Combining chains of Bayesian models with Markov melding

Andrew A. Manderson[*], **Robert J. B. Goudie**[†]

Andrew A. Manderson: andrew.manderson@mrc-bsu.cam.ac.uk; Robert J. B. Goudie: robert.goudie@mrc-bsu.cam.ac.uk

[*]MRC Biostatistics Unit, University of Cambridge, United Kingdom, and The Alan Turing Institute

[†]MRC Biostatistics Unit, University of Cambridge, United Kingdom

Abstract

A challenge for practitioners of Bayesian inference is specifying a model that incorporates multiple relevant, heterogeneous data sets. It may be easier to instead specify distinct submodels for each source of data, then join the submodels together. We consider chains of submodels, where submodels directly relate to their neighbours via common quantities which may be parameters or deterministic functions thereof. We propose *chained Markov melding*, an extension of Markov melding, a generic method to combine chains of submodels into a joint model. One challenge we address is appropriately capturing the prior dependence between common quantities within a submodel, whilst also reconciling differences in priors for the same common quantity between two adjacent submodels. Estimating the posterior of the resulting overall joint model is also challenging, so we describe a sampler that uses the chain structure to incorporate information contained in the submodels in multiple stages, possibly in parallel. We demonstrate our methodology using two examples. The first example considers an ecological integrated population model, where multiple data sets are required to accurately estimate population immigration and reproduction rates. We also consider a joint longitudinal and time-to-event model with uncertain, submodel-derived event times. Chained Markov melding is a conceptually appealing approach to integrating submodels in these settings.

Keywords

Combining models; Markov melding; Bayesian graphical models; Multi-stage estimation; Model/ data integration; Integrated population model

1   Introduction

The Bayesian philosophy is appealing in part because the posterior distribution quantifies all sources of uncertainty. However, a joint model for all data and parameters is a prerequisite to posterior inference, and in situations where multiple, heterogeneous sources of data are available, specifying such a joint model is a formidable task. Models that consider such data are necessary to describe complex phenomena at a useful precision. One possible approach begins by specifying individual submodels for each source of data. These submodels could guide the statistician when directly specifying the joint model, but to use the submodels

only informally seems wasteful. Instead, it may be preferable to construct a joint model by formally joining the individual submodels together.

Some specific forms of combining data are well established. Meta-analyses and evidence synthesis methods are widely used to summarise data, often using hierarchical models (Ades and Sutton, 2006; Presanis et al., 2014). Outside of the statistical literature, a common name for combining multiple data is *data fusion* (Lahat et al., 2015; Kedem et al., 2017), though there are many distinct methods that fall under this general name. Interest in integrating data is not just methodological; applied researchers often collect multiple disparate data sets, or data of different modalities, and wish to combine them. For example, to estimate SARS-CoV-2 positivity Donnat et al. (2020) build an intricate hierarchical model that integrates both testing data and self-reported questionnaire data, and Parsons et al. (2021) specify a hierarchical model of similar complexity to estimate the number of injecting drug users in Ukraine. Both applications specify Bayesian models with data-specific components, which are united in a hierarchical manner. In conservation ecology, *integrated population models* (IPMs) (Besbeas et al., 2002; Brooks et al., 2004; Schaub and Abadi, 2011; Maunder and Punt, 2013; Zipkin and Saunders, 2018) are used to estimate population dynamics, e.g. reproduction and immigration rates, using multiple data on the same population. Such data have standard models associated with them, such as the Cormack-Jolly-Seber model (Lebreton et al., 1992) for capture-recapture data, and the IPM serves as the framework in which the standard models are combined. More generally, the applications we list illustrate the importance of generic, flexible methods for combining data to applied researchers.

Markov melding (Goudie et al., 2019) is a general statistical methodology for combining submodels. Specifically, it considers M submodels that share some common quantity ϕ, with each of the $m = 1, \ldots, M$ submodels possessing distinct parameters ψ_m, data Y_m, and form $p_m(\phi, \psi_m, Y_m)$. Goudie et al. (2019) then propose to combine the submodels into a joint model, denoted $p_{\text{meld}}(\phi, \psi_1, \ldots, \psi_M, Y_1, \ldots, Y_M)$. However, it is unclear how to integrate models where there is no single quantity ϕ common to all submodels, such as for submodels that are linked in a chain structure.

We propose an extension to Markov melding, which we call *chained Markov melding*[1], which facilitates the combination of M submodels that are in a chain structure. For example, when $M = 3$ we address the case in which submodel 1 and 2 share a common quantity $\phi_{1 \cap 2}$, and submodel 2 and 3 share a different quantity $\phi_{2 \cap 3}$. Our extension addresses previously unconsidered complications including the distinct domains (and possibly supports) of the common quantities, and the desire to capture possible prior correlation between them. Two examples serve to illustrate our methodology, which we introduce in the following section. The computational effort required to fit a complex, multi-response model is a burden to the model development process. We propose a multi-stage posterior estimation method that exploits the properties of our chained melded model to reduce this burden. We can parallelise aspects of the computation across the submodels, using less computationally expensive techniques for some submodels. Reusing existing software implementations of

---

[1] *"Chained graphs"* were considered by Lauritzen and Richardson (2002), however they are unrelated to our proposed model. We use "chained" to emphasise the nature of the relationships between submodels.

submodels, and subposterior samples where available, is also possible. Multi-stage samplers can aid in understanding the contribution of each submodel to the final posterior, and are used in many applied settings, including hierarchical modelling (Lunn et al., 2013) and joint models (Mauff et al., 2020).

One contribution of our work is to clarify the informal process commonly used in applied analyses of summarising and/or approximating submodels for use in subsequent analyses. The two most common approximation strategies seem to be (i) approximating the subposterior of the common quantity with a normal distribution for use in subsequent models (see, e.g. Jackson and White, 2018; Nicholson et al., 2021) and (ii) taking only a point estimate of the subposterior, and treating it as a known value in further models. These strategies may, but not always, produce acceptable approximations to the chained melded model. Both the chained melded model and these approximation strategies are examples of 'multi-phase' and 'multi-source' inference (Meng, 2014), with the melding approach most comprehensively accounting for uncertainty.

1.1 Example introduction

In this section we provide a high-level overview of two applications that require integrating a chain of submodels, with more details in Sections 4 and 5. Our first example decomposes a joint model into its constituent submodels and rejoins them. This simple situation allows us to compare the output from the chained melding process to the complete joint model, and is meant to illustrate both the 'chain-of-submodels' notion and the mechanics of chained melding. The second example is a realistic and complex setting in which the combining of submodels without chained Markov melding is nonobvious. Our comparator is the common technique of summarising previously considered submodels with point estimates, and demonstrates the importance of fully accounting for uncertainty.

An integrated population model for little owls—Integrated population models (IPMs) (Zipkin and Saunders, 2018) combine multiple data to estimate key quantities governing the dynamics of a specific population. Schaub et al. (2006) and Abadi et al. (2010) used an IPM to estimate fecundity, immigration, and yearly survival rates for a population of little owls. These authors collect and model three types of data, illustrated in Figure 1. Capture-recapture data Y_1, and associated capture-recapture submodel $p_1(\phi_{1\cap2}, \psi_1, Y_1)$, are acquired by capturing and tagging owls each year, and then counting the number of tagged individuals recaptured in subsequent years. Population counts Y_2 are obtained by observing the number of occupied nesting sites, and are modelled in $p_2(\phi_{1\cap2}, \phi_{2\cap3}, \psi_2, Y_2)$. Finally, nest-record data Y_3 counts both the number of reproductive successes and possible breading pairs, and is associated with a submodel for fecundity $p_3(\phi_{2\cap3}, \psi_3, Y_3)$. The population count model p_2 shares the parameter $\phi_{1\cap2}$ with the capture-recapture model p_1, and the parameter $\phi_{2\cap3}$ with the fecundity model p_3; each of the $m = 1, 2, 3$ submodels has distinct, submodel-specific parameters ψ_m. No single source of data is sufficient to estimate all quantities of interest, so it is necessary to integrate the three submodels into a single joint model to produce acceptably precise estimates of fecundity and immigration rates. We will show that the chained Markov melding framework developed in

Section 2 encapsulates the process of integrating these submodels, producing results that are concordant with the original joint IPM.

Survival analysis with time varying covariates and uncertain event times—Our second example considers the time to onset of respiratory failure (RF) amongst patients in intensive care units, and factors that influence the onset of RF. A patient can be said to be experiencing RF if the ratio of the partial pressure of arterial blood oxygen (PaO_2) to the faction of inspired oxygen (FiO_2) is less than 300mmHg (The ARDS Definition Task Force, 2012), though this is not the only definition of RF. Patients' PaO_2/FiO_2 (P/F) ratios are typically measured only a few times a day. The relative infrequency of P/F ratio data, when combined with the intrinsic variability in each individual's blood oxygen level, results in significant uncertainty in about the time of onset of RF.

Factors that influence the time to onset of RF are both longitudinal and time invariant. Both types of data can be considered in *joint models* (Rizopoulos, 2012), which are composed of two distinct submodels, one for each data type. However, existing joint models are not able to incorporate the uncertainty surrounding the event time, which may result in overconfident and/or biased estimates of the parameters in the joint model.

Chained Markov melding offers a conceptually straightforward, Bayesian approach to incorporating uncertain event times into joint models. Specifically, we consider the event time as a submodel-derived quantity from a hierarchical regression model akin to Lu and Meeker (1993). We call this submodel the *uncertain event time* submodel and denote it $p_1(\phi_{1 \cap 2}, \psi_1, Y_1)$, where $\phi_{1 \cap 2}$ incorporates the event time. The survival submodel $p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2, Y_2)$ uses the event time within $\phi_{1 \cap 2}$, the common quantity, as the response. We treat the longitudinal submodel, $p_3(\phi_{2 \cap 3}, \psi_3, Y_3)$, separately from the survival submodel, as is common in two-stage joint modelling (Mauff et al., 2020), and denote the subject-specific parameters that also appear in the survival model as $\phi_{2 \cap 3}$. Each of the $m = 1, 2, 3$ has submodel-specific data Y_m and parameters ψ_m. The high level submodel relationships are displayed as a DAG in Figure 2.

It is in examples such as this one that we foresee the most use for chained Markov melding; a fully Bayesian approach is desired and the submodels are nontrivial in complexity, with no previously existing or obvious joint model.

1.2 Markov melding

We now review Markov melding (Goudie et al., 2019) before detailing our proposed extension. As noted in the introduction, Markov melding is a method for combining M submodels $p_1(\phi, \psi_1, Y_1), \ldots, p_M(\phi, \psi_M, Y_M)$ which share the same ϕ. When the sub-model prior marginals $p_m(\phi)$ are identical, i.e. $p_m(\phi) = p(\phi)$ for all m, it is possible to combine the submodels using *Markov combination* (Dawid and Lauritzen, 1993; Massa and Lauritzen, 2010)

$$p_{comb}(\phi, \psi_1, ..., \psi_M, Y_1, ..., Y_M) = p(\phi) \prod_{m=1}^{M} p_m(\psi_m, Y_m \mid \phi),$$

$$= \frac{\prod_{m=1}^{M} p_m(\phi, \psi_m, Y_m)}{p(\phi)^{M-1}}. \tag{1}$$

Markov combination is not immediately applicable when submodel prior marginals are distinct, so Goudie *et al.* define a *marginal replacement* procedure, where individual submodel prior marginals are replaced with a common marginal $p_{pool}(\phi) = h(p_1(\phi), ..., p_M(\phi))$ which is the result of a pooling function h that appropriately summarises all prior marginals (the choice of which is described below). The result of marginal replacement is

$$p_{repl,m}(\phi, \psi_m, Y_m) = p_{pool}(\phi) \frac{p_m(\phi, \psi_m, Y_m)}{p_m(\phi)}. \tag{2}$$

Goudie *et al.* show that $p_{repl,m}(\phi, \psi_m, Y_m)$ minimises the Kullback–Leibler (KL) divergence between a distribution $q(\phi, \psi_m, Y_m)$ and $p_m(\phi, \psi_m, Y_m)$ under the constraint that $q(\phi) = p_{pool}(\phi)$, and that marginal replacement is valid when ϕ is a deterministic function of the other parameters in submodel m. Markov melding joins the submodels via the Markov combination of the marginally replaced submodels

$$p_{meld}(\phi, \psi_1, ..., \psi_M, Y_1, ..., Y_M) = p_{pool}(\phi) \prod_{m=1}^{M} p_{repl,m}(\psi_m, Y_m \mid \phi),$$

$$= p_{pool}(\phi) \prod_{m=1}^{M} \frac{p_m(\phi, \psi_m, Y_m)}{p_m(\phi)}. \tag{3}$$

Pooled prior—Goudie *et al.* proposed forming $p_{pool}(\phi)$ using linear or logarithmic prior pooling (O'Hagan et al., 2006; Genest et al., 1986)

$$p_{pool,lin}(\phi) = \frac{1}{K_{lin}(\lambda)} \sum_{m=1}^{M} \lambda_m p_m(\phi), \quad K_{lin}(\lambda) = \int \sum_{m=1}^{M} \lambda_m p_m(\phi) d\phi, \tag{4}$$

$$p_{pool,log}(\phi) = \frac{1}{K_{log}(\lambda)} \prod_{m=1}^{M} p_m(\phi)^{\lambda_m}, \quad K_{log}(\lambda) = \int \prod_{m=1}^{M} p_m(\phi)^{\lambda_m} d\phi, \tag{5}$$

where $\lambda = (\lambda_1, ..., \lambda_M)$ are nonnegative weights, which are chosen subjectively to ensure $p_{pool}(\phi)$ appropriately represents prior knowledge about the common quantity. Two special cases of pooling are of particular interest. *Product of experts (PoE) pooling* (Hinton, 2002) is a special case of logarithmic pooling that occurs when $\lambda_m = 1$ for all m. *Dictatorial pooling* is a special case of either pooling method when $\lambda_{m'} = 1$ and, for all $m \neq m'$, $\lambda_m = 0$.

2 Chained model specification

Consider $m = 1, \ldots, M$ submodels each with data Y_m and parameters θ_m denoted $\mathrm{p}_m(\theta_m, Y_m)$, with $M \geq 3$. We assume that the submodels are connected in a manner akin to a chain and so can be ordered such that only 'adjacent' submodels in the chain have parameters in common. Specifically we assume that submodels m and $m + 1$ have some parameter $\phi_{m \cap m+1}$ in common for $m = 1, \ldots, M - 1$. For notational convenience define $\phi_1 = \phi_{1 \cap 2}$, $\phi_M = \phi_{M-1 \cap M}$ and $\phi_m = (\phi_{m-1 \cap m}, \phi_{m \cap m+1})$ for $m = 2, \ldots, M-1$, so that $\phi_m \subseteq \theta_m$ denotes the parameters in model m shared with another submodel. The submodel-specific parameters of submodel m are thus $\psi_m = \theta_m \mid \phi_m$. Define the vector of all common quantities $\phi = \cup_{m=1}^{M} \phi_m = (\phi_{1 \cap 2}, \phi_{2 \cap 3}, \ldots, \phi_{M-1 \cap M})$ so that all elements in ϕ are unique. Further denote by ϕ_{-m} the subvector of ϕ excluding the mth element. It will also be convenient to define $\psi = (\psi_1, \ldots, \psi_M)$ and likewise $Y = (Y_1, \ldots, Y_M)$. Note that all components of ϕ, ψ and Y may themselves be multivariate. Additionally, because $\phi_{m \cap m+1}$ may be a deterministic function of either θ_m or θ_{m+1} we refer to $\phi_{m \cap m+1}$ as a common parameter or a common quantity as appropriate.

All submodels, and marginal and conditional distributions thereof, have density functions that are assumed to exist and integrate to one. When considering conditional distributions we assume that the parameter being conditioned on has support in the relevant region. We define the mth *subposterior* as $\mathrm{p}_m(\phi_m, \psi_m \mid Y_m)$.

2.1 Extending marginal replacement

We now define the chained melded model by extending the marginal replacement procedure to submodels linked in a chain-like way. The proposed chained marginal replacement operation modifies the submodels to enforce a common prior for ϕ. This consistency allows us to employ Markov combination to unite the submodels.

Specifically, the mth marginally replaced submodel is

$$\mathrm{p}_{\mathrm{repl},m}(\phi, \psi_m, Y_m) = \mathrm{p}_{\mathrm{pool}}(\phi)\mathrm{p}_m(\psi_m, Y_m \mid \phi) = \mathrm{p}_{\mathrm{pool}}(\phi)\frac{\mathrm{p}_m(\phi_m, \psi_m, Y_m)}{\mathrm{p}_m(\phi_m)}, \tag{6}$$

where $\mathrm{p}_{\mathrm{pool}}(\phi) = g(\mathrm{p}_1(\phi_1), \mathrm{p}_2(\phi_2), \ldots, \mathrm{p}_m(\phi_m))$ is a pooling function that appropriately summarises all submodel prior marginals. The second equality in Equation (6) is because of the conditional independence $(\psi_m, Y_m \perp\!\!\!\perp \phi_{-m}) \mid \phi_m$ that exists due to the chained relationship between submodels. It is important to note that $\mathrm{p}_{\mathrm{repl},m}(\phi, \psi_m, Y_m)$ is defined on a larger parameter space than $\mathrm{p}_m(\phi_m, \psi_m, Y_m)$, as it includes ϕ_{-m}.

Define $\mathrm{p}_{\mathrm{repl},m}(\phi_m, \psi_m, Y_m) = \int \mathrm{p}_{\mathrm{repl},m}(\phi, \psi_m, Y_m) \mathrm{d}\phi_{-m}$. Each marginally replaced submodel, as defined in Equation (6), minimises the following KL divergence[2]

$$\mathrm{p}_{\mathrm{repl},m}(\phi_m, \psi_m, Y_m) = \arg\min_{\mathrm{q}}\{D_{\mathrm{KL}}(\mathrm{q} \,\|\, \mathrm{p}_m) \mid \mathrm{q}(\phi_m) = \mathrm{p}_{\mathrm{pool}}(\phi_m) \text{ for all } \phi_m\}, \tag{7}$$

---

[2]This is shown in Appendix B of the online supplement to Goudie et al. (2019).

where $p_{pool}(\phi_m) = \int p_{pool}(\phi)d\phi_{-m}$. We can thus interpret $p_{repl,m}(\phi_m, \psi_m, Y_m)$ as a minimally modified $p_m(\phi_m, \psi_m, Y_m)$ which admits $p_{pool}(\phi_m)$ as a marginal. Note that it is the combination of $p_{repl,m}(\phi_m, \psi_m, Y_m)$ and $p_{pool}(\phi_{-m} / \phi_m)$ that uniquely determine (6).

We form the chained melded model by taking the Markov combination of the marginally replaced submodels

$$p_{meld}(\phi, \psi, Y) = p_{pool}(\phi) \prod_{m=1}^{M} p_{repl,m}(\psi_m, Y_m \mid \phi), \tag{8}$$

$$= p_{pool}(\phi) \prod_{m=1}^{M} \frac{p_m(\phi_m, \psi_m, Y_m)}{p_m(\phi_m)}. \tag{9}$$

Rewriting (9) in terms of $\phi_{m \cap m+1}$ for $m = 1, \ldots, M-1$ yields

$$\begin{aligned}
p_{meld}(\phi, \psi, Y) &= p_{pool}(\phi) \frac{p_1(\phi_{1\cap2}, \psi_1, Y_1)}{p_1(\phi_{1\cap2})} \frac{p_M(\phi_{M-1\cap M}, \psi_M, Y_M)}{p_M(\phi_{M-1\cap M})}, \\
&\times \prod_{m=2}^{M-1} \left(\frac{p_m(\phi_{m-1\cap m}, \phi_{m\cap m+1}, \psi_m, Y_m)}{p_m(\phi_{m-1\cap m}, \phi_{m\cap m+1})} \right).
\end{aligned} \tag{10}$$

Finally, we use *chained melded posterior* $p_{meld}(\phi, \psi / Y) \propto p_{meld}(\phi, \psi, Y)$ to refer to posterior of the chained melded model conditioned on all data.

2.2 Pooled prior

Specifying (9) requires a joint prior for ϕ. As in Markov melding we form the joint prior by pooling the marginal priors, selecting a pooling function g that appropriately represents prior knowledge about the common quantities. We define $p_{pool}(\phi)$ as a generic function of all prior marginals

$$p_{pool}(\phi) = g(p_1(\phi_1), p_2(\phi_2), \ldots, p_M(\phi_M)), \tag{11}$$

$$= g(p_1(\phi_{1\cap2}), p_2(\phi_{1\cap2}, \phi_{2\cap3}), \ldots, p_M(\phi_{M-1\cap M})), \tag{12}$$

because we do not always wish to assume independence between the components of ϕ.

Two special cases of Equation (12) are noteworthy. Firstly, if all components of ϕ are independent, then we can form $p_{pool}(\phi)$ as the product of $M-1$ standard pooling functions h_m defined in Section 1.2

$$p_{pool}(\phi) = \prod_{m=1}^{M-1} p_{pool,m}(\phi_{m\cap m+1}), \tag{13}$$

$$p_{\text{pool},m}(\phi_{m \cap m+1}) = h_m(p_m(\phi_{m \cap m+1}), p_{m+1}(\phi_{m \cap m+1})). \tag{14}$$

A second case, in between complete dependence (12) and independence (14), is that if $p_m(\phi_{m-1 \cap m}, \phi_{m \cap m+1}) = p_m(\phi_{m-1 \cap m})p_m(\phi_{m \cap m+1})$ then we can define

$$p_{\text{pool}}(\phi) = g_1(p_1(\phi_{1 \cap 2}), \ldots, p_m(\phi_{m-1 \cap m}))g_2(p_m(\phi_{m \cap m+1}), \ldots, p_M(\phi_M)), \tag{15}$$

without any additional assumptions. That is, if any two consecutive components of ϕ are independent in the submodel containing both of them, we can divide the pooled prior specification problem into two pooling functions. The smaller number of arguments to g_1 and g_2 make it easier to choose appropriate forms for those functions.

Selecting a specific form of g is not trivial given the many choices of functional form and pooling weights (the latter of which we discuss momentarily). One complication is that standard linear and logarithmic pooling, as defined in Equations (4) and (5), are not immediately applicable when the submodel marginal distributions consider different quantities. We now propose extensions to logarithmic, linear, and dictatorial pooling for use in the case of chained melding.

Chained logarithmic pooling—Extending logarithmic pooling for chained Markov melding is straightforward. We define the logarithmically pooled prior to be

$$p_{\text{pool,log}}(\phi) = \frac{1}{K_{\log}(\lambda)} \prod_{m=1}^{M} p_m(\phi_m)^{\lambda_m}, \tag{16}$$

with $K_{\log}(\lambda) = \int \prod_{m=1}^{M} p_m(\phi_m)^{\lambda_m} d\phi$ for nonnegative weight vector $\lambda = (\lambda_1, \ldots, \lambda_M)$ and $\sum_{m=1}^{M} \lambda_m \geq 1$. Note that (16) does not imply independence between the elements of ϕ because

$$\prod_{m=1}^{M} p_m(\phi_m)^{\lambda_m} = p_1(\phi_{1 \cap 2})^{\lambda_1} \prod_{m=2}^{M-1} \left(p_m(\phi_{m-1 \cap m}, \phi_{m \cap m+1})^{\lambda_m} \right) p_M(\phi_{M-1 \cap M})^{\lambda_M}. \tag{17}$$

When $\lambda_1 = \lambda_2 = \ldots = \lambda_M = 1$ we obtain a special case which we call product-of-experts (PoE) pooling (Hinton, 2002).

Chained linear pooling—Our generalisation of linear pooling to handle marginals of different quantities is a two step procedure. The first step forms intermediary pooling densities via standard linear pooling, using appropriate marginals of the relevant quantity

$$p_{\text{pool},m}(\phi_{m \cap m+1}) \propto \lambda_{m,1} p_m(\phi_{m \cap m+1}) + \lambda_{m,2} p_{m+1}(\phi_{m \cap m+1}), \tag{18}$$

where $\lambda_m = (\lambda_{m,1}, \lambda_{m,2})$ are nonnegative pooling weights, and for $m = 2, \ldots, M-1$

$$p_m(\phi_{m \cap m+1}) = \int p_m(\phi_{m-1 \cap m}, \phi_{m \cap m+1}) d\phi_{m-1 \cap m}. \tag{19}$$

For $m = 1$ and $m = M$ the relevant marginals are $p_1(\phi_{1 \cap 2})$ and $p_M(\phi_{M-1 \cap M})$. In step two we form the pooled prior as the product of the intermediaries

$$p_{\text{pool, lin}}(\phi) = \frac{1}{K_{\text{lin}}(\lambda)} \prod_{m=1}^{M-1} p_{\text{pool},m}(\phi_{m \cap m+1}),\tag{20}$$

with $K_{\text{lin}}(\lambda) = \int \prod_{m=1}^{M-1} p_{\text{pool},m}(\phi_{m \cap m+1}) d\phi$, for $\lambda = (\lambda_1, \ldots, \lambda_M)$. Clearly, this assumes prior independence amongst all components of ϕ which may be undesirable, particularly if this independence was not present under one or more of the submodel priors. We discuss extensions to linear pooling that enable prior dependence between the components of ϕ in Section 6.

Dictatorial pooling—Chained Markov melding does not admit a direct analogue to dictatorial pooling as defined in Section 1.2 because not all submodel prior marginals contain all common quantities. For example, consider the logarithmically pooled prior of Equation (16) with, say, the mth entry in λ set to 1 and all others set to 0. This choice of λ results in $p_{\text{pool}}(\phi) = p(\phi_m)$, which is flat for ϕ_{-m}. It seems reasonable to require any generalisation of dictatorial pooling to result in a reasonable prior for all components in ϕ. Such a generalisation should also retain the original intention of dictatorial pooling, i.e. '*the authoritative prior for ϕ_m is $p_m(\phi_m)$*'.

We propose two possible forms of dictatorial pooling that satisfy the aforementioned criteria. *Partial dictatorial pooling* enforces a single submodel prior for the relevant components of ϕ, with no restrictions on the pooling of the remaining components; and *complete dictatorial pooling* which requires selecting one of the two possible submodel priors for each component of ϕ.

Partial dictatorial pooling considers $p_m(\phi_m)$ as the authoritative prior for $\phi_m = (\phi_{m-1 \cap m}, \phi_{m \cap m+1})$. This results in,

$$\begin{aligned} p_{\text{pool,dict}}(\phi) = {} & g_1(p_1(\phi_{1 \cap 2}), \ldots, p_{m-1}(\phi_{m-2 \cap m-1})) \\ & \times p_m(\phi_{m-1 \cap m}, \phi_{m \cap m+1}) \\ & \times g_2(p_{m+1}(\phi_{m+1 \cap m+2}), \ldots, p_M(\phi_{M-1 \cap M})), \end{aligned}\tag{21}$$

where g_1 and g_2 are linear or logarithmic pooling functions as desired[3].

Complete dictatorial pooling requires the marginal pooled prior for each component in ϕ to be chosen solely on the basis of only one of the two priors specified for it under the submodels. For $m = 1, \ldots, M-1$, the mth marginal of the pooled prior is either

$$p_{\text{pool,dict}}(\phi_{m \cap m+1}) := \begin{cases} p_m(\phi_{m \cap m+1}) & \text{or} \\ p_{m+1}(\phi_{m \cap m+1}). \end{cases}\tag{22}$$

---

[3]Some care is required if the authoritative submodel is p_m for $m \in \{1, 2, M-1, M\}$. If it is taken to be $m \in \{1, 2\}$, then g_1 does not exist, and additionally in the $m = 1$ case $p_1(\phi_{0 \cap 1}, \phi_{1 \cap 2}) := p_1(\phi_{1 \cap 2})$. The $m \in \{M-1, M\}$ cases have analogous definitions.

If two consecutive marginals are chosen to have the same submodel prior, then we wish to retain the dependence between $\phi_{m-1 \cap m}$ and $\phi_{m \cap m+1}$ present in p_m. We thus redefine consecutive terms so that

$$
\begin{aligned}
p_{\text{pool,dict}}(\phi_{m-1 \cap m}) p_{\text{pool,dict}}(\phi_{m \cap m+1}) &= p_m(\phi_{m-1 \cap m}) p_m(\phi_{m \cap m+1}) && \text{(FROM EQ. (22);} \\
p_{\text{pool,dict}}(\phi_{m-1 \cap m}) p_{\text{pool,dict}}(\phi_{m \cap m+1}) &:= p_m(\phi_{m-1 \cap m}, \phi_{m \cap m+1}). && \text{(REDEFINED)}
\end{aligned}
\tag{23}
$$

The complete dictatorially pooled prior is thus

$$
p_{\text{pool,dict}}(\phi) = \prod_{m=1}^{M-1} p_{\text{pool,dict}}(\phi_{m \cap m+1}),
\tag{24}
$$

where, subject to the potential modification in Equation (23), the terms in the product are as defined in Equation (22). For example, if $M = 5$ and we wish to ignore p_2 and p_4 when constructing the pooled prior and instead associate $\phi_{1 \cap 2}$ with p_1, both $\phi_{2 \cap 3}$ and $\phi_{3 \cap 4}$ with p_3, and $\phi_{4 \cap 5}$ with p_5, then

$$
\begin{aligned}
p_{\text{pool,dict}}(\phi) &= \overbrace{p_{1,\text{dict}}(\phi_{1 \cap 2}) P_{3,\text{dict}}(\phi_{2 \cap 3}) p_{3,\text{dict}}(\phi_{3 \cap 4})}^{\text{APPLY EQ. (23)}} p_{5,\text{dict}}(\phi_{4 \cap 5}) \\
&= p_1(\phi_{1 \cap 2}) p_3(\phi_{2 \cap 3}, \phi_{3 \cap 4}) p_5(\phi_{4 \cap 5}).
\end{aligned}
\tag{25}
$$

Pooling weights—Choosing values for the pooling weights is an important step in specifying the pooled prior (Carvalho et al., 2022; Abbas, 2009; Rufo et al., 2012a,b). Because appropriate values for the weights depend on the submodels being pooled and the information available *a priori*, universal recommendations are impossible, so we illustrate the impact of different choices in a straightforward example. It is important that prior predictive visualisations of the pooled prior are produced (Gabry et al., 2019; Gelman et al., 2020) to guide the choice of pooling weights and ensure that the result suitably represents the available information. Figure 3 illustrates how λ and the choice of pooling method impacts $p_{\text{pool}}(\phi)$ when pooling normal distributions. Specifically, we consider $M = 3$ submodels and pool

$$
\begin{aligned}
p_1(\phi_{1 \cap 2}) &= N(\phi_{1 \cap 2}; \mu_1, \sigma_1^2), \quad p_3(\phi_{2 \cap 3}) = N(\phi_{2 \cap 3}; \mu_3, \sigma_3^2), \\
p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}) &= N\left(\begin{bmatrix} \phi_{1 \cap 2} \\ \phi_{2 \cap 3} \end{bmatrix}; \begin{bmatrix} \mu_{2,1} \\ \mu_{2,2} \end{bmatrix}, \begin{bmatrix} \sigma_2^2 & \rho\sigma_2^2 \\ \rho\sigma_2^2 & \sigma_2^2 \end{bmatrix} \right),
\end{aligned}
\tag{26}
$$

where $N(\phi; \mu, \sigma^2)$ is the normal density function with mean μ and variance σ^2 (or covariance matrix where appropriate). The two dimensional density function p_2 has an additional parameter ρ, which controls the intra-submodel marginal correlation. We set $\mu_1 = -2.5$, $\mu_2 = [\mu_{2,1} \, \mu_{2,2}]' = [0 \, 0]'$, $\mu_3 = 2.5$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ and $\rho = 0.8$. In the logarithmic case we set $\lambda_1 = \lambda_3$ and parameterise $\lambda_2 = 1 - 2\lambda_1$, so that $\lambda_1 + \lambda_2 + \lambda_3 = 1$ whilst limiting ourselves to varying only λ_1. Similarly, in the linear case we set $\lambda_{1,1} = \lambda_{2,2} = \lambda_1$ and $\lambda_{1,2} = \lambda_{2,1} = 1 - 2\lambda_1$. We consider 5 evenly spaced values of $\lambda_1 \in [0, 0.5]$.

For both pooling methods, as the weight λ_1 associated with models p_1 and p_3 increases, the relative contributions of $p_1(\phi_{1\cap2})$ and $p_3(\phi_{2\cap3})$ increase. Note the lack of correlation in p_{pool} under linear pooling (right column of Figure 3) due to Equation (20). A large, near-flat plateau is visible in the $\lambda_1 = 0.25$ and $\lambda_1 = 0.375$ cases, which is a result of the mixture of four, 2-D normal distributions that linear pooling produces in this example. The logarithmic pooling process produces a more concentrated prior for small values of λ_1, and does not result in *a priori* independence between $\phi_{1\cap2}$ and $\phi_{2\cap3}$. Appendix A shows analytically that λ_2 controls the quantity of correlation present in p_{pool} in this setting.

3 Posterior estimation

We now present a multi-stage MCMC method for generating samples from the melded posterior. Whilst the melded posterior is a standard Bayesian posterior and so can, in principle, be targeted using any suitable Monte Carlo method, in practice this may be cumbersome or infeasible. More specifically, it may be feasible to fit each submodel separately using standard methods, but when the submodels are combined – either through Markov melding, or by expanding the definition of one submodel to include another – the computation required to estimate the posterior in a single step poses an insurmountable barrier. In such settings we can employ multi-stage posterior estimation methods including Tom et al. (2010), Lunn et al. (2013), Hooten et al. (2019), and Mauff et al. (2020). We propose a multi-stage strategy that uses the chain-like relationship to both avoid evaluating all submodels simultaneously, and parallelise the computation required in the first stage to produce posterior samples in less time than an equivalent sequential method[4]. Avoiding concurrently evaluating all submodels also enables the reuse of existing software, minimising the need for custom submodel and/or sampler implementations.

We also describe an approximate method, where stage one submodels are summarised by normal distributions for use in stage two.

We consider the $M = 3$ case, as this setting includes both of our examples. Our approach can be extended to $M > 3$ settings, although we anticipate that it is unlikely to be suitable for large M settings. We discuss some of difficulties associated with generic, parallel methodology for efficient posterior sampling in Section 6.

3.1 Parallel sampler

Our proposed strategy involves obtaining in stage one samples from submodels 1 and 3 in parallel. Stage two reuses these samples in a Metropolis-within-Gibbs sampler, which targets the full melded posterior. The stage specific targets are displayed in Figure 4.

The parallel sampler assumes that the pooled prior decomposes such that

$$p_{pool}(\phi) = p_{pool,1}(\phi_{1\cap2})p_{pool,2}(\phi_{1\cap2}, \phi_{2\cap3})p_{pool,3}(\phi_{2\cap3}). \tag{27}$$

---

[4]For completeness, Appendix B describes such a sequential MCMC sampler. We do not use the sequential sampler in this paper.

All pooled priors trivially satisfy (27) by assuming $p_{pool,1}(\phi_{1 \cap 2})$ and $p_{pool,3}(\phi_{2 \cap 3})$ are improper and/or flat distributions. Alternatively we may choose $p_{pool,1}(\phi_{1 \cap 2}) = p_1(\phi_{1 \cap 2})$ and $p_{pool,3}(\phi_{2 \cap 3}) = p_3(\phi_{2 \cap 3})$, with appropriate adjustments to $p_{pool,2}(\phi_{1 \cap 2}, \phi_{2 \cap 3})$. This choice targets, in stage one, the subposteriors of p_1 and p_3 under their original prior for $\phi_{1 \cap 2}$ and $\phi_{2 \cap 3}$ respectively.

Stage one—Two independent, parallel sampling processes occur in stage one. Terms from the melded model that pertain to p_1 and p_3 are isolated

$$p_{meld,1}(\phi_{1 \cap 2}, \psi_1 \mid Y_1) \propto p_{pool,1}(\phi_{1 \cap 2}) \frac{p_1(\phi_{1 \cap 2}, \psi_1, Y_1)}{p_1(\phi_{1 \cap 2})}, \tag{28}$$

$$p_{meld,3}(\phi_{2 \cap 3}, \psi_3 \mid Y_3) \propto p_{pool,3}(\phi_{2 \cap 3}) \frac{p_3(\phi_{2 \cap 3}, \psi_3, Y_3)}{p_3(\phi_{2 \cap 3})}, \tag{29}$$

and targeted using standard MCMC methodology. Assuming that the stage one chains converge and after discarding warmup iterations –possibly thinning them, if within-chain correlation is high– we obtain N_1 samples from $\{(\phi_{1 \cap 2}, \psi_1)_n\}_{n=1}^{N_1}$ from $p_{meld,1}(\phi_{1 \cap 2}, \psi_2 / Y_1)$, and N_3 samples $\{(\phi_{2 \cap 3}, \psi_3)_n\}_{n=1}^{N_3}$ from $p_{meld,3}(\phi_{2 \cap 3}, \psi_3 / Y_3)$. For well mixing stage one Markov chains targeting the correct stationary distribution, and large values of N_1 or N_3, the stage one samples accurately approximate the subposteriors.

Stage two—Stage two targets the melded posterior of Equation (9) using a Metropolis-within-Gibbs sampler, where the proposal distributions are

$$\phi_{1 \cap 2}^*, \psi_1^* \mid \phi_{2 \cap 3}, \psi_2, \psi_3 \sim p_{meld,1}(\phi_{1 \cap 2}^*, \psi_1^* \mid Y_1), \tag{30}$$

$$\phi_{2 \cap 3}^*, \psi_3^* \mid \phi_{1 \cap 2}, \psi_1, \psi_2 \sim p_{meld,3}(\phi_{2 \cap 3}^*, \psi_3^* \mid Y_3), \tag{31}$$

$$\psi_2^* \mid \phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_1, \psi_3 \sim q(\psi_2^* \mid \psi_2), \tag{32}$$

where $q(\psi_2^* \mid \psi_2)$ is a generic proposal distribution for ψ_2. We draw an index n_1^* uniformly from $\{1, \ldots, N_1\}$ and use the corresponding value $(\phi_{1 \cap 2}^*, \psi_1^*)_{n_1^*}$ as the proposal, doing likewise for n_3^* and $(\phi_{2 \cap 3}^*, \psi_3^*)_{n_3^*}$. The acceptance probabilities for these updates are

$$\alpha((\phi_{1 \cap 2}^*, \psi_1^*)_{n_1^*}, (\phi_{1 \cap 2}, \psi_1)_{n_1}) = \frac{p_{pool,2}(\phi_{1 \cap 2}^*, \phi_{2 \cap 3}) p_2(\phi_{1 \cap 2}^*, \phi_{2 \cap 3}, \psi_2, Y_2) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3})}{p_{pool,2}(\phi_{1 \cap 2}, \phi_{2 \cap 3}) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2, Y_2) p_2(\phi_{1 \cap 2}^*, \phi_{2 \cap 3})}, \tag{33}$$

$$\alpha((\phi_{2 \cap 3}^*, \psi_3^*)_{n_3^*}, (\phi_{2 \cap 3}, \psi_3)_{n_3}) = \frac{p_{pool,2}(\phi_{1 \cap 2}, \phi_{2 \cap 3}^*) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}^*, \psi_2, Y_2) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3})}{p_{pool,2}(\phi_{1 \cap 2}, \phi_{2 \cap 3}) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2, Y_2) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}^*)}, \tag{34}$$

$$\alpha(\psi_2^*, \psi_2) = \frac{p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2^*, Y_2)}{p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2, Y_2)} \frac{q(\psi_2 \mid \psi_2^*)}{q(\psi_2^* \mid \psi_2)}, \tag{35}$$

where $a(x, z)$ denotes the probability associated with a move from z to x. Note that all stage two acceptance probabilities only contain terms from the second submodel and the pooled prior, and thus do not depend on ψ_1 or ψ_3. If a move is accepted then we also store the index, i.e. n_1^* or n_3^*, associated with the move, otherwise we store the current value of the index. The stored indices are used to appropriately resample ψ_1 and ψ_3 from the stage one samples.

3.2 Normal approximations to submodel components

Normal approximations are commonly employed to summarise submodels for subsequent use in more complex models. For example, two-stage meta-analyses often use a normal distribution centred on each studies' effect estimate (Burke et al., 2017). Suppose we employ such an approximation to summarise the prior and posterior of $\phi_{1 \cap 2}$ and $\phi_{2 \cap 3}$ under p_1 and p_3 respectively. In addition, assume that (a) such approximations are appropriate for $p_1(\phi_{1 \cap 2})$, $p_1(\phi_{1 \cap 2} / Y_1)$, $p_3(\phi_{2 \cap 3})$, and $p_3(\phi_{2 \cap 3} / Y_3)$, (b) we are not interested in ψ_1 and ψ_3, and can integrate them out of all relevant densities, and (c) we employ our second form of dictatorial pooling and choose $p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3})$ as the authoritative prior. The latter two assumptions imply that the melded posterior of interest is proportional to

$$p_{meld}(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2 \mid Y) \propto \frac{p_1(\phi_{1 \cap 2} \mid Y_1)}{p_1(\phi_{1 \cap 2})} p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2 \mid Y_2) \frac{p_3(\phi_{2 \cap 3} \mid Y_3)}{p_3(\phi_{2 \cap 3})} . \tag{36}$$

Denote the normal approximation of $p_1(\phi_{1 \cap 2} \mid Y_1)$ as $\widehat{P}_1\big(\phi_{1 \cap 2} \mid \hat{\mu}_1, \widehat{\Sigma}_1\big)$ which is a normal distribution with mean $\hat{\mu}_1$ and covariance matrix $\widehat{\Sigma}_1$. The corresponding normal approximation of the prior $p_1(\phi_{1 \cap 2})$ is $\widehat{P}_1\big(\phi_{1 \cap 2} \mid \hat{\mu}_{1,0}, \widehat{\Sigma}_{1,0}\big)$. The equivalent approximations for the subposterior and prior of p_3 are $\widehat{P}_3\big(\phi_{2 \cap 3} \mid \hat{\mu}_3, \widehat{\Sigma}_3\big)$ and $\hat{p}_3\big(\phi_{2 \cap 3} \mid \hat{\mu}_{3,0}, \widehat{\Sigma}_{3,0}\big)$ respectively. Substituting in the approximations and using standard results for Gaussian density functions (see Bromiley (2003) and Appendix C) results in

$$\hat{p}_{meld}(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2 \mid Y) \propto \hat{p}\big((\phi_{1 \cap 2}, \phi_{2 \cap 3}) \mid \hat{\mu}, \widehat{\Sigma}\big) p_2(\phi_{1 \cap 2}, \phi_{2 \cap 3}, \psi_2 \mid Y_2), \tag{37}$$

where

$$\hat{\mu}_{nu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_3 \end{bmatrix}, \quad \widehat{\Sigma}_{nu} = \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_3 \end{bmatrix}, \quad \hat{\mu}_{de} = \begin{bmatrix} \hat{\mu}_{1,0} \\ \hat{\mu}_{3,0} \end{bmatrix}, \quad \widehat{\Sigma}_{de} = \begin{bmatrix} \widehat{\Sigma}_{1,0} & 0 \\ 0 & \widehat{\Sigma}_{3,0} \end{bmatrix},$$
$$\widehat{\Sigma} = \big(\widehat{\Sigma}_{nu}^{-1} - \widehat{\Sigma}_{de}^{-1}\big)^{-1}, \quad \hat{\mu} = \widehat{\Sigma}\big(\widehat{\Sigma}_{nu}^{-1}\hat{\mu}_{nu} - \widehat{\Sigma}_{de}^{-1}\hat{\mu}_{de}\big). \tag{38}$$

Standard MCMC methods can be used to sample from the approximate melded posterior. If instead we opt for product-of-experts pooling, all $\hat{\mu}_{de}$ and $\widehat{\sum}_{de}$ terms disappear from the parameter definitions in Equation (38).

4 An integrated population model for little owls

We now return to the integrated population model (IPM) for the little owls introduced in Section 1.1. Finke et al. (2019) consider a number of variations on the original model of Schaub et al. (2006) and Abadi et al. (2010): here we consider only the variant from Finke et al. (2019) with the highest marginal likelihood (Model 4 of their online supplement). This example is particularly interesting to us as, for a certain choice of pooling function and pooling weights, the chained Markov melded model and the IPM are identical. This coincidence allows us to use the posterior from the IPM as a benchmark for our multi-stage sampler.

Before we detail the specifics of each submodel, we must introduce some notation. Data and parameters are stratified into two age-groups $a \in \{J, A\}$ where J denotes juvenile owls (less than one year old) and A adults, two sexes $s \in \{M, F\}$, and observations occur annually at times $t \in \{1, \ldots, T\}$, with $T = 25$. The sex- and age-specific probability of an owl surviving from time t to $t + 1$ is $\delta_{a,s,t}$, and the sex-specific probability of a previously captured owl being recaptured at time $t + 1$ is $\pi_{s,t+1}$ so long as the owl is alive at time $t + 1$.

4.1 Capture recapture: p_1

Capture-recapture data pertain to owls that are released at time t (having been captured and tagged), and then recaptured at time $u = t + 1, \ldots, T$, or not recaptured before the conclusion of the study, in which case $u = T + 1$. Define $M_{a,s,t,u}$ as the number of owls of age-group a and sex s released at time t and recaptured at time u. We aggregate these observations into age- and sex-specific matrices $M_{a,s}$, with T rows, corresponding to release times, and $T+1$ columns, corresponding to recapture times. Let $R_{a,s,t} = \sum_{u=1}^{T+1} M_{a,s,t,u}$ be the number of owls released at time t, i.e. a vector containing the row-wise sum of the entries in $M_{a,s}$. The recapture times for owls released at time t follow an age- and sex-specific multinomial likelihood

$$(M_{a,s,t,1}, \ldots, M_{a,s,t,T+1}) \sim \text{Multinomial}(R_{a,s,t}, Q_{a,s,t}), \tag{39}$$

with probabilities $Q_{a,s,t} = (Q_{a,s,t,1}, \ldots, Q_{a,s,t,T+1})$ such that

$$Q_{a,s,t,u} = \begin{cases} 0, & \text{for } u = 1, \ldots, t \\ \delta_{a,s,t}\pi_{s,u}\prod_{r=t+1}^{u-1}\delta_{a,s,r}(1 - \pi_{s,r}), & \text{for } u = t+1, \ldots, T \\ 1 - \sum_{r=1}^{T} Q_{a,s,t,r}, & \text{if } u = T+1. \end{cases} \tag{40}$$

4.2 Count data model: p_2

To estimate population abundance, a two level model is used: the first level models the observed (counted) number of females at each point in time denoted y_t, with a second, latent process modelling the total number of females in population. The observation model is

$$y_t \mid x_t \sim \text{Poisson}(x_t), \tag{41}$$

where we denote the number of juvenile and adult females in the population at time t as $x_{J,t}$ and $x_{A,t}$ respectively, with $x_t = x_{J,t} + x_{A,t}$. If sur_t adult females survive from time $t-1$ to time t, and imm_t adult females immigrate over the same time period, then the latent, population level model is

$$
\begin{aligned}
x_{J,t} \mid x_{t-1}, \rho, \delta_{J,F,t-1} &\sim \text{Poisson}\!\left(x_{t-1}\frac{\rho}{2}\delta_{J,F,t-1}\right), \\
\text{sur}_t \mid x_{t-1}, \delta_{A,F,t-1} &\sim \text{Binomial}(x_{t-1}, \delta_{A,F,t-1}), \\
\text{imm}_t \mid x_{t-1}, \eta_t &\sim \text{Poisson}(x_{t-1}\eta_t), \\
x_{A,t} &= \text{sur}_t + \text{imm}_t,
\end{aligned}
\tag{42}
$$

where η_t is the immigration rate. The initial population sizes $x_{J,1}$ and $x_{A,1}$ have independent discrete uniform priors on $\{0, 1, \ldots, 50\}$. If $x_{t-1} = 0$ then we assume that the Poisson and binomial distributions become point masses at zero.

4.3 Fecundity: p_3

The fecundity submodel considers the number of breeding females at time t denoted N_t, and the number of chicks produced that survive and leave the nest denoted n_t. A Poisson model is employed to estimate fecundity (reproductive rate) ρ

$$n_t \sim \text{Poisson}(N_t\rho). \tag{43}$$

4.4 Parameterisation and melding quantities

Abadi et al. (2010) parameterise the time dependent quantities via linear predictors to minimise the number of parameters in the submodels. The specific parameterisation of Finke et al. (2019) we employ is

$$
\begin{aligned}
\text{logit}(\delta_{a,s,t}) &= \alpha_0 + \alpha_1 \, \mathbb{I}\,(s = M) + \alpha_2 \, \mathbb{I}\,(a = A), \quad \log(\eta_t) = \alpha_6, \\
\text{logit}(\pi_{s,u}) &= \alpha_4 \, \mathbb{I}\,(s = M) + \alpha_{5,u}, \quad \text{for } u = 2, \ldots T,
\end{aligned}
\tag{44}
$$

thus the quantities in common between the submodels are $\phi_{1 \cap 2} = (\alpha_0, \alpha_2)$ and $\phi_{2 \cap 3} = \rho$. To align the notation of this example with our chained melding notation we define, for all permitted values of a, s and t, $Y_1 = (M_{a,s})$, $\psi_1 = \left(\alpha_1, \alpha_4, (\alpha_{5,u})_{u=2}^T\right)$ $Y_2 = (y_t)$, $\psi_2 = (x_{J,t}, \alpha_6, \text{sur}_t, \text{imm}_t)$; and $Y_3 = (N_t, n_t)$, $\psi_3 = \varnothing$. Note that the definition of $\phi_{1 \cap 2}$ does not include α_1 as it is male specific and does not exist in p_2. The model variant of Finke et al. (2019) we consider does not include α_3, and for comparability we keep the other parameter indices the same.

4.5 Priors

We use the priors of Finke et al. (2019) for the parameters in each submodel. Denote $a = (a_0, a_1, a_2, a_4, a_6)$. In both p_1 and p_2 the elements of a are assigned independent Normal$(0, 2^2)$ priors truncated to $[-10, 10]$. The time varying recapture probabilities $a_{5,u}$ also have Normal$(0, 2^2)$ priors truncated to $[-10, 10]$. A Uniform$(0, 10)$ prior is assigned to ρ in p_2 and p_3.

To completely specify p_{meld} we must choose how to form $p_{pool}(\phi_{1 \cap 2}, \phi_{2 \cap 3})$. We form $p_{pool}(\phi_{1 \cap 2}, \phi_{2 \cap 3})$ using three different pooling methods and estimate the melded posterior in each case. The first pooling method is product-of-experts (PoE) pooling, which is logarithmic pooling with $\lambda = (1, 1, 1)$, and we denote the melded posterior as $p_{meld,PoE}$. We also use logarithmic pooling with $\lambda = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, which is denoted $p_{meld,log}$ and results in the chained melded model being identical to the IPM. The final pooling method is linear pooling with $\lambda = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, denoted $p_{meld,lin}$.

4.6 Posterior estimation

We estimate the melded posterior – $p_{meld}(\phi, \psi \mid Y)$, proportional to Equation (9) – using both the parallel sampler (Section 3.1) and normal approximation (Section 3.2). This allows us to use pre-existing implementations of the submodels. Specifically, the capture-recapture submodel is written in BUGS (Lunn et al., 2009) and sampled via `rjags` (Plummer, 2019). The fecundity submodel is written in Stan (Carpenter et al., 2017) and sampled via `rstan` (Stan Development Team, 2021). The count data submodel is also written in BUGS, and we reuse this implementation in stage two of the multi-stage sampler via NIMBLE (de Valpine et al., 2017) and its R interface (NIMBLE Development Team, 2019). The approximate melded posterior obtained by Section 3.2 is sampled using `rjags`. Code and data for this example, as well as trace plots and numerical convergence measures (Vehtari et al., 2020) for both stages of the parallel sampling process, are available in the accompanying online repository[5].

4.7 Results

We empirically validate our methodology and sampler by comparing the melded posterior samples to a large sample – 6 chains, each containing 1×10^5 post-warmup iterations – from the original IPM posterior. Similarity in the posteriors is expected as the IPM is effectively the joint model we wish to approximate with the chained melded model. It is simply fortunate, from a modelling standpoint, that this example's joint model is easy to construct and computationally feasible with standard tools. Note that under logarithmic pooling with $\lambda = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ the melded posterior is identical to the original IPM, so any differences between the two posteriors are attributable to the multi-stage sampler. Figure 5 depicts the posterior credible intervals (Gabry et al., 2021; Kay, 2020) for the common quantities from the individual submodels, the melded models, and the original IPM. The top row in Figure 5

---

[5] https://doi.org/10.5281/zenodo.6552714

indicates that the count data alone (p_2) contain minimal information about a_0, a_2 and ρ; incorporating the data from the other submodels is essential for precise estimates.

The multi-stage sampler works well by producing melded posterior estimates generally similar to the original IPM estimate, and are near identical for logarithmic pooling. PoE pooling produces the posterior most different from the original IPM, as it yields a prior for (a_0, a_2) that is more concentrated around zero than the other pooling methods. The lack of large differences between the melded posteriors that use different pooled priors indicates that the prior has almost no effect on the posterior. The similarity of the approximate approach ($\widehat{p}_{\text{meld}}$ - bottom row of Figure 5) to the melding approaches suggests that the normal approximations are good summaries of the subposteriors, and that the approximate melding procedure of Section 3.2 is suitable for this example.

5 Survival analysis with time varying covariates and uncertain event times

We return now to the respiratory failure example introduced in Section 1.1. Our intention is to illustrate the application of chained Markov melding to an example of realistic complexity, and explore empirically the importance of accounting for all sources of uncertainty by comparing chained Markov melding to equivalent analyses which use only a point estimate summary of the uncertainty. Specifically, event times and indicators are a noninvertible function of other parameters in the first submodel, and are an uncertain response in the survival submodel. Chained Markov melding enables us to specify a suitable joint model despite these complications.

There are $i = 1, \ldots, N$ individuals in the data set. Each individual is admitted to the ICU at time 0, and is discharged or dies at time C_i. See Appendix I for information on how the $N = 37$ individuals were selected from MIMIC-III (Johnson et al., 2016).

5.1 P/F ratio submodel (B-spline): p_1

The first submodel fits a B-spline to the PaO_2/FiO_2 data to calculate if and when an individual experiences respiratory failure. Each individual has PaO_2/FiO_2 ratio observations $z_{i,j}$ (in units of mmHg) at times $t_{i,j}$, with $j = 1, \ldots, J_i$. For each individual denote the vector of observations $z_i = (z_{i,1}, \ldots, z_{i,J_i})$ and observation times $t_i = (t_{i,1}, \ldots, t_{i,J_i})$. To improve computational performance, we standardise the P/F ratio data for each individual such that $z_{i,j} = \frac{\tilde{z}_{i,j} - \bar{z}_i}{\hat{s}_i}$, where $\tilde{z}_{i,j}$ is the underlying unstandardised observation with mean $\bar{z}_i$ and standard deviation $\hat{s}_i$. Similarly we rescale the threshold for respiratory failure: $\tau_i = \frac{300 - \bar{z}_i}{\hat{s}_i}$.

We choose to model the P/F ratio using cubic B-splines and 7 internal knots, and do not include an intercept column in the spline basis (for background on B-splines see: Chapter 2 in Hastie and Tibshirani, 1999; and the supplementary material of Wang and Yan, 2021). The internal knots are evenly spaced between two additional boundary knots at $\min(t_i)$ and $\max(t_i)$. These choices result in $k = 1, \ldots, 10$ spline basis terms per individual, with coefficients $\zeta_{i,k}$ where $\zeta_i = (\zeta_{i,1}, \ldots, \zeta_{i,10})$. We denote the individual specific B-spline basis evaluated at time $t_{i,j}$ as $B_i(t_{i,j}) \in [0, \infty)^{10}$ so that the submodel can be written as

$$z_{i,j} = \beta_{0,i} + B_i(t_{i,j})^\top \zeta_i + \varepsilon_{i,j}. \tag{45}$$

We employ a weakly informative prior for the intercept $\beta_{0,i} \sim \mathrm{N}(0, 1^2)$, a heavy tailed distribution for the error term[6] $\varepsilon_{i,j} \sim t_5(0, \omega_i)$, and a weakly informative half-normal prior for the unknown scale parameter $\omega_i \sim \mathrm{N}_+(0, 1^2)$. For the spline basis coefficients we set $\zeta_{i,1} \sim \mathrm{N}(0, 0.1^2)$, and for $k = 2, \ldots, 10$ we employ the random-walk prior $\zeta_{i,k} \sim \mathrm{N}(\zeta_{i,k-1}, 0.1^2)$ from Kharratzadeh (2017).

We identify that a respiratory failure event occurred (which we denote by $d_i = 1$) at event time T_i if a solution to the following optimisation problem exists

$$T_i = \min_t \{ \tau_i = \beta_{0,i} + B_i(t)\,\zeta_i \mid t \in [\max(0, \min(t_i)),\ \max(t_i)] \}, \tag{46}$$

We attempt to solve Equation 46 using a standard multiple root finder (Soetaert et al., 2020). If there are no roots then the individual died or was discharged before respiratory failure occurred so we set $T_i = C_i$ and $d_i = 0$. The relationship between T_i and other model coefficients is displayed in the left hand panel of Figure 6.

5.2 Cumulative fluid submodel (piecewise linear) p₃

The rate of fluid administration reflects the clinical management of patients by ICU staff, and hence changes to the rate reflect decisions to change treatment strategy. We employ a breakpoint regression model to capture the effect of such decisions, and consider only one breakpoint as this appears sufficient to fit the observed data. Specifically, we model the 8-hourly cumulative fluid balance data $x_{i,l}$ (in litres) at times $u_{i,l}$, $l = 1, \ldots, L_i$.

The cumulative data are derived from the raw fluid input/output observations, which we detail in Appendix D. We denote the complete vector of observations by $x_i = (x_{i,1}, \ldots, x_{i,L_i})$ and times by $u_i = (u_{i,1}, \ldots, u_{i,L_i})$.

We assume a piecewise linear model with $\eta_{0,i}$ as the value at the breakpoint at time κ_i, slope $\eta_{1,i}^b$ before the breakpoint, and slope $\eta_{1,i}^a$ after the breakpoint. We write this submodel as

$$x_{i,l} = m_i(u_{i,l}) + \epsilon_{i,l},$$
$$m_i(u_{i,l}) = \eta_{0,i} + \eta_{1,i}^b (u_{i,l} - \kappa_i)\,\mathbb{1}_{\{u_{i,l} < \kappa_i\}} + \eta_{1,i}^a (u_{i,l} - \kappa_i)\,\mathbb{1}_{\{u_{i,l} \geq \kappa_i\}}. \tag{47}$$

It will be useful to refer to the fitted value of this submodel at arbitrary time as $m_i(t)$. We assume a weakly informative prior for the error term $\epsilon_{i,l} \sim \mathrm{N}(0, \sigma_{x,i}^2)$, with individual-specific error variances $\sigma_{x,i} \sim \mathrm{N}_+(0, 5^2)$, and specific, informative priors for the slope before the breakpoint $\eta_{1,i}^b \sim \mathrm{Gamma}(1.53, 0.24)$ and after $\eta_{1,i}^a \sim \mathrm{Gamma}(1.53, 0.24)$. An appropriate prior

---

[6]P/F data contain many outliers for, amongst many possible reasons, arterial/venous blood sample mislabelling; incorrectly recorded oxygenation support information; and differences between sample collection time, lab result time, and the observation time as recorded in the EHR.

for κ_i and $\eta_{0,i}$ is challenging to specify due to the relationship between the two parameters and the individual-specific support for κ_i. We address both challenges by reparameterisation, resulting in a prior for κ_i that, in the absence of other information, places the breakpoint in the middle of an individual's ICU stay, and a prior for $\eta_{0,i}$ that captures the diverse pathways into ICU that an individual can experience. Details and justifications for all the informative priors are available in Appendix E. Figure 6 displays the parameters and their relationship to the fitted regression line.

5.3 Survival submodel p$_2$

The rate at which fluid is administered is thought to influence the time to respiratory failure (Seethala et al., 2017), so we explore this relationship using a survival model. Individuals experience respiratory failure ($d_i = 1$) at time $0 < t < C_i$, or are censored ($d_i = 0$, $t = C_i$). We assume a Weibull hazard with shape parameter γ for the event times. All individuals have baseline (time invariant) covariates $w_{i,a}$, $a = 1, \ldots, A$, with $w_i = (1, w_{i,1}, \ldots, w_{i,A})$ (i.e. including an intercept term), and common coefficients $\theta = (\theta_0, \ldots, \theta_A)$. The hazard is assumed to be influenced by these covariates and the rate of increase $\frac{\partial}{\partial t} m_i(t)$ in the cumulative fluid balance. The strength of the latter relationship is captured by α. Hence, the hazard is

$$h_i(t) = \gamma t^{\gamma - 1} \exp\left\{ w_i^\top \theta + \alpha \frac{\partial}{\partial t} m_i(t) \right\}, \tag{48}$$

$$\frac{\partial}{\partial t} m_i(t) = \eta_{1,i}^b \mathbb{1}_{\{t < \kappa_i\}} + \eta_{1,i}^a \mathbb{1}_{\{t \geq \kappa_i\}}, \tag{49}$$

The survival function at an individual's observed event time and status, (T_i, d_i), denoted $S_i(T_i) = \exp\left\{ -\int_0^{T_i} h_i(u) \mathrm{d}u \right\}$, has an analytic form which we derive in Appendix F.

Thus, the likelihood for individual i is

$$p\left(T_i, d_i \mid \gamma, \theta, \alpha, \kappa_i, \eta_{1,i}^b, \eta_{1,i}^a, w_i\right) = h_i(T_i)^{d_i} S_i(T_i), \tag{50}$$

where we suppress the dependence on the parameters on the right hand side for brevity.

Our priors, which we justify in Appendix G, for the submodel specific parameters are $\gamma \sim$ Gamma(9.05, 8.72), $\alpha \sim$ SkewNormal(0, 0.5, −2), $\theta_a \sim$ SkewNormal(0, 0.5, −1), and $\theta_0 \sim$ N($\hat{E}$, 0.5^2) where $\hat{E}$ is the log of the crude event rate (Brilleman et al., 2020). We adopt the same priors as the cumulative fluid balance submodel for κ_i, $\eta_{1,i}^b$, and $\eta_{1,i}^a$.

5.4 Chained Markov melding details

To combine the submodels with chained Markov melding we must define the common quantities $\phi_{1\cap2}$ and $\phi_{2\cap3}$. We meld p$_1$ and p$_2$ by treating the derived event times and indicators $\{(T_i, d_i)\}_{i=1}^N$ under p$_1$ as the "response", i.e. event times, in p$_2$. Care is required

when defining $\phi_{1\cap2}$ under p_1 as it is a deterministic function of $\beta_{0,i}$ and ζ_i. Define $\chi_{1,i}$ $= (\beta_{0,i}, \zeta_i)$ and $\phi_{1\cap2,i} = f(\chi_{1,i}) = (T_i, d_i)$, where f is the output from attempting to solve Equation (46), so that $\phi_{1\cap2} = (f(\chi_{1,i}), \ldots, f(\chi_{1,N}))$. The parameters shared by Equations (47) and (49) constitute $\phi_{2\cap3} = (\eta_{1,i}^b, \eta_{1,i}^a, \kappa_i)_{i=1}^N$.

To completely align with our chained melding notation we also define, for the P/F submodel, $Y_1 = (z_i, t_i)_{i=1}^N$ and $\psi_1 = (\omega_i)_{i=1}^N$, noting that ψ_1 and $(\chi_{1,i}, \ldots, \chi_{1,N})$ have no components in common. For the cumulative fluid submodel we define $Y_3 = (x_i, u_i)_{i=1}^N$, and $\psi_3 = (\eta_{0,i}, \sigma_{x,i}^2)_{i=1}^N$. Finally, for the survival submodel we define $Y_2 = (w_i)_{i=1}^N$ and $\psi_2 = (\gamma, \theta, a)$.

5.5 Pooling and estimation

We consider logarithmic pooling with $\lambda = \left(\frac{4}{5}, \frac{4}{5}, \frac{4}{5}\right)$ (any smaller value of λ results in a prior that is so uninformative that it causes computational problems) and with $\lambda = (1, 1, 1)$ (Product-of-Experts). Because the correlation between $\phi_{1\cap2}$ and $\phi_{2\cap3}$ in $p_2(\phi_{1\cap2}, \phi_{2\cap3})$ is important, we do not consider linear pooling in this example. Logarithmic pooling requires estimates of $p_1(\phi_{1\cap2})$ and $p_2(\phi_{1\cap2}, \phi_{2\cap3})$. Because these are mixed distributions, with both discrete and continuous components, standard kernel density estimation, as suggested by Goudie et al. (2019), is inappropriate. Instead we fit, to transformed versions of $\phi_{1\cap2}$ and $\phi_{2\cap3}$, a mixture containing a discrete component and either a Gaussian or beta distribution, depending on the transformation. Further details for all the mixture distribution estimates are contained in Appendix H.

We use the parallel multi-stage sampler with $p_{\text{pool},1}(\phi_{1\cap2}) = p_1(\phi_{1\cap2})$, $p_{\text{pool},3}(\phi_{2\cap3}) = p_3(\phi_{2\cap3})$ and $p_{\text{pool},2}(\phi_{1\cap2}, \phi_{2\cap3}) = p_{\text{pool}}(\phi) / (p_1(\phi_{1\cap2})p_3(\phi_{2\cap3}))$. That is, in stage one we target the subposteriors $p_1(\phi_{1\cap2}, \psi_1 / Y_1)$ and $p_3(\phi_{2\cap3}, \psi_3 / Y_3)$; in stage two we target the full melded model. Targeting $p_1(\phi_{1\cap2}, \psi_1 / Y_1)$ in stage one alleviates the need to solve Equation (46) within an MCMC iteration, instead turning the production of $\phi_{1\cap2}$ into an embarrassingly parallel, post-stage-one processing step. Attempting to sample the melded posterior directly would involve solving (46) many times within each iteration, presenting a sizeable computational hurdle which we avoid. It is crucial for the convergence of our multi-stage sampler that the components of $\phi_{1\cap2}$ and $\phi_{2\cap3}$ are updated *individual-at-a-time* in stage two. This is possible due to the conditional independence between individuals in the stage one posterior, and Appendix K contains the details of this scheme. The stage one subposteriors are sampled using Stan, using 5 chains with 10^3 warm-up iterations and 10^4 post warm-up iterations. We use Stan to sample ψ_2 where, in every MH-within-Gibbs step, we run Stan for 9 warm-up iterations and 1 post warm-up iteration[7]. We run 5 chains of 10^4 iterations for all stage two targets. Visual and numerical diagnostics (Vehtari et al., 2020) are assessed and are available in the repository accompanying this paper[8].

---

[7]We also initialise Stan at the previous value of ψ_2, and disable all adaptive procedures as the default (identity) mass matrix and step size are suitable for this example.
[8] https://doi.org/10.5281/zenodo.6552714

5.6 Results

We first inspect the subposterior fitted values for p_1 and p_3. The top row of Figure 7 displays the P/F data, the fitted submodel, and derived event times for individuals $i = 17$ and 29. The spline appears to fit the raw P/F data well, with the heavy tailed error term accounting for the larger deviations away from the fitted value. It is interesting to see the relatively wide, multimodal distribution for (T_{29}, d_{29}) (there is a second mode at $(T_{29} = C_{29}, d_{29} = 0)$ and for other individuals not shown here). The bottom row of Figure 7 displays the cumulative fluid data and the fitted submodel, with the little noise in the data resulting in minimal uncertainty about the fitted value and a concentrated subposterior distribution.

To assess the importance of fully accounting for the uncertainty in $\phi_{1 \cap 2}$ and $\phi_{2 \cap 3}$, we compare the posterior for ψ_2 obtained using the chained melding approach with the posterior obtained by fixing $\phi_{1 \cap 2}$ and $\phi_{2 \cap 3}$. Plugging in a point estimate reflects common applied statistical practice when combining submodels, particularly when a distributional approximation is difficult to obtain (as it is for $p_1(\phi_{1 \cap 2} / Y_1)$). Additionally, standard survival models and software typically do not permit uncertainty in event times and indicators, rendering such a plug-in approach necessary.

Specifically, we fix $\phi_{1 \cap 2}$ to the median value[9] for each individual under $p_1(\phi_{1 \cap 2} / Y_1)$ and denote it $\hat{\phi}_{1 \cap 2}$ and use the subposterior mean of $p_3(\phi_{2 \cap 3} / Y_3)$ denoted $\hat{\phi}_{2 \cap 3}$. With these fixed values we sample $p(\psi_2 \mid \hat{\phi}_{1 \cap 2}, \hat{\phi}_{2 \cap 3}, Y_2)$. We also compare the melded posterior to the submodel marginal prior $p_2(\psi_2)$, but we note that this comparison is difficult to interpret, as the melding process alters the prior for ψ_2. Figure 8 displays the aforementioned densities for $(\theta_3, \theta_{17}, \gamma, a) \subset \psi_2$, with (θ_3, θ_{17}) chosen as they exhibit the greatest sensitivity to the fixing of $\phi_{1 \cap 2}$ and $\phi_{2 \cap 3}$. For the baseline coefficients (θ_3, θ_{17}) the chained melding posterior differs slightly in location from $p(\psi_2 \mid \hat{\phi}_{1 \cap 2}, \hat{\phi}_{2 \cap 3}, Y_2)$ with a small increase in uncertainty. A more pronounced change is visible for a, where the melding process has added a notable degree of uncertainty and shifted the posterior leftwards.

To investigate which part of the melding process causes this change in the posterior of a, we consider fixing either one of $\phi_{1 \cap 2}$ and $\phi_{2 \cap 3}$ to their respective point estimates. That is, we employ Markov melding as described in Section 1.2, using either logarithmic or PoE pooling, to obtain $p_{meld}(a \mid \hat{\phi}_{1 \cap 2}, Y_2, Y_3)$ and $p_{meld}(a \mid \hat{\phi}_{2 \cap 3}, Y_1, Y_2)$. Figure 9 displays the same distributions for a as Figure 8, and adds the posteriors obtained using one fixed value $(\hat{\phi}_{1 \cap 2}\, \text{or}\, \hat{\phi}_{2 \cap 3})$ whilst melding the other non-fixed parameter.

Evident for both choices of pooling is the importance of incorporating the uncertainty in $\phi_{1 \cap 2}$. This is expected given the large uncertainty and multimodal nature of $\phi_{1 \cap 2}$ compared to $\phi_{2 \cap 3}$ (see Figure 7). We suspect that it is the multimodality in $p_1(\phi_{1 \cap 2} / Y_1)$ that produces the shift in posterior mode of $\phi_{1 \cap 2}$, with the width of $p_1(\phi_{1 \cap 2} / Y_1)$ affecting the increase in uncertainty. Because we prefer the chained melded posterior, under either

---

[9]For each individual the samples of $(T_i, d_i)_{i=1}^{N}$-pairs are sorted by T_i, and the $\left\lfloor \frac{N}{2} \right\rfloor^{\text{th}}$ tuple $(\hat{T}_i, \hat{d}_i)$ is chosen as the median.

pooling method, for its full accounting of uncertainty we conclude that $p\left(\alpha \mid \hat{\phi}_{1 \cap 2}, \hat{\phi}_{2 \cap 3}, Y_2\right)$ is both overconfident and biased.

The marginal changes to the components of ψ_2 visible in Figure 8 appear small, however the cumulative effect of such changes becomes apparent when inspecting the posterior of the survival function. Figure 10 displays the model-based, mean survival function under the melded posterior (using PoE pooling), and corresponding draws of $\phi_{1 \cap 2}$ converted into survival curves using the Kaplan-Meier estimator. Also shown are the Kaplan-Meier estimate of $\hat{\phi}_{1 \cap 2}$ and the mean survival function computed using $p\left(\psi_2 \mid \hat{\phi}_{1 \cap 2}, \hat{\phi}_{2 \cap 3}, Y_2\right)$. The posterior survival functions differ markedly, with the 95% intervals overlapping only for small values of time. It is also interesting to see that $\hat{\phi}_{1 \cap 2}$, despite being a reasonable point estimate of $p_1(\phi_{1 \cap 2} / Y_1)$, is not very likely under the melded posterior. Figure 10 also suggests that the Weibull hazard is insufficiently flexible for this example. We discuss the complexities of other hazards in Section 6.

6   Conclusion

This paper introduces the chained Markov melded model. In doing so we make explicit the notion of submodels related in a chain-like way, describe a generic methodology for joining together any number of such submodels and illustrate its application with our examples. Our examples also demonstrate the importance of quantifying the uncertainty when joining submodels; not doing so can produce biased, over-confident inference. We also present the choices, and their impacts, that users of chained Markov melding must make which include: the choice of pooling function, and where required the pooling weights; the choice of posterior sampler and the design thereof, including the apportionment of the pooled prior over the stages and stage-specific MCMC techniques.

We have introduced extensions to linear and logarithmic pooling to marginals of different but overlapping quantities. Linear pooling, introduced in Section 1.2, could be extended to induce dependence between the components of ϕ using multivariate or vine copulas (Kurowicka and Joe, 2011; Nelsen, 2006), or other techniques (Lin et al., 2014). Copula methods are particularly appealing as, depending on the choice of copula, they yield computationally cheap to evaluate expressions for the density function, are easy to sample, and induce correlation between an arbitrary number of marginals.

Our parallel multi-stage sampler currently only considers $M = 3$ submodels, rather than the fully generic definition of chained Markov melding in Equation (10). Whilst we anticipate needing more complex methods in large M settings, the value of M at which the performance of our multi-stage sampler becomes unacceptable will depend on the specific submodels and data under consideration. A general method would consider a large and arbitrary number of submodels in a chain, and initially split the chain into more/fewer pieces depending on the computational resources available. Designing such a method is complex, as it would have to:

- avoid requiring the inverse of any component of ϕ with a noninvertible definition,

- estimate the relative cost of sampling each submodel's subposterior, to split the chain of submodels into steps/jobs of approximately the same computational cost,

- decide the order in which pieces of the chain are combined.

These are substantial challenges. It may be possible to use combine the ideas in Lindsten et al. (2017) and Kuntz et al. (2021), who propose a parallel Sequential Monte Carlo method, with the aforementioned constraints to obtain a generic methodology. Ideally we would retain the ability to use existing implementations of the submodels, however the need to recompute the weights of the particles, and hence reevaluate previously considered submodels, may preclude this requirement. Our current sampler is also sensitive to large differences in location or scale of the target distribution between the stages. The impact of these differences can be ameliorated using the methodology of Manderson and Goudie (2022), and, more generally, Sequential Monte Carlo samplers are likely to perform better in these settings.

Our chained Markov melding methodology is general and permits any form of uncertainty in the common quantities. In Section 5 we use our chained melded model to incorporate uncertainty in the event times and indicators into a survival submodel. Some specific forms of uncertainty in the event times have been considered in previous work. These include Wang et al. (2020), who consider uncertain event times arising from record linkage, where the event time is assumed to be one of a finite number of event times arising from the record linkage; and Giganti et al. (2020), Oh et al. (2018), and Oh et al. (2021), who leverage external validation data to account for measurement error in the event time. However, the general and Bayesian nature of our methodology readily facilitates any form of uncertainty in the event times and the event indicators; uncertainty in the latter is not considered in the cited papers.

The example in Section 5 has three more interesting aspects to discuss. Firstly, the P/F ratio data used in the first submodel is obtained by finding all blood gas measurements from arterial blood samples. Approximately 20% of the venous/arterial labels are missing. In these instances a logistic regression model, fit by the MIMIC team[10], is used to predict the missing label based on other covariates. It is theoretically possible to refit the model in a Bayesian framework and use the chained melded model to incorporate the uncertainty in the predicted sample label – adding another 'link' to the chain.

Secondly, the application of our multi-stage sampler to this example is similar to the two-stage approach used for joint longitudinal and time-to-event models (see Mauff et al., 2020, for a description of this approach). In the two-stage approach, the longitudinal model is fit using standard MCMC methods in stage one, and the samples are reused in stage two when considering the time-to-event data. This can significantly reduce the computational effort required to fit the joint model. However, unlike our multi-stage sampler, the typical

---

[10] The coefficients, classification threshold, and the imputation used in the case of missing data are supplied in the `blood-gasses.sql` file in the GitHub repository accompanying this paper. No other information is available about this model (the data used to produce the coefficients, and the performance of the fitted model).

two-stage approach does not target the full posterior distribution, which can lead to biased estimates (though Mauff et al. (2020) extend the typical two-stage approach to reduce this bias).

Thirdly, we observe a lack of flexibility the baseline hazard, visible in Figure 10. More complex hazards could be employed, e.g. modelling the (log-)hazard using a (penalised) B-spline (Royston and Parmar, 2002; Rosenberg, 1995; Rutherford et al., 2015). However, this increased flexibility precludes an analytic form for the survival function. Whilst numerical integration is possible it is not trivial, particularly when the hazard is discontinuous, as our hazard is at the breakpoint. Splines also have more coefficients than the single parameter of the Weibull hazard. Identifiability issues arise with a small number of individuals, many of whom are censored, and are compounded when there are a relatively large number of other parameters (α, θ). Whilst we do not believe these costs are worth incurring for our example, for settings with a larger number of patients and more complicated longitudinal submodels the increased flexibility may be vital.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Abadi F, Gimenez O, Ullrich B, Arlettaz R, Schaub M. Estimation of Immigration Rate Using Integrated Population Models. Journal of Applied Ecology. 2010; 47 (2) 393–400.

Abbas AE. A Kullback-Leibler View of Linear and Log-Linear Pools. Decision Analysis. 2009.

Ades AE, Sutton AJ. Multiparameter Evidence Synthesis in Epidemiology and Medical Decision-Making: Current Approaches. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2006; 169 (1) 5–35.

Belgorodski N, Greiner M, Tolksdorf K, Schueller K. rriskDistributions: Fitting Distributions to given Data or Known Quantiles. R package version 2.1.2. 2017.

Besbeas P, Freeman SN, Morgan BJT, Catchpole EA. Integrating Mark–Recapture–Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. Biometrics. 2002; 58 (3) 540–547. [PubMed: 12229988]

Brilleman S. Simsurv: Simulate Survival Data. R package version 1.0.0. 2021.

Brilleman SL, Elci EM, Novik JB, Wolfe R. Bayesian Survival Analysis Using the Rstanarm R Package. 2020. arXiv:2002.09633

Bromiley P. Products and Convolutions of Gaussian Probability Density Functions. Tina-Vision Memo. 2003; 3 (4) 1.

Brooks SP, King R, Morgan BJT. A Bayesian Approach to Combining Animal Abundance and Demographic Data. Animal Biodiversity and Conservation. 2004; 27 (1)

Burke DL, Ensor J, Riley RD. Meta-Analysis Using Individual Participant Data: One-Stage and Two-Stage Approaches, and Why They May Differ. Statistics in Medicine. 2017; 36 (5) 855–875. [PubMed: 27747915]

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017; 76 (1) 1–32. [PubMed: 36568334]

Carvalho LM, Villela DAM, Coelho FC, Bastos LS. Bayesian Inference for the Weights in Logarithmic Pooling. Bayesian Analysis. 2022. 1–29.

Crowther MJ, Lambert PC. Simulating Biologically Plausible Complex Survival Data. Statistics in Medicine. 2013; 32 (23) 4118–4134. [PubMed: 23613458]

Dawid AP, Lauritzen SL. Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models. The Annals of Statistics. 1993; 21 (3) 1272–1317.

de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. Programming with Models: Writing Statistical Algorithms for General Model Structures with NIMBLE. Journal of Computational and Graphical Statistics. 2017; 26 (2) 403–413.

Donnat, C; Miolane, N; Bunbury, F; Kreindler, J. A Bayesian Hierarchical Network for Combining Heterogeneous Data Sources in Medical Diagnoses; Proceedings of the Machine Learning for Health NeurIPS Workshop; 2020. 53–84. Proceedings of Machine Learning Research

Finke A, King R, Beskos A, Dellaportas P. Efficient Sequential Monte Carlo Algorithms for Integrated Population Models. Journal of Agricultural, Biological and Environmental Statistics. 2019; 24 (2) 204–224.

Gabry J, Mahr T, Bürkner P-C, Modrák M, Barrett M, Weber F, Sroka EC, Vehtari A. Bayesplot: Plotting for Bayesian Models. R package version 1.8.0. 2021.

Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian Workflow. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2019; 182 (2) 389–402.

Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P-C, Modrák M. Bayesian Workflow. 2020. arXiv:2011.01808

Genest C, McConway KJ, Schervish MJ. Characterization of Externally Bayesian Pooling Operators. The Annals of Statistics. 1986; 14 (2) 487–501.

Giganti MJ, Shaw PA, Chen G, Bebawy SS, Turner MM, Sterling TR, Shepherd BE. Accounting for Dependent Errors in Predictors and Time-to-Event Outcomes Using Electronic Health Records, Validation Samples, and Multiple Imputation. Annals of Applied Statistics. 2020; 14 (2) 1045– 1061. [PubMed: 32999698]

Goudie RJB, Presanis AM, Lunn D, De Angelis D, Wernisch L. Joining and Splitting Models with Markov Melding. Bayesian Analysis. 2019; 14 (1) 81–109. [PubMed: 30631389]

Hastie, T, Tibshirani, R. Generalized Additive Models. Boca Raton, Fla: Chapman & Hall/CRC; 1999.

Hinton GE. Training Products of Experts by Minimizing Contrastive Divergence. Neural Computation. 2002; 14 (8) 1771–1800. [PubMed: 12180402]

Hooten MB, Johnson DS, Brost BM. Making Recursive Bayesian Inference Accessible. The American Statistician. 2019. 1–10.

Jackson D, White IR. When Should Meta-Analysis Avoid Making Hidden Normality Assumptions? Biometrical Journal. 2018; 60 (6) 1040–1058. [PubMed: 30062789]

Johnson AEW, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG. MIMIC-III, a Freely Accessible Critical Care Database. Scientific Data. 2016; 3 (1) 160035 [PubMed: 27219127]

Kay M. Tidybayes: Tidy Data and Geoms for Bayesian Models. R package version 2.0.2. 2020.

Kedem, B, De Oliveira, V, Sverchkov, M. Statistical Data Fusion. World Scientific; 2017.

Kharratzadeh M. Splines in Stan. Stan Case Studies. 2017; 4

Kuntz J, Crucinio FR, Johansen AM. The Divide-and-Conquer Sequential Monte Carlo Algorithm: Theoretical Properties and Limit Theorems. 2021. arXiv:2110.15782

Kurowicka, D, Joe, H, editors. Dependence Modeling: Vine Copula Handbook. Singapore: World Scientific; 2011.

Lahat D, Adali T, Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. Proceedings of the IEEE. 2015; 103 (9) 1449–1477.

Lauritzen SL, Richardson TS. Chain Graph Models and Their Causal Interpretations. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64 (3) 321–348.

Lebreton J-D, Burnham KP, Clobert J, Anderson DR. Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies. Ecological Monographs. 1992; 62 (1) 67–118.

Lin G, Dou X, Kuriki S, Huang J-S. Recent Developments on the Construction of Bivariate Distributions with Fixed Marginals. Journal of Statistical Distributions and Applications. 2014; 1 (1) 14.

Lindsten F, Johansen AM, Naesseth CA, Kirkpatrick B, Schön TB, Aston JAD, Bouchard-Côté A. Divide-and-Conquer with Sequential Monte Carlo. Journal of Computational and Graphical Statistics. 2017; 26 (2) 445–458.

Lu CJ, Meeker WQ. Using Degradation Measures to Estimate a Time-to-Failure Distribution. Technometrics. 1993; 35 (2) 161–174.

Lunn D, Barrett J, Sweeting M, Thompson S. Fully Bayesian Hierarchical Modelling in Two Stages, with Application to Meta-Analysis. Journal of the Royal Statistical Society Series C. 2013; 62 (4) 551–572.

Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS Project: Evolution, Critique and Future Directions. Statistics in Medicine. 2009; 28 (25) 3049–3067. [PubMed: 19630097]

Manderson AA, Goudie RJB. A Numerically Stable Algorithm for Integrating Bayesian Models Using Markov Melding. Statistics and Computing. 2022; 32 (2) 24. [PubMed: 35310545]

Massa, MS, Lauritzen, SL. Algebraic Methods in Statistics and Probability II. Vol. 516. Amer. Math. Soc; Providence, RI: 2010. 239–259. Contemp Math

Mauff K, Steyerberg E, Kardys I, Boersma E, Rizopoulos D. Joint Models with Multiple Longitudinal Outcomes and a Time-to-Event Outcome: A Corrected Two-Stage Approach. Statistics and Computing. 2020; 30 (4) 999–1014.

Maunder MN, Punt AE. A Review of Integrated Analysis in Fisheries Stock Assessment. Fisheries Research. 2013; 142: 61–74.

Meng, X-L. Past, Present, and Future of Statistical Science. Lin, X, Genest, C, Banks, DL, Molenberghs, G, Scott, DW, Wang, J-L, editors. Chapman and Hall/CRC; 2014. 561–586.

Nelsen, RB. An Introduction to Copulas. second edition. Springer; New York: 2006.

Nicholson G, Blangiardo M, Briers M, Diggle PJ, Fjelde TE, Ge H, Goudie RJB, Jersakova R, King RE, Lehmann BCL, Mallon A-M, et al. Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality. Statistical Science. 2021.

NIMBLE Development Team. NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling. R package manual version 0.9.0. 2019.

Oh EJ, Shepherd BE, Lumley T, Shaw PA. Considerations for Analysis of Time-to-Event Outcomes Measured with Error: Bias and Correction with SIMEX. Statistics in medicine. 2018; 37 (8) 1276–1289. [PubMed: 29193180]

Oh EJ, Shepherd BE, Lumley T, Shaw PA. Raking and Regression Calibration: Methods to Address Bias from Correlated Covariate and Time-to-Event Error. Statistics in Medicine. 2021; 40 (3) 631–649. [PubMed: 33140432]

O'Hagan, A, Buck, C, Daneshkhah, A, Eiser, J, Garthwaite, P, Jenkinson, D, Oakley, J, Rakow, T. Statistics in Practice. Wiley; 2006.

Parsons J, Niu X, Bao L. A Bayesian Hierarchical Modeling Approach to Combining Multiple Data Sources: A Case Study in Size Estimation. 2021. arXiv:2012.05346

Plummer M. Rjags: Bayesian Graphical Models Using MCMC. R package version 4-10. 2019.

Presanis AM, Pebody RG, Birrell PJ, Tom BDM, Green HK, Durnall H, Fleming D, De Angelis D. Synthesising Evidence to Estimate Pandemic (2009) A/H1N1 Influenza Severity in 2009-2011. Annals of Applied Statistics. 2014; 8 (4) 2378–2403.

Rizopoulos, D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. CRC Press; 2012.

Rosenberg PS. Hazard Function Estimation Using B-splines. Biometrics. 1995; 51 (3) 874–887. [PubMed: 7548706]

Royston P, Parmar MKB. Flexible Parametric Proportional-Hazards and Proportional-Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects. Statistics in Medicine. 2002; 21 (15) 2175–2197. [PubMed: 12210632]

Rufo MJ, Martín J, Pérez CJ. Log-Linear Pool to Combine Prior Distributions: A Suggestion for a Calibration-Based Approach. Bayesian Analysis. 2012a; 7 (2) 411–438.

Rufo MJ, Párez CJ, Martín J. A Bayesian Approach to Aggregate Experts' Initial Information. Electronic Journal of Statistics. 2012b; 6: 2362–2382.

Rutherford MJ, Crowther MJ, Lambert PC. The Use of Restricted Cubic Splines to Approximate Complex Hazard Functions in the Analysis of Time-to- Event Data: A Simulation Study. Journal of Statistical Computation and Simulation. 2015; 85 (4) 777–793.

Schaub M, Abadi F. Integrated Population Models: A Novel Analysis Framework for Deeper Insights into Population Dynamics. Journal of Ornithology. 2011; 152 (1) 227–237.

Schaub M, Ullrich B, Knötzsch G, Albrecht P, Meisser C. Local Population Dynamics and the Impact of Scale and Isolation: A Study on Different Little Owl Populations. Oikos. 2006; 115 (3) 389–400.

Seethala RR, Hou PC, Aisiku IP, Frendl G, Park PK, Mikkelsen ME, Chang SY, Gajic O, Sevransky J. Early Risk Factors and the Role of Fluid Administration in Developing Acute Respiratory Distress Syndrome in Septic Patients. Annals of Intensive Care. 2017; 7 (1) 11. [PubMed: 28116595]

Soetaert K, Hindmarsh AC, Eisenstat SC, Moler C, Dongarra J, Saad Y. Rootsolve: Nonlinear Root Finding, Equilibrium and Steady-State Analysis of Ordinary Differential Equations. R package version 1.8.2.1. 2020.

Stan Development Team. RStan: The R Interface to Stan. R package version 2.26. 2021.

The ARDS Definition Task Force. Acute Respiratory Distress Syndrome: The Berlin Definition. JAMA. 2012; 307 (23) 2526–2533. [PubMed: 22797452]

Tom JA, Sinsheimer JS, Suchard MA. Reuse, Recycle, Reweigh: Combating Influenza through Efficient Sequential Bayesian Computation for Massive Data. The Annals of Applied Statistics. 2010; 4 (4) 1722–1748. [PubMed: 26681992]

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-Normalization, Folding, and Localization: An Improved R for Assessing Convergence of MCMC. Bayesian Analysis. 2020.

Wang W, Aseltine R, Chen K, Yan J. Integrative Survival Analysis with Uncertain Event Times in Application to a Suicide Risk Study. Annals of Applied Statistics. 2020; 14 (1) 51–73.

Wang W, Yan J. Shape-Restricted Regression Splines with R Package Splines2. Journal of Data Science. 2021; 19 (3) 498–517.

Zipkin EF, Saunders SP. Synthesizing Multiple Data Types for Biological Conservation Using Integrated Population Models. Biological Conservation. 2018; 217: 240–250.
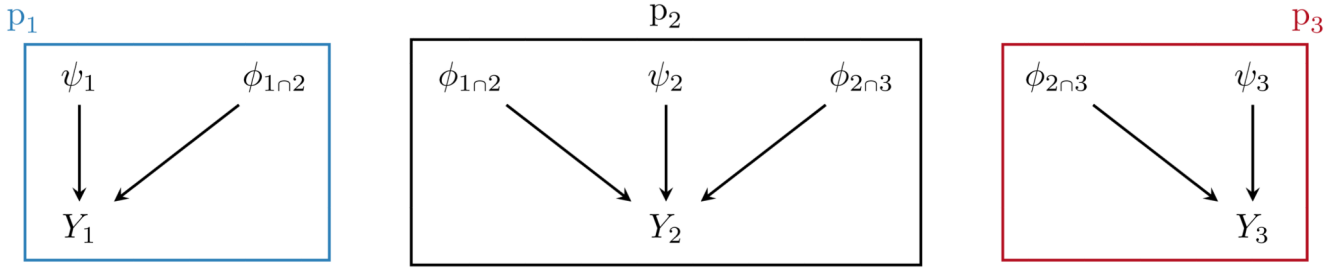
p_1 p_2 p_3

ψ_1 $\phi_{1\cap2}$ $\phi_{1\cap2}$ ψ_2 $\phi_{2\cap3}$ $\phi_{2\cap3}$ ψ_3

Y_1 Y_2 Y_3

Figure 1.
A simplified DAG of the integrated population model (IPM) for the little owls. The capture-recapture submodel (p_1) is surrounded by the blue line, the population count submodel (p_2) by the black line, and the fecundity submodel (p_3) by the red line. The capture-recapture and population count submodels share parameters affecting the juvenile and adult survival rate ($\phi_{1\cap2}$), whilst the parameter for fecundity is common to both the population count and fecundity submodels ($\phi_{2\cap3}$). The combination of all the submodels forms the IPM.
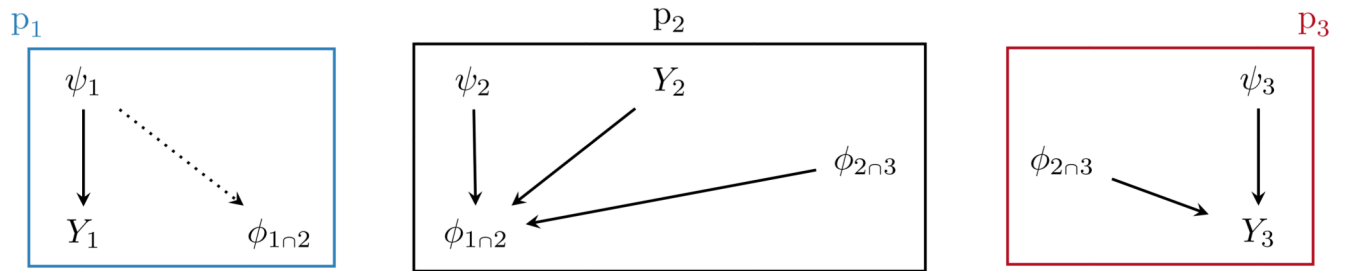
Figure 2.
A simplified DAG of the submodels considered in the survival analysis example. The event time submodel p_1 defines the event time $\phi_{1\cap2}$ as noninvertible function of the other model parameters (denoted by the dotted line), whilst the survival submodel p_2 considers $\phi_{1\cap2}$ as the response. The longitudinal submodel p_3 has parameters $\phi_{2\cap3}$ in common with the survival submodel.
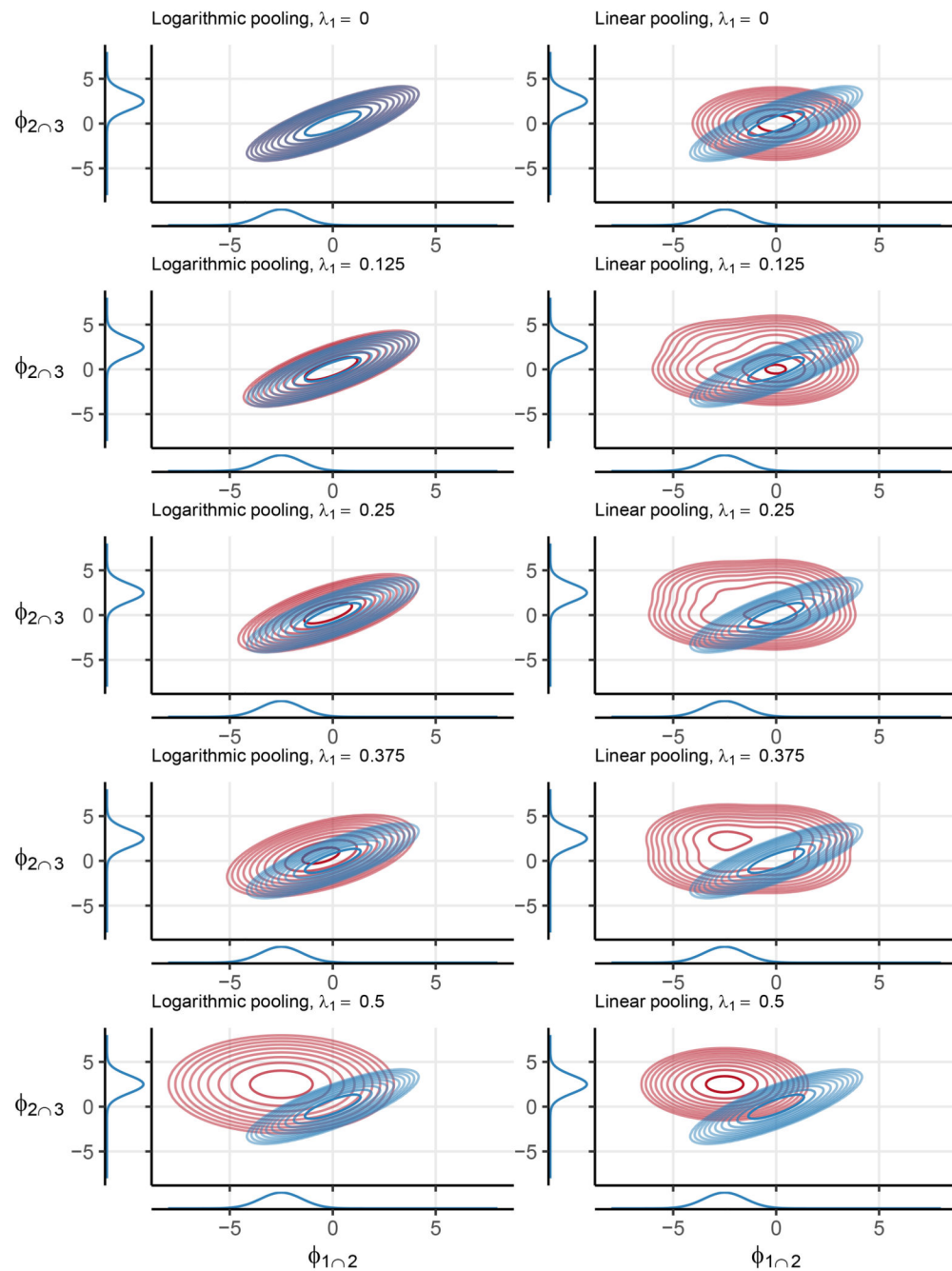
Figure 3.
Contour plots of $p_{pool}(\phi)$ (red) under logarithmic and linear pooling (left and right column respectively). The three original densities $p_1(\phi_{1\cap2})$, $p_3(\phi_{2\cap3})$ and $p_2(\phi_{1\cap2}, \phi_{2\cap3})$ are shown in blue, with the univariate densities shown on the appropriate axis. The pooling weight parameter λ_1 is indicated in the plot titles.
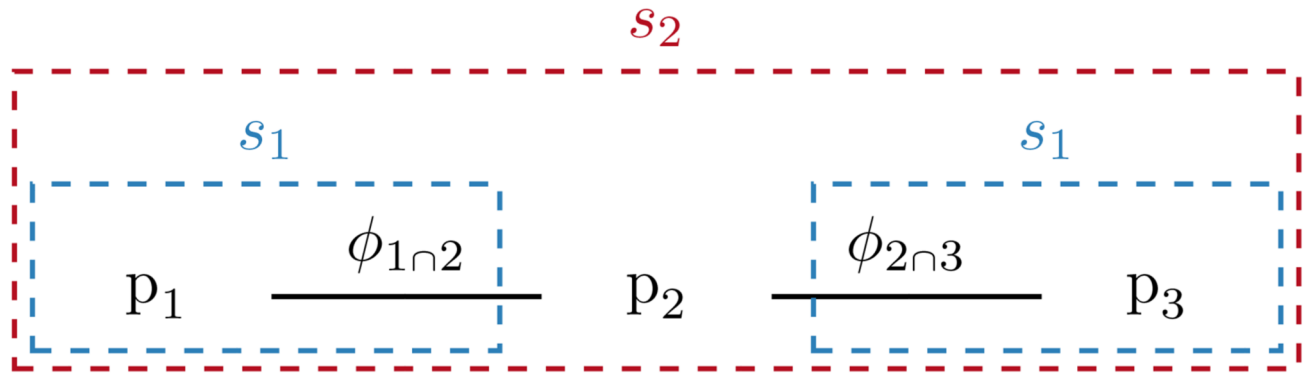
Figure 4.
A graphical depiction of the submodels and their shared quantities, with the parallel sampling strategy overlaid. The stage one (s_1) targets are surrounded by blue dashed lines, with the stage two (s_2) target in red.
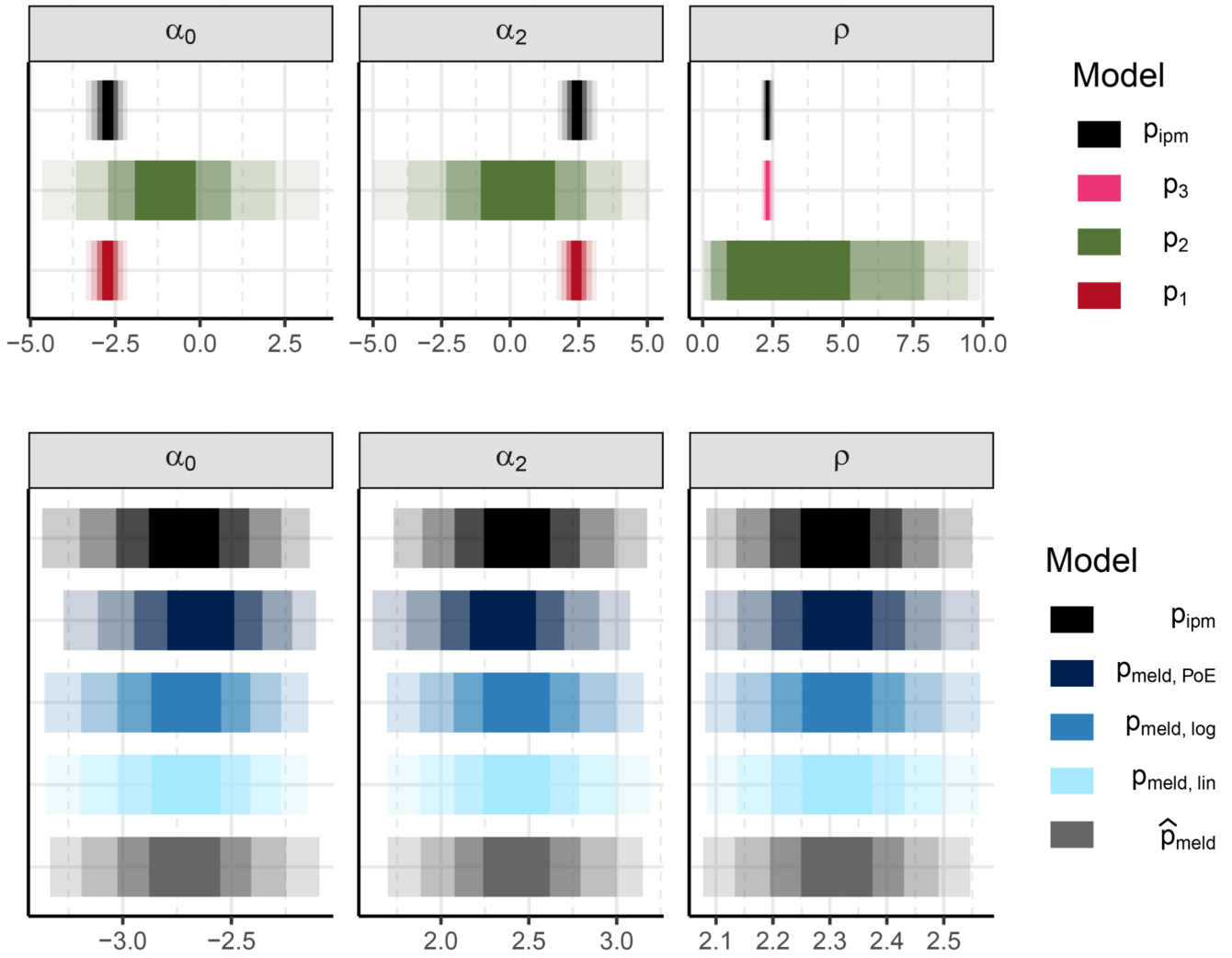
Figure 5.
Top row: credible intervals for $\phi_{1\cap2} = (\alpha_0, \alpha_2)$ and $\phi_{2\cap3} = \rho$ from the posterior of the original integrated population model p_{ipm}, and the individual subposteriors from submodels p_1, p_2, and p_3. Bottom row: credible intervals for the same quantities, but with a different x-axis scale, from the original IPM (repeated from top row); the chained melded posteriors using product-of-experts pooling, logarithmic pooling, and linear pooling denoted p_{meld}, $p_{meld,log}$ and $p_{meld,lin}$; and the melded posterior using the normal approximation $\hat{P}_{meld}$. Intervals are 50%, 80%, 95%, and 99% wide.
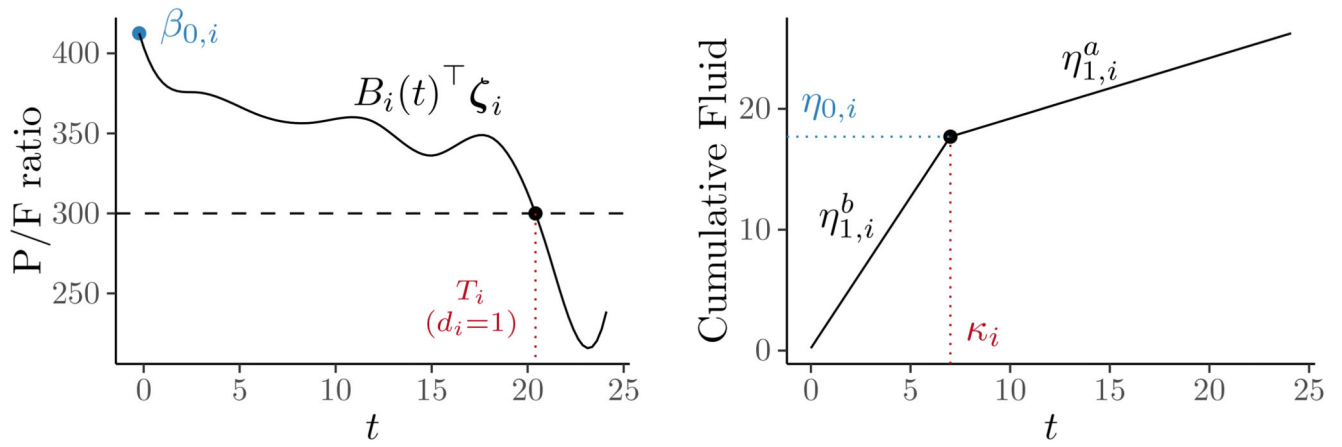
Figure 6. Parameters and form for the P/F ratio submodel (p$_1$, left) and cumulative fluid submodel (p$_3$, right).
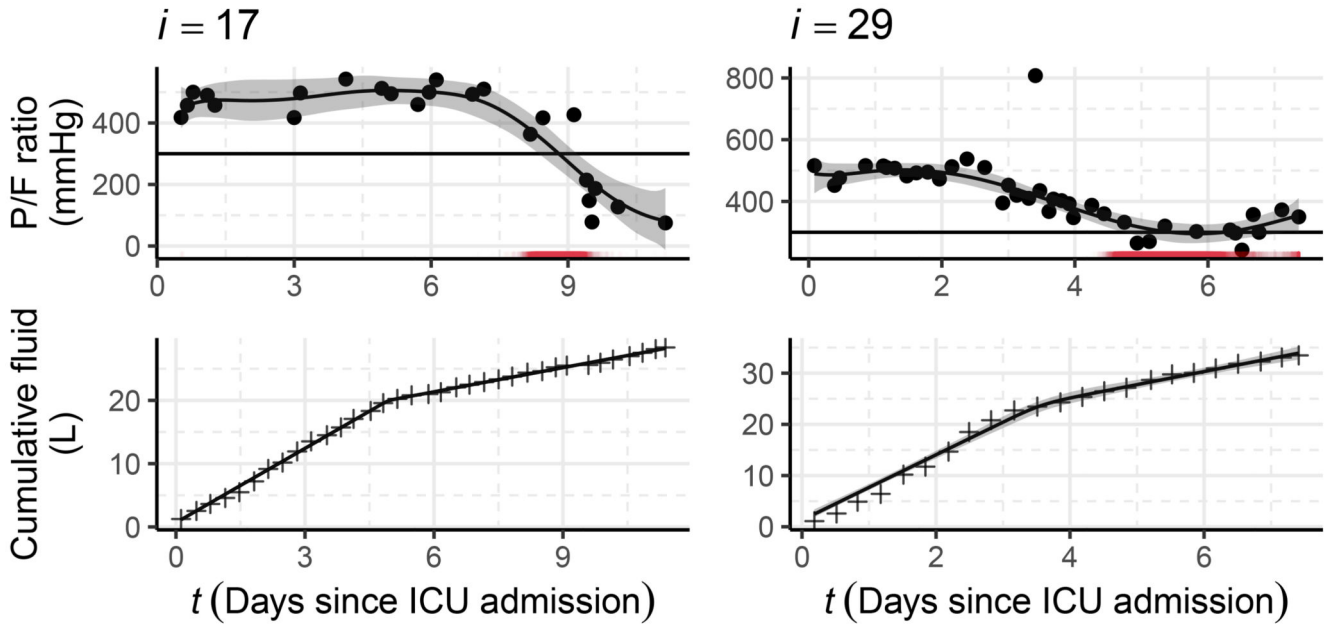
Figure 7.
The P/F ratio data (Y_1, top row); cumulative fluid data (Y_3, bottom row); subposterior means and 95% credible intervals for each of the submodels (black solid lines and grey intervals); and stage one event times (T_i, red rug in the top row) for individuals $i = 17$ and 29.
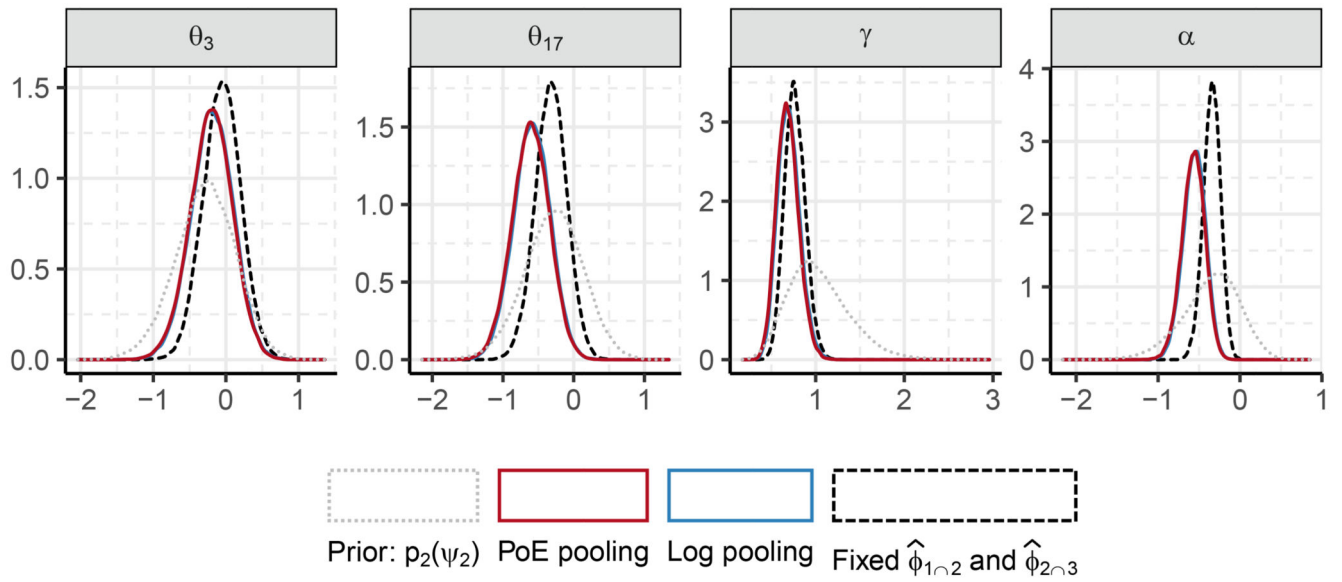
Figure 8.

Density estimates for a subset of ψ_2. The submodel marginal prior $p_2(\psi_2)$ is shown as the grey dotted line (note that this is not the marginal prior under the melded model). The figure also contains the subposteriors obtained from chained melding using PoE pooling (red, solid line) and logarithmic pooling (blue, solid line), as well as the posterior using the fixed values $p(\psi_2 \mid \hat{\phi}_{1 \cap 2}, \hat{\phi}_{2 \cap 3}, Y_2)$ (black, dashed line).
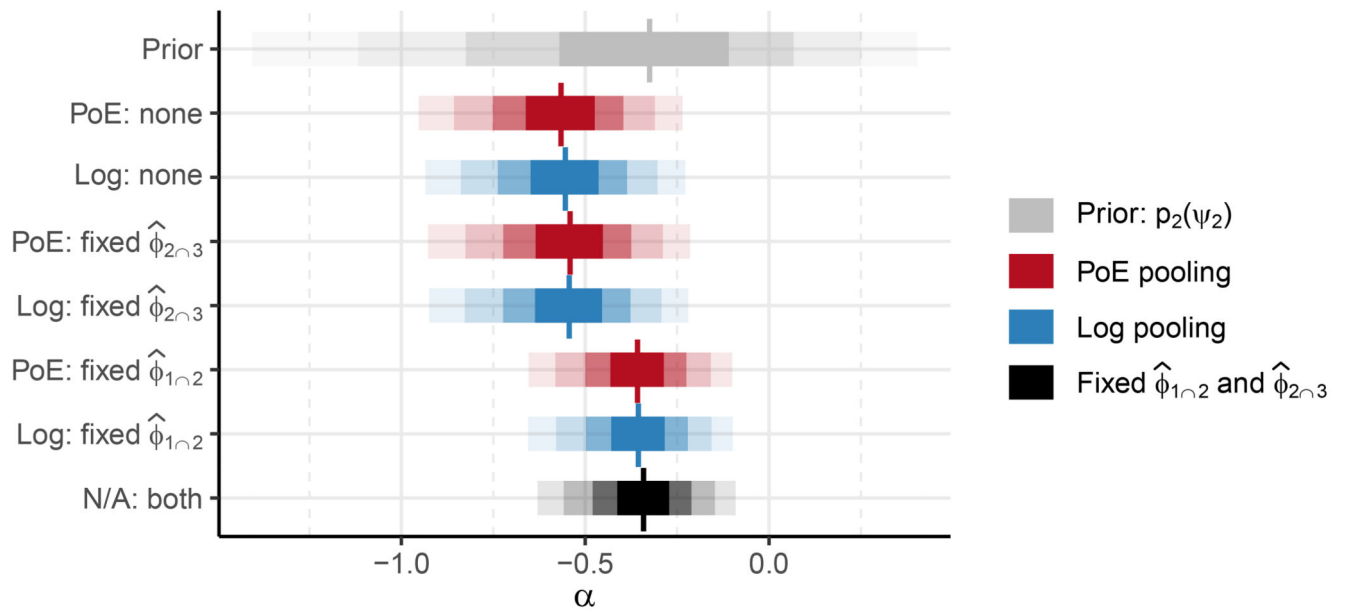
Figure 9.

Median (vertical line), 50%, 80%, 95%, and 99% credible intervals (least transparent to most transparent) for α. The marginal prior (grey, top row) and posterior using fixed $\hat{\phi}_{1 \cap 2}$ and $\hat{\phi}_{2 \cap 3}$ (black, bottom row) are as in Figure 8. For the chained melded posteriors (red and blue, rows 2 and 3) and the melded posteriors (red and blue, rows $4-7$), the tick label on the y-axis denotes the type of pooling used, and which of $\phi_{1 \cap 2}$ and/or $\phi_{2 \cap 3}$ are fixed.
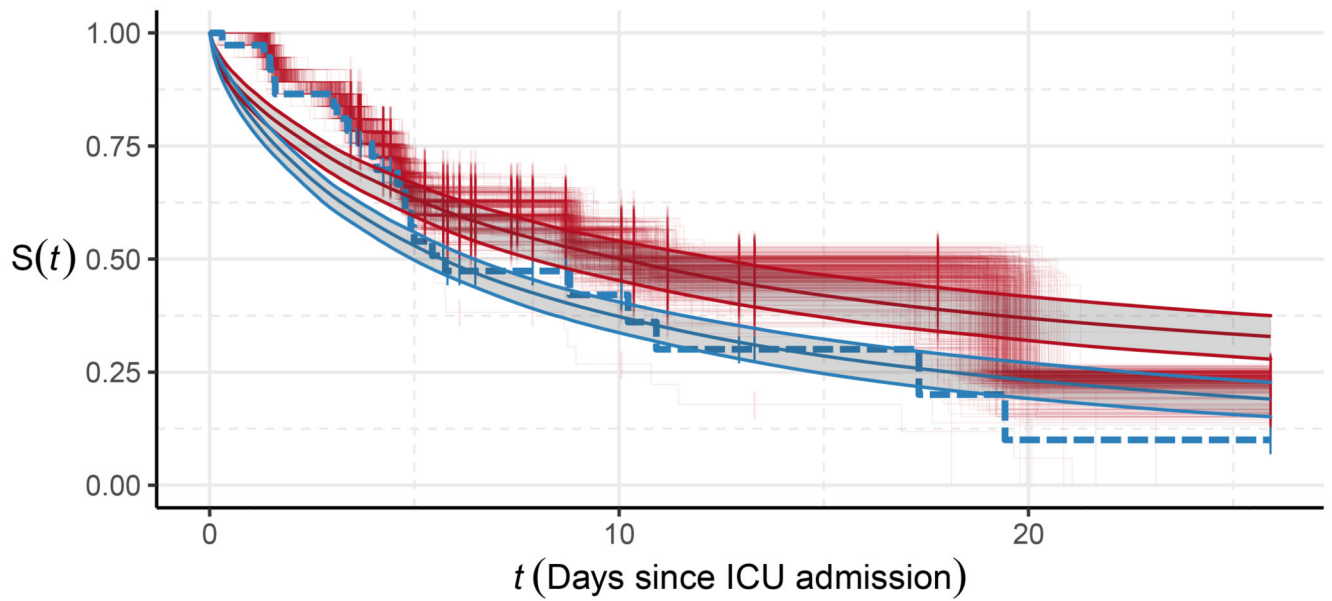
Figure 10.

Survival curves and mean survival function at time t. The red, stepped lines are draws of $\phi_{1\cap2}$ from the melded posterior using PoE pooling, converted into survival curves via the Kaplan-Meier estimator. The smooth red line and interval (posterior mean and 95% credible interval) denote the model-based, mean survival function obtained from the melded posterior (PoE pooling) values of ψ_2 and $\phi_{2\cap3}$. The blue dashed line is the Kaplan-Meier estimate of $\hat{\phi}_{1\cap2}$, and the blue solid line and interval are the corresponding model-based estimate from $p\left(\psi_2 \mid \hat{\phi}_{1\cap2}, \hat{\phi}_{2\cap3}, Y_2\right)$.