline weighting designs were correlated with other potentially informative predictors of seabird bycatch risk that were not available for inclusion in our study. Examples of such potentially informative covariates include: seabird density attending vessels, species composition of the exposed seabird complex attending the fishing vessels, seabird species functional groups such as shallow or deep-diving species, environmental conditions, and other operational characteristics of fishing methods and gear designs (Brothers et al., 1999; Jimenez et al., 2014; Soriano-Redondo et al., 2016; Bi et al., 2020). Designs E and G had broad definitions that could include a wide range of combinations of weight amounts and leader lengths that could have large ranges of efficacies at reducing seabird bycatch risk — with larger sample sizes these categories could be split into narrower design definitions.

4.4. Conclusions

Bycatch in pelagic longline fisheries is a major threat to some populations of pelagic seabirds, particularly albatrosses and petrels (Anderson et al. 2011; Phillips et al., 2016). Managing longline weighting designs is one approach for seabird bycatch mitigation that also supports flexible compliance monitoring. Using a Bayesian multilevel network meta-regression modelling approach, our study synthesized available evidence to assess the relative efficacies of alternative branchline weighting designs, including those prescribed by RFMOs and ACAP. All six assessed weighting designs were significantly more effective at reducing seabird bycatch than a reference design of no weight within 5m of the hook, ranging from reductions of *ca.* 60% (95% HDI: 7-96) to *ca.* 89% (95% HDI: 63-99) in the estimated seabird bycatch rates. There was a >97% probability that all weighting designs significantly reduced seabird bycatch compared to the reference design. But some RFMO and ACAP prescribed designs had significantly different relative seabird catch rates. To augment seabird conservation, weighting designs with relatively low or uncertain predicted efficacies could be assigned lower preferences or eliminated. When combined with additional key inputs for bycatch management strategy evaluation (Supplemental Material Section S8), these estimates of the relative efficacies of alternative branchline weighting designs support the robust evaluation of alternative bycatch management strategies for regional pelagic longline fisheries.

ACKNOWLEDGEMENTS

We thank Mr. Ding-Rong Lin of the Fisheries Agency of Taiwan for authorizing access to observer program data and processing a data query. We are grateful for the assistance of Gary Decossas and Eric Forney for processing observer program data queries for U.S. pelagic longline fisheries operating in the Atlantic and Pacific Oceans, respectively. We thank the New Zealand government for providing access to observer program data. We are grateful to Noriyosi Sato of Tokai University for providing additional information on the findings of Sato et al. (2016). We thank Yu-Min Yeh and Dimas Gianuca for authorizing our referencing their restricted ACAP documents. This study was made possible through funding from the National Fish and Wildlife Foundation Pacific Seabirds Program and the U.S. National Marine Fisheries Service Bycatch Reduction Engineering Program.

Supporting Information

This article includes online Supplemental Material.

Conflict of Interest Statement

The authors declare that they have no competing financial interests or personal relationships that influenced the work reported in this article.

Data Availability Statement

Data from journal articles and grey literature materials used in this study are available in Supplemental Material Table S2 and the original publications. Table S2 excludes records derived from national observer programs included in the study for pelagic longline fisheries of New Zealand, Taiwan, U.S. Atlantic Ocean, U.S. Pacific Ocean shallow-set swordfish fishery, and U.S. Pacific Ocean deep-set tuna fishery because these records are subject to confidentiality restrictions that prevent the authors from making the observer data publicly available.

5. REFERENCES

ACAP. 2023. ACAP Review of Mitigation Measures and Best Practice Advice for Reducing the Impact of Pelagic Longline Fisheries on Seabirds. Agreement on the Conservation of Albatrosses and Petrels, Hobart, Australia.

Ades, A., Welton, N., Dias, S., Phillippo, D., & Caldwell, D. 2024. Twenty years of network meta-analysis: Continuing controversies and recent developments. *Research Synthesis Methods* (early view): 1-26

Anderson, O. et al. 2011. Global seabird bycatch in longline fisheries. *Endangered Species Research, 14*, 91-106.

Bakker, V. et al. 2018. The albatross of assessing and managing risk for long-lived pelagic seabirds. Biol. Conserv. 217, 83–95.

Barrington, J.H.S., Robertson, G. Candy S.G., 2016. Categorising branch line weighting for pelagic longline fishing according to sink rates. SBWG7 Doc 07. Agreement on the Conservation of Albatrosses and Petrels, Hobart, Australia.

Bentley, L., Kato, A., Ropert-Coudert, Y., Manica, A. & Phillips, R. 2021. Diving behaviour of albatrosses: Implications for foraging ecology and bycatch susceptibility. *Mar. Biol.* **168**, 36.

Betancourt M (2017) The QR decomposition for regression models. https://mc-stan.org/users/documentation/case-studies/qr_regression.html

Beverly S, Chapman L, Sokimi W (2003) Horizontal Longline Fishing Methods and Techniques. A Manual for Fishermen. Secretariat of the Pacific Community, Noumea, New Caledonia.

Bi R, Jiao Y, Bakka H, Browder J. 2020. Long-term climate ocean oscillations inform seabird bycatch from pelagic longline fishery. ICES Journal of Marin Science 77: 668-679.

Brothers, N., Cooper, J. & Lokkeberg, S. *The Incidental Catch of Seabirds by Longline Fisheries: Worldwide Review and Technical Guidelines for Mitigation*. FAO Fisheries Circular No. 937. (Food and Agriculture Organization of the United Nations, 1999)

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: a probabilistic programming language. Journal of Statistical Software 76(1): 1-32

Chu H, Lin L, Wang Z, Wang Z, Chen Y, Cappelleri J (2024) A review and comparison of arm-based versus contrast-based network meta-analysis for binary outcomes — Understanding their differences and limitations. WIREs Computational Statistics 16: e1639

Conway J, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33: 2938–2940

Cox T, Lewison R, Zydelis R, Crowder L, Safina C, Read A (2007) Comparing effectiveness of experimental and implemented bycatch reduction measures: the ideal and the real. Conser Biol 21:1155–1164. doi:10.1111/j.1523-1739.2007.00772.x

Currier B, Mcleod J, Banfield L, Bayene J, Welton N, D'Souza A et al (2023) Resistance training prescription for muscle strength and hypertrophy in healthy adults: a systematic review and Bayesian network meta-analysis. British Journal of Sports Medicine 57: 1211-1220

Daly C, Neupane B, Beyene J, Thabane L, E Straus S, Hamid J (2019) Empirical evaluation of SUCRA-based treatment ranks in network meta-analysis: quantifying robustness using Cohen's kappa. BMJ Open 9: e024625

Dias, M., Martin, R., Pearmain, E., et al. Threats to seabirds: A global assessment. Biol. Conserv. 237, 525–537 (2019).

Dias S, Welton N, Sutton A, Ades A (2011) NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. National Institute for Health and Care Excellence. https://www.sheffield.ac.uk/nice-dsu

Dicks L, Hodge I, Randall N, Scharlemann J, et al (2014) A transparent process for "evidence-informed" policy making. Conserv Lett 7:119–125. doi:10.1111/conl.12046

Frankish, C., Manica, A., Navarro, J. & Phillips, R. 2021. Movements and diving behaviour of white-chinned petrels: Diurnal variation and implications for bycatch mitigation. *Aquat. Conserv.* **31**, 1715–1729

Freeman S, Saeedi E, Ordóñez-Mena J, Nevill C, Hartmann-Boyce J, Caldwell D, Welton N, Cooper N, Sutton A (2023) Data visualisation approaches for component network meta-analysis: visualising the data structure. BMC Medical Research Methodology 23: 208

Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A (2019) Visualization in Bayesian workflow. Journal of the Royal Society Series A 182 (1): 1-14

Gianuca, D., Peppes, F., César, J., Marques, C., Neves, T., 2011. The effect of leaded swivel position and light toriline on bird attack rates in Brazilian pelagic longline. SBWG-4 Doc 40 Rev1. Agreement on the Conservation of Albatrosses and Petrels, Hobart, Australia.

Gilman, E., Beverly, S., Musyl, M. & Chaloupka, M. Commercial viability of alternative designs placing pelagic longline branchline weights at the hook to reduce seabird bycatch. *Endanger. Species Res.* **43**, 223-233 (2020)

Gilman, E., Brothers, N. & Kobayashi, D. 2005. Principles and approaches to abate seabird bycatch in longline fisheries. *Fish Fish.* **6**, 35–49.

Gilman, E., Chaloupka, M. 2023. Applying a sequential evidence hierarchy, with caveats, to support prudent fisheries bycatch policy. Reviews in Fish Biology and Fisheries doi: 10.1007/s11160-022-09745-4.

Gilman, E., Chaloupka, M., Wiedoff, B., Willson, J. 2014b. Mitigating seabird bycatch during hauling by pelagic longline vessels. *PLOS ONE* 9: e84499. doi:10.1371/journal.pone.0084499.

Gilman E, Legorburu G, Fedoruk A, Heberer C, Zimring M, Barkai A (2019) Increasing the functionalities and accuracy of fisheries electronic monitoring systems. *Aquatic Conservation: Marine and Freshwater Ecosystems* 29: 901-926

Gilman E, Passfield K, Nakamura K (2014a) Performance of regional fisheries management organizations: ecosystem-based governance of bycatch and discards. Fish Fish 15:327-351

Guilford, T., Padget, O., Maurice, L. & Catry, P. 2022. Unexpectedly deep diving in an albatross. *Curr. Biol.* **32**, R26–R28

Haddaway, N., Macura, B., Whaley, P., Pullin, A. 2018. ROSES RepOrting standards for Systematic Evidence Syntheses: *pro forma*, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ. Evid.* **7**, doi: 10.1186/s13750-018-0121-7.

Hall, M., Gilman, E., Minami, H., Mituhasi, T., Carruthers, E. 2017. Mitigating bycatch in tuna fisheries. *Reviews in Fish Biology and Fisheries.* 27: 881-908.

Heppell, S.S., Caswell, H., Crowder, L.B., 2000. Life histories and elasticity patterns: perturbation analysis for species with minimal demographic data. Ecology 81, 654–665

Hooper L, Summerbell C, Higgins J, Thompson R, Clements G, Capps N, Davey Smith G, Riemersma R, Ebrahim S (2000) Reduced or modified dietary fat for preventing cardiovascular disease. Cochrane Database of Systematic Reviews 2. https://doi.org/10.1002/14651858.CD002137

Hutton B, Salanti G, Caldwell D, Chaimani A, Schmid C et al (2015) The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Annals of Internal Medicine 162: 777–784

IATTC. 2011. Resolution to Mitigate the Impact on Seabirds of Fishing for Species Covered by the IATTC. Resolution C-11-02. Inter-American Tropical Tuna Commission, La Jolla, USA.

ICCAT. 2007. Recommendation by ICCAT on Reducing Incidental By-catch of Seabirds in Longline Fisheries. Recommendation 07-07. International Commission for the Conservation of Atlantic Tunas, Madrid.

ICCAT. 2011. Supplemental Recommendation by ICCAT on Reducing Incidental By-catch of Seabirds in ICCAT Longline Fisheries. Recommendation 11-09. International Commission for the Conservation of Atlantic Tunas, Madrid.

IOTC. 2023. On Reducing the Incidental Bycatch of Seabirds in Longline Fisheries. Resolution 23/07. Indian Ocean Tuna Commission, Mahe, Seychelles.

Jimenez, S., Domingo, A., Abreu, M. & Brazeiro, A. 2012. Bycatch susceptibility in pelagic longline fisheries: Are albatrosses affected by the diving behaviour of medium-sized petrels?. *Aquat. Conserv.* **22**, 436–445

Jimenez, S., Domingo, A., Forselledo, R., Sullivan, B. & Yates, O. 2019. Mitigating bycatch of threatened seabirds: The effectiveness of branch line weighting in pelagic longline fisheries. *Anim. Conserv.* **22**, 376–385

Jimenez, S., Phillips, R.A., Brazeiro, A., Defeo, O., Domingo, A. 2014. Bycatch of great albatrosses in pelagic longline fisheries in the southwest Atlantic: contributing factors and implications for management. Biological Conservation 171: 9–20. doi 10.1016/j.biocon.2013.12.035

Juan-Jorda, M., Murua, H., Arrizabalga, H., Dulvy, N., Restrepo, V. 2018. Report card on ecosystem-based fisheries management in tuna regional fisheries management organizations. Fish and Fisheries 19: 321-339.

Kappes M, Shaffer S, Tremblay Y, Foley D, Palacios D, Bograd S, Costa D (2015) Reproductive constraints influence habitat accessibility, segregation, and preference of sympatric albatross species. Movement Ecology 3: 34

Kay M (2023) tidybayes: Tidy data and geoms for Bayesian models. R package version 3.0.6. http://mjskay.github.io/tidybayes. https://doi.org/10.5281/zenodo.1308151

Kay M (2024) ggdist: Visualizations of distributions and uncertainty in the Grammar of Graphics. IEEE Transactions on Visualization and Computer Graphics 30: 414-424

Kazama, K., Harada, T., Deguchi, T., Suzuki, H. & Watanuki, Y. 2019, Foraging behavior of black-footed albatross Phoebastria nigripes rearing chicks on the Ogasawara Islands. Ornithol. Sci. 18, 27–37.

Khorsan R, Crawford C (2014) External validity and model validity: A conceptual approach for systematic review methodology. Evid-based Complement Altern Med 2014:1-12. doi:10.1155/2014/694804

Krassowski M (2020) krassowski/complex-upset. http://doi.org/10.5281/zenodo.3700590

Kruschke J, Liddell T (2018) The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review 25: 178–206

Lajeunesse M (2011) On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology 92: 2049-2055

Lemoine N (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos 128: 912-928

Liu Y, Béliveau A, Wei Y, Chen M, Record-Lemon R, Kuo P, Pritchard E, Tang X, Chen G (2023) A gentle introduction to Bayesian network meta-analysis using an automated R package. Multivariate Behavioral Research 58: 706–722

Luján J, Todt O (2021) Evidence based methodology: a naturalistic analysis of epistemic policies in regulatory science. Eur J Philos Sci 11:26. doi:10.1007/s13194-020-00340-7

Melvin, E., Guy, T. & Read, L. 2013. Reducing seabird bycatch in the South African tuna fishery using bird-scaring lines, branch line weighting and nighttime setting of hooks. *Fish. Res.* **147**, 72–82.

Melvin, E., Guy, T. & Read, L. 2014. Best practice seabird bycatch mitigation for pelagic longline fisheries targeting tuna and related species. *Fish. Res.* **149**, 5–18

Miyake M, Guillotreau P, Sun C, Ishimura G. 2010. Recent Developments in the Tuna Industry. Stocks, Fisheries, Management, Processing, Trade and Markets. FAO Fisheries and Aquaculture Technical Paper 543. Food and Agriculture Organization of the United Nations, Rome.

Nichols J, Kendall W, Boomer G (2019) Accumulating evidence in ecology: once is not enough. Ecol Evol 9:13991-14004. doi:10.1002/ece3.5836

Pappalardo P, Ogle K, Hamman E, Bence J, Hungate B, Osenberg C (2020) Comparing traditional and Bayesian approaches to ecological meta-analysis. Methods in Ecology and Evolution 11: 1286– 1295

Pedersen T (2020) patchwork: The composer of plots. R package version 1.1.0. https://CRAN.R-project.org/package=patchwork

Phillippo D (2024) multinma: Bayesian network meta-analysis of individual and aggregate data. doi:10.5281/zenodo.3904454. https://doi.org/10.5281/zenodo.3904454. R package version 0.6.1. https://dmphillippo.github.io/multinma

Phillippo D, Dias S, Ades A, Belger M, Brnabic A, Schacht A, Saure D, Kadziola Z, Welton N (2020) Multilevel network meta-regression for population-adjusted treatment comparisons. Journal of the Royal Statistical Society: Series A (Statistics in Society) 183: 1189-1210

Phillippo D, Dias S, Welton N, Caldwell D, Taske N, Ades A (2019) Threshold analysis as an alternative to GRADE for assessing confidence in guideline recommendations based on network meta-analyses. Annals of Internal Medicine 170: 538-546

Phillips R, Gales R, Baker G, Double M, et al (2016). The conservation status and priorities for albatrosses and large petrels. Biol Conserv 201:169–183

Phillips R, Wood A. 2020. Variation in live-capture rates of albatrosses and petrels in fisheries, post-release survival and implications for management. Biological Conservation 247: 108641.

Prince, P., Huin, N. & Weimerskirch, H. 1994. Diving depths of albatrosses. Antarct. Sci. 6, 353–354.

Pullin, A., Frampton, G., Livoreil, B. & Petrokofsky, G., Eds. 2020. Section 5. Conducting a Search. Key CEE Standards for Conduct and Reporting. In Pullin, A., Frampton, G., Livoreil, B., Petrokofsky, G., Eds. *Guidelines and Standards for Evidence Synthesis in Environmental Management*. Version 5.0. Collaboration for Environmental Evidence.

Pullin, A., Frampton, G., Livoreil, B. & Petrokofsky, G., Eds. 2021. Section 3. Planning a CEE Evidence Synthesis. In Pullin, A., Frampton, G., Livoreil, B., Petrokofsky, G., Eds. *Guidelines and Standards for Evidence Synthesis in Environmental Management*. Version 5.0. Collaboration for Environmental Evidence.

Robertson, G., Candy, S. & Hall, S. 2013. New branch line weighting regimes to reduce the risk of seabird mortality in pelagic longline fisheries without affecting fish catch. *Aquat. Conserv.* 23: 885–900

Robertson, G., Candy, S., Wienecke, B. & Lawton, K. 2010. Experimental determinations of factors affecting the sink rates of baited hooks to minimize seabird mortality in pelagic longline fisheries. *Aquat. Conserv.* **20**, 632–643.

Rollinson D, Dilley B, Davies D, Ryan P. 2016. Diving behaviour of Grey Petrels and its relevance for mitigating long-line by-catch. Emu doi: 10.1071/MU15032

Santos, R., Silva-Costa, A., Sant'Ana, R, Gianuca, D., Yates, O., Marques, C., Neves, T. 2019. Improved line weighting reduces seabird bycatch without affecting fish catch in the Brazilian pelagic longline fishery. *Aquat. Conserv.* **29**, 442–449

Sato N, Minami H, Katsumata N, Ochi D, Yokawa K. 2013. Comparison of the effectiveness of paired and single tori lines for preventing bait attacks by seabirds and their bycatch in pelagic longline fisheries. Fisheries Research 140: 14-19.

Sato N, Katsumata N, Yokota K, Uehara, T., Fusejima, I., Minami, H. (2016) Tori-lines with weighted branch lines reduce seabird bycatch in eastern South Pacific longline fishery. Aquatic Conserv: Mar Freshw Ecosyst 26: 95–107.

Seide S, Jensen K, Kieser M (2020) A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. Research Synthesis Methods 11: 363–378

Singh J, Gsteiger S, Wheaton L, Riley R, Abrams K, Gillies C, Bujkiewicz S (2022) Bayesian network meta-analysis methods for combining individual participant data and aggregate data from single arm trials and randomised controlled trials. BMC Medical Research Methodology 22: 186

SIOFA. 2022. Conservation and Management Measure on Mitigation of Seabirds Bycatch in Demersal and Pelagic Longlines and other Demersal Fishing Gears Fisheries (Mitigation of Seabirds Bycatch). CMM 2022/13. Southern Indian Ocean Fisheries Agreement, Saint-Denis, La Reunion.

Soriano-Redondo A, Cortes V, Reyes-Gonzalez J., et al. (2016) Relative abundance and distribution of fisheries influence risk of seabird bycatch. Scientific Reports doi: 10.1038/srep37373.

Spineli L (2022) A revised framework to evaluate the consistency assumption globally in a network of interventions. Medical Decision Making 42: 637-648

Turner R, Jackson D, Wei, Y, Thompson S, Higgins J (2015) Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Statistics in Medicine 34: 984–998

van de Schoot R, Depaoli S, King R, Kramer B, Märtens K, Tadesse M, Vannucci M, Gelman A, Veen D, Willemsen J, Yau C (2021) Bayesian statistics and modelling. Nature Reviews Methods Primers 1(1): 1-26. https://doi.org/10.1038/s43586-021-00017-2

van der Bles A, van der Linden S, Freeman A, Mitchell J, Galvao A, Zaval L, Spiegelhalter D (2019) Communicating uncertainty about facts, numbers and science. Royal Society Open Science 6: 181870

Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27: 1413–1432

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P (2021) Rank-normalization, folding, and localization: An improved Rhat for assessing convergence of MCMC (with Discussion). Bayesian Analysis 16: 667–718

Veroniki A, Straus S, Rücker G, Tricco A (2018) Is providing uncertainty intervals in treatment ranking helpful in a network meta-analysis? Journal of Clinical Epidemiology 100:122-129

Veroniki A, Vasiliadis H, Higgins J, Salanti G (2013) Evaluation of inconsistency in networks of interventions. International Journal of Epidemiology 42: 332-345

Wang X, Ryan Y, Faraway J (2018) Bayesian regression modeling with INLA. CRC Press, USA.

WCPFC. 2006. Report and Information Paper of the Voluntary Small Working Group on Seabird Bycatch Mitigation. WCPFC3-2006/IP03. Western and Central Pacific Fisheries Commission, Kolonia, Federated States of Micronesia.

WCPFC. 2018. Conservation and Management Measure to Mitigate the Impact of Fishing for Highly Migratory Fish Stocks on Seabirds. CMM 2018-03. Western and Central Pacific Fisheries Commission, Kolonia, Federated States of Micronesia.

Williams, P., Ruaia, T. 2022. Overview of Tuna Fisheries in the Western and Central Pacific Ocean, Including Economic Conditions—2021. WCPFC-SC18-2022/GN IP-1 Western and Central Pacific Fisheries Commission, Kolonia, Federated States of Micronesia.

Veroniki A, Straus S, Rücker G, Tricco A (2018) Is providing uncertainty intervals in treatment ranking helpful in a network meta-analysis? Journal of Clinical Epidemiology 100:122-129

Veroniki A, Vasiliadis H, Higgins J, Salanti G (2013) Evaluation of inconsistency in networks of interventions. International Journal of Epidemiology 42: 332-345

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. 2nd Edition. Springer-Verlag, New York, USA

Yao Y, Vehtari A, Simpson D, Gelman A (2018) Using stacking to average Bayesian predictive distributions (with Discussion). Bayesian Analysis 13: 917–1003

Zeileis A, Fisher J, Hornik K, Ihaka R, McWhite C, Murrell P, Stauffer R, Wilke C (2020) colorspace: A toolbox for manipulating and assessing colors and palettes. Journal of Statistical Software 96: 1-4

Figure Captions

Figure 1. NETWORK GEOMETRY. Diagram showing the 7 (A-G) branchline weighting designs at the nodes connected by 21 direct network interactions or edges between those nodes that were sourced from 18 individual studies. This shows that our network is fully connected and suitable for subsequent network meta-analysis. Node size is proportional to sample size with connecting line (edge) thickness proportional to number of studies in our assembled dataset that included summary data for comparison between any 2 specific designs (nodes). There are three treatment (design) classes of: a reference design A (0 weight attached ≤5m from hook), designs B through F, which are prescribed by RFMOs or ACAP, and G is a non-prescribed weighting design. Definitions of designs are in the Methods section 2.2 *Branchline Weighting Design Treatments*.

Figure 2. DESIGN-SPECIFIC CONDITIONAL EFFECTS. Summary of the predicted branchline weighting design-specific conditional effects relative to the reference design A (no weight within 5m of the hook) based on the best-fit Bayesian network meta-regression model. Coloured polygon shows the posterior density distribution summary (i.e., a wide and thin polygon indicates low precision from a small sample size, and narrow and tall indicates there was high precision and large sample size.), solid dots = median of the posterior density distribution, thick horizontal line below each polygon shows the 80% highest posterior density interval (HDI) for the density distribution while the thin horizontal line is the 95% HDI. Log rate (response) ratios < 0 = decreased seabird bycatch risk. Although a small fraction of 5 posterior density distributions were > 0, (salmon shading) all the 6 contrasts with the reference design (A) were significantly < 0.

Figure 3. DESIGN-SPECIFIC RANK-ORDER. Posterior rank summaries for each of the branchline weighting designs (A-G) based on the best-fit network meta-regression model. Solid dot = posterior median rank, thin horizontal line = 95% credible interval summary of the posterior density distribution for that design-specific ranking simulations.

Figure 4. DESIGN-SPECIFIC CONSTRASTS. Summary of the predicted log rate ratios for all 21 branchline weighting design contrasts based on the best-fit network meta-regression model. Solid dots = conditional median effect estimate, thick horizontal line = 80% credible interval while the thin horizontal line = 95% credible interval. Log rate ratio (response ratio) < 0 = decreased seabird bycatch risk.

Figures

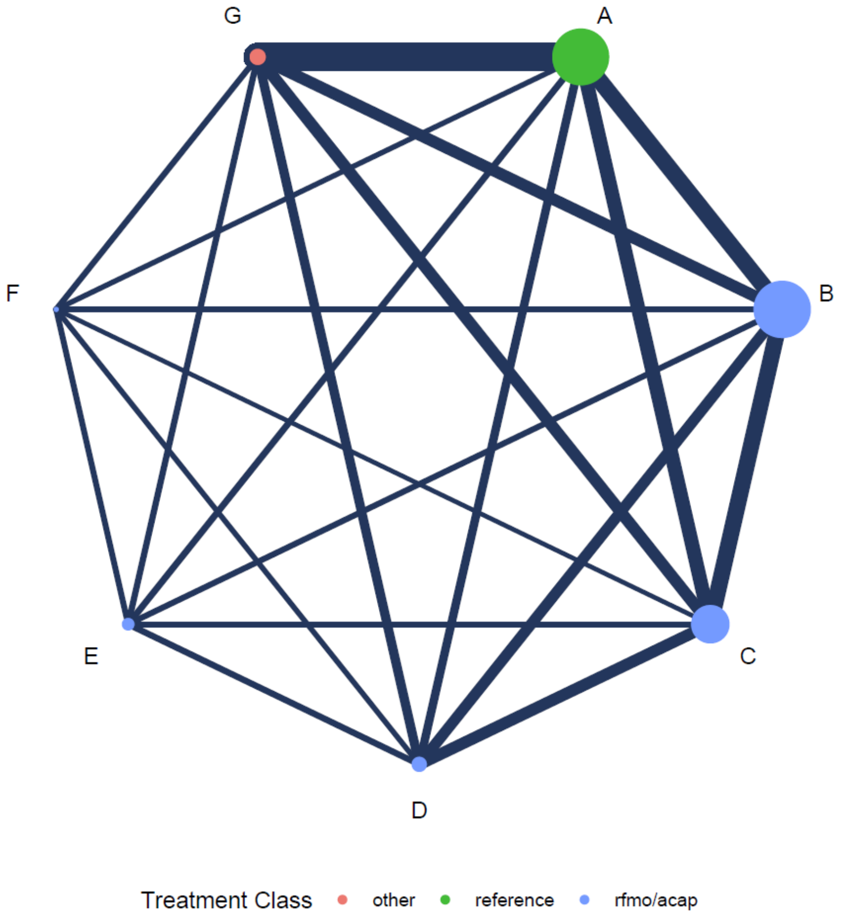

Figure 1

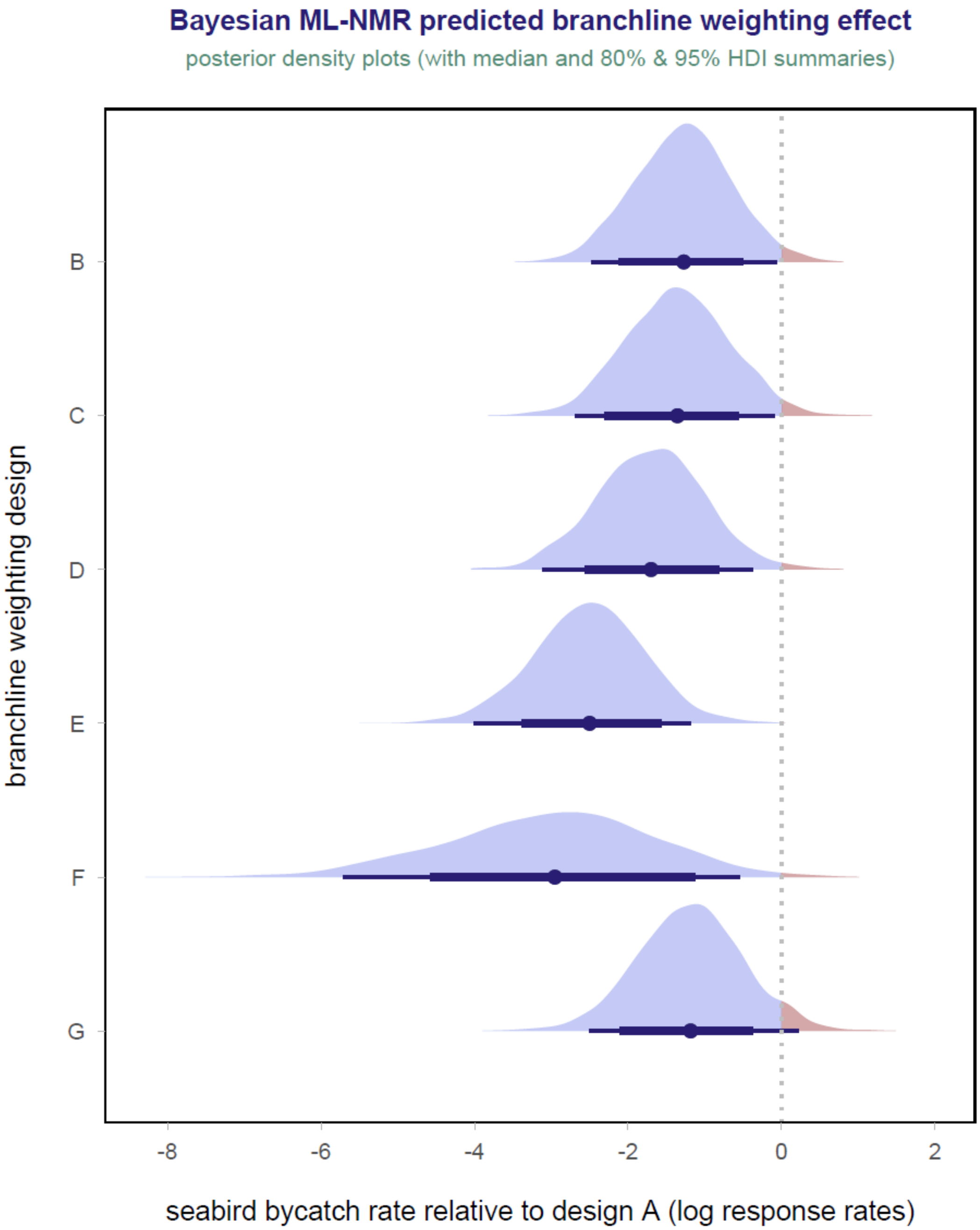
Bayesian ML-NMR predicted branchline weighting effect
posterior density plots (with median and 80% & 95% HDI summaries)
branchline weighting design
B
C
D
E
F
G
-8
-6
-4
-2
0
2
seabird bycatch rate relative to design A (log response rates)

Figure 2

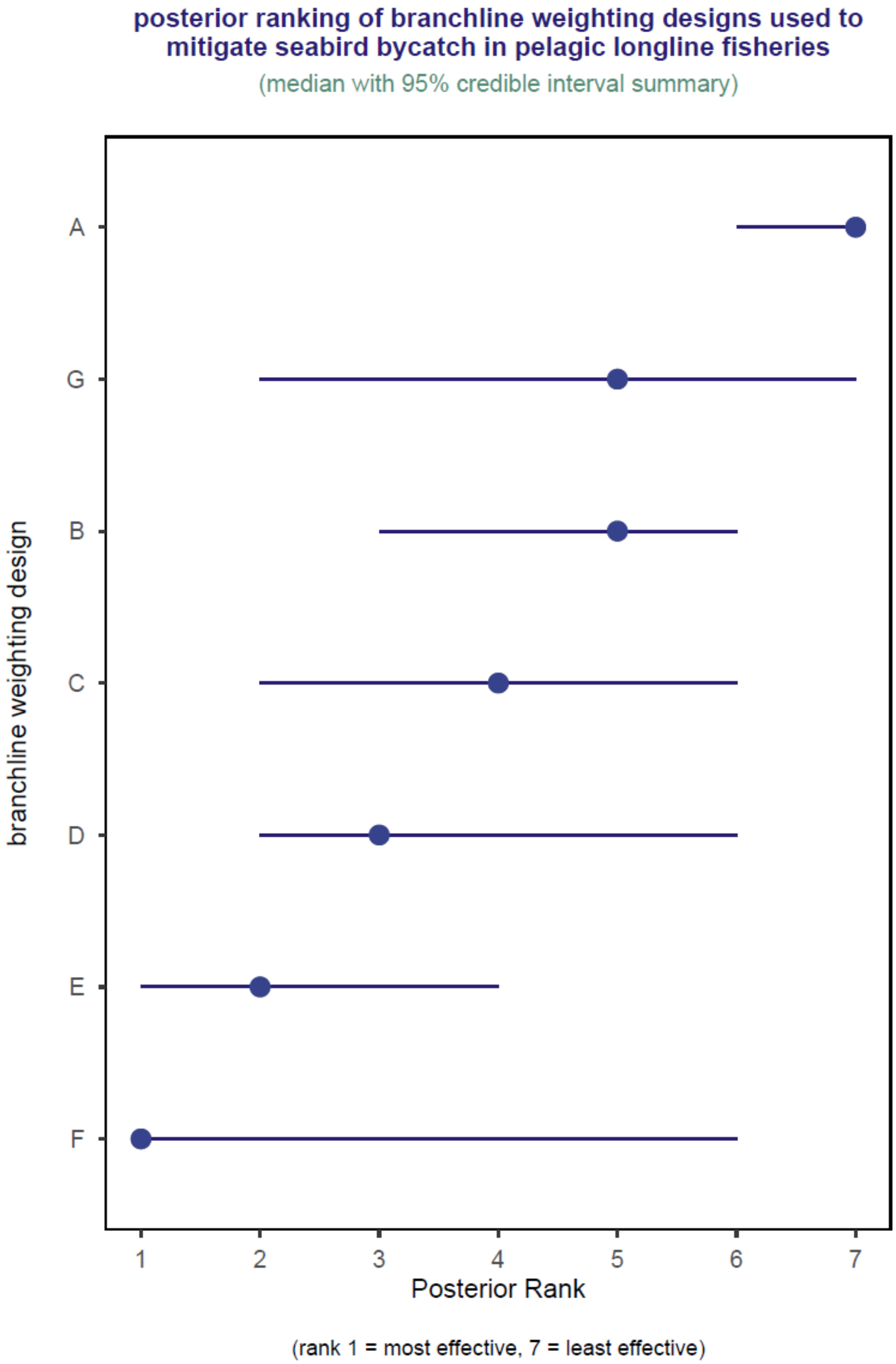
posterior ranking of branchline weighting designs used to mitigate seabird bycatch in pelagic longline fisheries
(median with 95% credible interval summary)
branchline weighting design
A
G
B
C
D
E
F
1
2
3
4
5
6
7
Posterior Rank
(rank 1 = most effective, 7 = least effective)

Figure 3

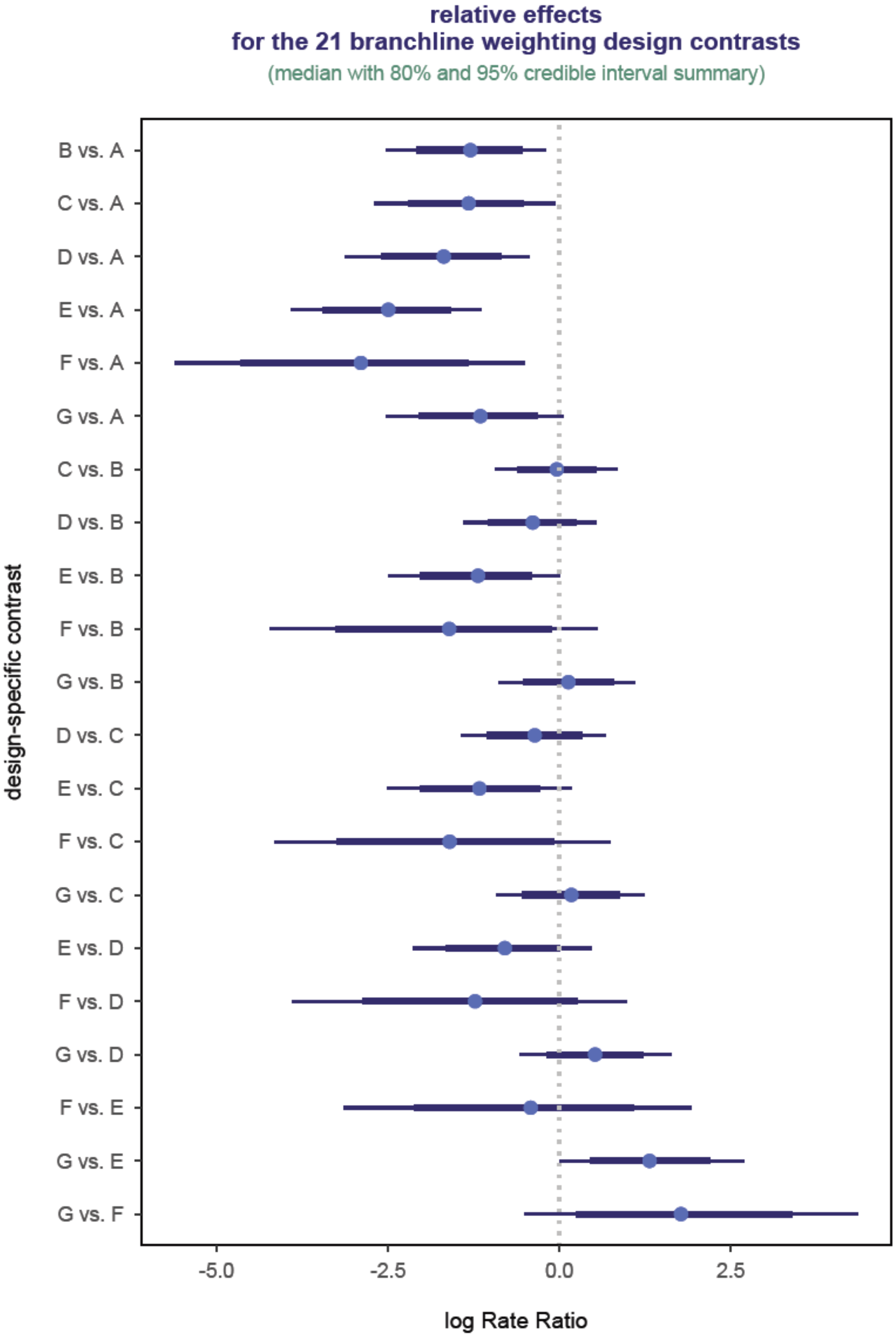
relative effects
for the 21 branchline weighting design contrasts
(median with 80% and 95% credible interval summary)
B vs. A
C vs. A
D vs. A
E vs. A
F vs. A
G vs. A
C vs. B
D vs. B
E vs. B
F vs. B
G vs. B
D vs. C
E vs. C
F vs. C
G vs. C
E vs. D
F vs. D
G vs. D
F vs. E
G vs. E
G vs. F
design-specific contrast
-5.0
-2.5
0.0
2.5
log Rate Ratio

Figure 4