LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Andreas Bender, Fabian Scheipl, Wolfgang Hartl, Andrew G. Day, Helmut Küchenhoff

Modeling Exposure-Lag-Response Associations with Penalized Piece-wise Exponential Models

Technical Report Number 192, 2016
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Modeling Exposure-Lag-Response Associations with Penalized Piece-wise Exponential Models

Andreas Bender, Fabian Scheipl, Wolfgang Hartl, Andrew G. Day, Helmut Küchenhoff

February 3, 2017

Abstract

We propose a novel approach for the flexible modeling of exposure-lag-response associations in time-to-event data, where multiple past exposures are cumulatively associated with the hazard within a defined time window. Our method is an extension of the piece-wise exponential model and allows for the estimation of a wide variety of effects, including potentially smooth and smoothly time-varying effects as well as cumulative effects with leads and lags, taking advantage of the advanced inference methods that have recently been developed for generalized additive mixed models. Our research has been motivated by a large multi-center study with the goal of analyzing the association between artificial nutrition intake and short term survival of critically ill patients in intensive care units.

1 Introduction

Critically ill patients admitted to the intensive care unit (ICU) often undergo mechanical ventilation (MV) and the need for artificial feeding is generally recognized. The optimal timing and amount of the nutrition supplied on the other hand is unclear, to the extent of differing guidelines with respect to the recommended amount being used in Europe and North America (Singer et al., 2009; McClave et al., 2009). More recent studies have also yielded contradictory results (see Heyland et al. (2011) for an overview), which can partly be attributed to differences in study design and methodology, and more generally to the difficulty of modeling the relationship between nutritional intake and health outcomes.

Modeling the association between nutritional intake and survival is particularly challenging for several reasons: Firstly, the amount of artificial nutritional intake during the ICU stay varies daily on a per patient basis. Secondly, the effect of the received intake is likely to vary over time and hazard rates at a particular point in time may depend on multiple past intakes. Moreover, intake may have a delayed impact on the outcome (Berger and Pichard, 2012) which presumably "wears off" after some time.

In more technical terms, this implies the need for an approach that can incorporate time-dependent covariates (TDC) and model their possibly non-linear, possibly time-varying, cumulative effects on survival with lead and lag times. Here, "lag time" refers to the amount of time after exposure until the delayed impact occurs, and "lead time" refers to the amount of time after exposure until the effect vanishes. Additionally, we need to adjust for heterogeneity due to different ICUs via frailty effects and other possibly non-linear, possibly smoothly time-varying effects of confounders recorded at baseline. In previous work in this field, Berhane et al. (2008) used tensor product smooths to model the association between survival and protracted exposure to radiation. Sylvestre and Abrahamowicz (2009) presented the *weighted cumulative exposure* model, where the effect of exposure at time t is the sum of weighted past exposures and the weight function is estimated using B-Splines whose smoothness is controlled through comparison of models based on different number of interior knots with respect to the BIC. Xiao et al. (2014) applied this approach to marginal structural Cox models. Gasparrini (2014) introduced an approach based on distributed lag non-linear models and coined the term *exposure-lag-response associations* (ELRA) for the type of relationship described above, which

we will adopt in this article. However, we will use this terminology in a broader sense, to refer to cumulative effects of time-dependent covariates in general. Note that "exposure" refers to the caloric intake of the patients in the following, as the intake is provided externally by the hospital staff.

We propose a flexible, novel approach for the modeling of the aforementioned *exposure-lag-response associations* that extends previous research by allowing penalized estimation of these potentially non-linear associations, taking into account all three dimensions relevant to the effect of a time-dependent exposure: time of exposure, amount of exposure as well as the time since exposure, and thereby a subject's *exposure history*. The method, an extension of the piece-wise exponential model (PEM), is described in detail in section 2. By embedding the concept of PEMs into the framework of generalized additive mixed models (GAMM) (cf. section 2.2), we establish a flexible model class for survival analysis and ELRA in particular that inherits most of the flexible tools for modeling, estimation and validation of GAMMs. Practical usefulness of this approach is further increased due to readily available, robust and efficient software implementations of these methods (Wood, 2006a, 2011). We extend existing methodology regarding confidence intervals and testing procedures for smooth terms to derive respective measures and test statistics for ELRAs and particularly for the comparison of hazard ratios (or log hazard differences) resulting from different patterns of a TDC. Section 3 demonstrates an application of our approach to a large multi-center observational data set of almost $10,000$ critically ill patients. In section 4, we present results of an extensive simulation study to assess properties of the proposed modeling approach and to investigate its behavior under deviations from modeling assumptions. In section 5, we review the proposed method and discuss advantages as well as disadvantages of the approach and sketch further possibilities for development in this field.

2 Methods and Model

2.1 Piece-wise Exponential Models

Let

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta}), \tag{1}$$

a general proportional hazards model with $i = 1, \ldots, n$, where n is the number of subjects under study and $\mathbf{x}_{i,\cdot}' = (x_{i,1}, \ldots, x_{i,P})$ the row-vector of time-constant covariates $x_{i,p}, p = 1, \ldots P$.
A piece-wise exponential model (PEM) is obtained by partitioning the follow-up period $(0, t_{\max}]$ into J intervals with $J+1$ cut-points $0 = \kappa_0 < \ldots < \kappa_J = t_{\max}$, where $t_{\max}$ is the maximal (observed) follow-up time. The j-th interval is given by $(\kappa_{j-1}, \kappa_j]$. Assuming the hazard rate in each interval j to be constant, such that $\lambda_0(t) = \lambda_j, \forall t \in (\kappa_{j-1}, \kappa_j], t > 0$, equation (1) (in log-linear form) simplifies to

$$\log(\lambda_i(t|\mathbf{x}_i)) = \log(\lambda_j) + \mathbf{x}_i'\boldsymbol{\beta} \;\forall\; t \in (\kappa_{j-1}, \kappa_j]. \tag{2}$$

Let T_i denote the true survival time and C_i the non-informative censoring time, then $t_i := \min(T_i, C_i)$ is the observed right-censored time under risk for subject i. Given intervals $1, \ldots, J$, Whitehead (1980) and Friedman (1982) established the equivalence of model (2) and the likelihood of the Poisson GLM (3) with
(a) one observation for each interval $j = 1, \ldots, J$ under risk for each subject i,
(b) responses $y_{ij} = 1$ if $t_i \in (\kappa_{j-1}, \kappa_j] \wedge t_i = T_i$, else $y_{ij} = 0$ as event indicators for subject i for interval j, and
(c) $t_{ij} = \min(t_i - \kappa_{j-1}, \kappa_j - \kappa_{j-1})$, representing the time subject i spends under risk in interval j:

$$\log(\mathbb{E}(y_{ij}|\mathbf{x}_i)) = \log(\lambda_{ij}t_{ij}) = \log(\lambda_j) + \mathbf{x}_i'\boldsymbol{\beta} + \log(t_{ij}). \tag{3}$$

Equation (2) then follows from (3) by trivial transformation and defining $\lambda_i(t|\mathbf{x}_i) := \lambda_{ij}\; \forall\, t \in (\kappa_{j-1}, \kappa_j]$. The likelihood of model (3) is proportional to the likelihood of (2), thus the two models are equivalent with respect to the ML estimation of the model parameters $\boldsymbol{\beta}$. In practice, when fitting the respective Poisson regression model, $\log(\lambda_j)$ is incorporated in the linear predictor $\mathbf{x}_i'\boldsymbol{\beta}$ and $\log(t_{ij})$ enters as an offset. A major advantage of this model structure

is that it lends itself easily to include TDC, as a covariate can change its value in each interval. Alternatively, the interval cut-points could be chosen as the time-points at which a change in the TDC is recorded. Then (2) can be extended to $\log(\lambda_i(t|\mathbf{x}_{ij})) = \log(\lambda_j) + \mathbf{x}'_{ij}\boldsymbol{\beta}$. Additionally, time-varying effects can be incorporated by creating a TDC for time itself, e.g. by using the interval midpoints $\tilde{t} := (\kappa_j - \kappa_{j-1})/2$, and including interaction terms of selected covariates with time $\tilde{t}$, or transformations thereof, in the linear predictor.

2.2 Piece-wise Exponential Additive Model

Transitioning from the framework of generalized linear models to the framework of generalized additive mixed models (GAMM), model (3) can be further extended to include smoothly time-varying effects of time-constant or time-dependent covariates (TDC). In reference to the acronyms for piece-wise exponential models (PEM) and generalized additive models (GAM) we denote these models (4) with PAM (*p*enalized piece-wise exponential *a*dditive *m*odel). For the sake of notational simplicity, we present the model with only one TDC that we refer to as *exposure* and denote this covariate of primary interest by z. An extension to multiple ELRAs, however, is straight forward, as will be illustrated in the application example (cf. section 3). We first present the general model specification and discuss individual terms in subsequent sections.

Let $\mathcal{Z}_i(t)$ denote a subset of past exposures that affect the hazard at time t (cf. section 2.2.3 for more details), $\ell = 1, \ldots, L$ the index for different clusters and ℓ_i the cluster to which subject i belongs. We model the hazard rate λ at time t for individual i from cluster ℓ_i as:

$$\log\left(\lambda_i(t|\mathbf{x}_i, \mathcal{Z}_i(t), \ell_i)\right) = f_0(t) + \sum_{p=1}^{P} f_p(x_{i,p}, t) + g(\mathcal{Z}_i(t), t) + b_{\ell_i} \tag{4}$$

where

- $f_0(t)$ represents the baseline hazard rate (cf. section 2.2.1),
- $f_p(x_{i,p}, t)$, $p = 1, \ldots, P$, are potentially smooth non-linear and smoothly time-varying effects (cf. section 2.2.2) of time-constant covariates $\mathbf{x}_{i,p}$,

- $g(\mathcal{Z}_i(t), t)$ denotes the *exposure-lag-response association* and will be discussed in detail in section 2.2.3.

- b_{ℓ_i} is a Gaussian random effect for cluster ℓ_i.

2.2.1 Baseline hazard

In the classical definition of PEMs (3), the baseline hazard is a step function and interval-specific hazards λ_j are estimated by including dummy variables for the individual intervals in the model matrix. A disadvantage of this approach is the potentially arbitrary choice of interval cut-points (Demarqui et al., 2008). Additionally, choosing a very high number of cut-points increases the number of parameters that need to be estimated and reduces stability of the individual estimates $\hat{\lambda}_j$. Representing the baseline hazard as a regression spline over the interval mid-points (or end-points) $\tilde{t}$ ameliorates this issue. Given a sufficiently large number of spline basis functions and intervals, the hazard can be estimated flexibly and efficiently, while overfitting is avoided due to penalization (cf. section 2.3). As hazard rates in medical studies tend to change quickly in the beginning of the follow-up and become more stable towards the end of the observation period, adaptive spline smooths (Wood, 2011, p. 21) can be employed to allow the smoothness of the baseline hazard to vary over time.

2.2.2 Smooth non-linear, smoothly time-varying effects

The summands $f_p(x_{i,p}, t)$ in the second term in equation (4) represent possibly non-linear, possibly time-varying effects of time-constant covariates. In the simplest case, when effects are assumed to be linear and not time-varying, this would reduce to a linear effect $x_{i,p}\beta_p$. Time-varying effects are modeled as interaction terms between the variable of interest $\mathbf{x}_{\cdot,p}$ and time t. Table 1 shows possible representations of time-varying effects. Depending on the specification of the interaction term, flexibility can increase from linear effects with linear time-variation $\beta_p x_{i,p} + \beta_{p:t}(x_{i,p} \cdot t)$, to varying coefficients $x_{i,p} f_p(t)$ or $f_p(x_{i,p})t$ (Hastie and Tibshirani, 1993), or nonlinear, smoothly time-varying covariate effects $f_p(x_{i,p}, t)$ modeled as bivariate function surfaces, parameterized as tensor product smooths (Wood et al., 2013). The smooth functions $f_p(\cdot)$ can be represented as splines of the form $f_p(\cdot) = \sum_{m=1}^{M} \gamma_{m,p} B_{m,p}(\cdot)$, where $B_{m,p}$ are covariate specific basis functions. The specification $x_{i,p} f_p(t)$ is particularly useful when

$x_{i,p}$ is a dummy variable coding for a certain level of a categorical variable, in which case a smoothly time-varying effect $f_p(t)$ is estimated for each category. One possible application is the evaluation of the effects of different treatment arms in clinical trials when the proportional hazards assumption is not fulfilled. Specification $f_p(x_{i,p}, t)$ is the most flexible and should be

Effect specification	Description
$\beta_p x_{i,p} + \beta_{p:t}(x_{i,p} \cdot t)$:	Linear, linearly time-varying effect
$f_p(x_{i,p}) \cdot t$:	Smooth, linearly time-varying effect
$x_{i,p} \cdot f_p(t)$:	Linear, smoothly time-varying effect
$f_p(x_{i,p}, t)$:	Smooth, smoothly time-varying effect

Table 1: Overview of possible time-varying effect specifications.

employed whenever prior information or domain specific knowledge regarding the relationship is absent. However, this latter option is also the most computationally demanding. In general, due to the model definition and respective estimation routine, the amount of parameters that can be feasibly and reliably estimated is limited by number of subjects n under study and the observed censoring rate. In addition, depending on the number of such components and their specification, identifiability issues may arise, especially since, in contrast to "standard" additive regression models, time t will typically appear in multiple model terms in PAMs (4). Therefore, appropriate nesting is necessary in these cases as discussed in Wood (2006b).

2.2.3 Exposure-lag-response Associations

For the specification of the ELRA $g(\mathcal{Z}_i(t), t)$ in (4) it is important to distinguish between time at risk t and time of exposure t_e, i.e., the time at which the hazard is evaluated and the time at which the value of the TDC is observed, respectively. Note that in the following, specifying t is equivalent to specifying j, as j is the interval for which $\kappa_{j-1} < t \leq \kappa_j$.

Let $z_i(t_e)$ denote the value of the TDC at *exposure time* t_e. To model the time-varying, cumulative effects of exposure histories $\mathcal{Z}_i(t)$, we:

1. Specify a time window $\mathcal{T}_e(j)$ of exposure-times t_e for which the time-dependent covariate $z(t_e)$ is assumed to affect survival in interval j, such that the exposure-history affecting the hazard at time t is defined by

$$\mathcal{Z}_i(t) := \{z_i(t_e) : t_e \in \mathcal{T}_e(j)\}. \tag{5}$$

To specify $\mathcal{T}_e(j)$ we first define the set of intervals j that can be affected by exposure at time t_e

$$\mathcal{J}(t_e) := \{(\kappa_{j-1}, \kappa_j] : \kappa_j \leq t_e + t_{\text{lag}} + t_{\text{lead}} \wedge \kappa_{j-1} > t_e + t_{\text{lag}}\}$$

where t_{lag} is the delay before exposure at time t_e can affect the hazard and t_{lead} is the maximal time for which the exposure still affects the hazard after being observed. Then, $\mathcal{T}_e(j)$ is defined as

$$\mathcal{T}_e(j) := \{t_e : (\kappa_{j-1}, \kappa_j] \in \mathcal{J}(t_e)\}. \tag{6}$$

This definition is explained more intuitively in section 3.2 of the application example (see also Figure 1).

2. Specify the shape of partial effects $g(z_i(t_e), t)$ representing the ELRA

$$g(\mathcal{Z}_i(t), t) = \int_{\mathcal{T}_e(j)} g(z_i(u), t)\mathrm{d}u \approx \sum_{k:t_{e,k} \in \mathcal{T}_e(j)} \Delta_k g(z_i(t_{e,k}), t), \tag{7}$$

with $t \in (\kappa_{j-1}, \kappa_j]$ and $\Delta_k = t_{e,k} - t_{e,k-1}$ the time between two consecutive exposures. Finally, the partial effects are represented as a bivariate smooth function in t_e and t

$$g(z_i(t_e), t) = f(t_e, t) \cdot w_{ij}, \tag{8}$$

where

$$w_{ij} = \begin{cases} z_i(t_e) & \text{if } t_e \in \mathcal{T}_e(j) \\ 0 & \text{else,} \end{cases} \tag{9}$$

and

$$f(t_e, t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \gamma_{mk} B_m(t_e) B_k(t) = \sum_{m,k} \gamma_{mk} B_{mk}(t_e, t) \tag{10}$$

is modeled as a tensor product spline smooth, with marginal bases $B_m(\cdot)$, $B_k(\cdot)$ evaluated at the respective values of t_e and t, $B_{mk}(\cdot, \cdot) = B_m(\cdot)B_k(\cdot)$, and spline coefficients γ_{mk} controlling the shape of $f(t_e, t)$. Note that the distinction in (9) is not really necessary, as

the cases where $w_{ij} = 0$ are implicitly excluded through specification of integration limits in (7), but making it explicit helps with practical data preparation (e.g. specification of the design matrix). The penalized estimation of the smooth terms (cf. section 2.3) implies the assumption of smoothness for $f(t_e, t)$, which ensures that effects of exposures at consecutive exposure times t_e, t'_e are similar and that effects of exposure $z(t_e)$ on the hazards in neighboring intervals j, j' are similar as well. Note that the information regarding the amount of exposure $z_i(t_e)$ is not included in the construction of the marginal bases $B(\cdot)$. This information is added to the design matrix through weights w_{ij} (9), specified beforehand (and therefore known). Lead and lag times are also implemented using these weights by setting the partial effects for exposures outside the relevant window $\mathcal{T}_e(j)$ to zero.

The above specification of the ELRA implies that effects of the TDC can be non-linear over the timing of exposure t_e and time t but not with respect to the value of $z_i(t_e)$, which enters linearly. An extension of the presented framework to non-linear ELRAs via three-dimensional smooth functions of the form $f(t_e, t, z_i(t_e))$ is straight forward (Wood, 2006a, sec. 4.1.8), but was not pursued in this work.

2.3 Estimation and Inference

Stable likelihood-based methods for the parameter estimation of the proposed model have been recently developed in Wood (2011) in the context of penalized models of the form $D(\boldsymbol{\gamma}) + \sum_p \lambda_p \boldsymbol{\gamma}' \boldsymbol{K}_p \boldsymbol{\gamma}$, where $D(\boldsymbol{\gamma})$ is the model deviance, $\boldsymbol{\gamma}$ contains all spline basis coefficients representing model (4), and λ_p and K_p are the smoothing parameters and penalty matrices for the individual smooths $f_p(\cdot)$, respectively. Given $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$, parameter estimates can be obtained by penalized iteratively reweighted least squares (P-IRLS). To guarantee convergence, Wood (2011) employs P-IRLS based on nested iterations, i.e. after each P-IRLS step, estimation of $\boldsymbol{\lambda}$ is updated given the current $\boldsymbol{\gamma}$ estimates. Subsequent papers (Marra and Wood, 2011, 2012; Wood, 2012) develop shrinkage based procedures for simultaneous smoothness and variable selection and methods for confidence intervals and significance tests for smooth components. In the following sections we extend these methods to the context of PAM and, particularly, exposure-lag-response associations.

2.3.1 Confidence Intervals

Confidence intervals (CI) with good coverage properties for smooth terms are developed in Marra and Wood (2012) and are directly applicable to ELRA in (4). Let $\boldsymbol{\gamma}_q$ be the vector of estimated basis coefficients associated with $f(t_e, t)$ in (10), and $\mathbf{V}_{\hat{\boldsymbol{\gamma}}_q}$ the empirical Bayesian covariance matrix of the estimated parameters $\hat{\boldsymbol{\gamma}}_q$. Let further $\mathbf{X}_q$ be the $J \times n_e$ design matrix for a specific exposure history $\mathcal{Z}(t)$, where J is the number of intervals into which the follow-up period has been partitioned, and n_e is the number of columns associated with the tensor-product smooth of the ELRA term. The confidence intervals are given by

$$\mathbf{X}_q\hat{\boldsymbol{\gamma}}_q \pm z_{1-\alpha/2}\sqrt{\text{diag}(\mathbf{X}_q\mathbf{V}_{\hat{\boldsymbol{\gamma}}_q}\mathbf{X}_q^T)} = \hat{\mathbf{f}}_q \pm z_{1-\alpha/2}\widehat{\mathbf{SE}}_q \tag{11}$$

In (11), $\hat{\mathbf{f}}_q$ as well as $\widehat{\mathbf{SE}}_q$ are vectors of length J, representing the estimated cumulative effect of the TDC and its standard errors in intervals $j = 1, \ldots, J$. By defining $\mathbf{X}_q := \mathbf{X}_{q_2} - \mathbf{X}_{q_1}$ in (11) we can obtain estimated *differences* of cumulative effects (and a respective CI) given different exposure histories $\mathcal{Z}_2(t)$ and $\mathcal{Z}_1(t)$. We demonstrate this approach in section 3.3 and investigate properties of such CIs for ELRAs by means of a simulation study in section 4.

2.3.2 Hypothesis testing

The method introduced in section 2.3.1 provides a way to assess differences in cumulative effects at individual time-points of the follow-up. In some applications, however, it is also of interest to assess the overall effect of individual ELRA terms or whether the difference of cumulative effects resulting from differing exposure histories $\mathcal{Z}^1(t)$, $\mathcal{Z}^2(t)$ is different from zero over the whole follow-up (or a predefined partition of it).

For the first question we can use significance tests for individual smooth terms of the form

$$H_0 : \mathbf{f}_q = \mathbf{0}, \tag{12}$$

where $\mathbf{f}_q$ could be any of the smooth components in (4) and particularly the ELRA (7). The general idea of the test is straight forward and uses the representation of the smooth term as a linear transformation of basis coefficients $\boldsymbol{\gamma}_q$ such that $\mathbf{f}_q = \mathbf{X}_q\boldsymbol{\gamma}_q$ and an appropriate

test-statistic has the familiar quadratic Wald-type form

$$T_r = \hat{\mathbf{f}}_q^T \mathbf{V}_{\mathbf{f_q}}^{\mathbf{r}^-} \hat{\mathbf{f}}_{\mathbf{q}}. \tag{13}$$

Here $\mathbf{V}_{\mathbf{f_q}}^{\mathbf{r}^-}$ is the rank-r pseudo inverse of $\mathbf{V}_{\mathbf{f_q}} = \mathbf{X_q} \mathbf{V}_{\boldsymbol{\gamma}_\mathbf{q}} \mathbf{X}_\mathbf{q}^\mathbf{T}$. The difficult part then becomes choosing the appropriate r in the context of penalized estimation, as naive choices (e.g. rank of $V_{\mathbf{f_q}}$) lead to reduced power (see Wood (2012) for details). Given r, which in this context can be a non-integer number, T_r follows a mixture of χ^2 distributions, from which p-values can be obtained routinely (Wood, 2012, p. 4). In section 4.4 we show that this *Overall Test* works well for testing individual ELRA terms.

Although this answers the question if there is any part of $\mathbf{f}_q$ that is not equal to zero with respect to the data set, i.e. for some patient-interval combinations, this may not be of particular interest for the practitioner. More interesting, from a medical or epidemiological point of view, may be the question whether specific, practically relevant exposure histories $\mathcal{Z}^1(t), \mathcal{Z}^2(t)$ lead to significantly different cumulative effects on the hazard rate. For example we would like to know if the differences of cumulative effects resulting from different nutrition profiles (cf. Table 3), are significantly different from 0.

One idea to construct such a test would be to extend the procedure for the overall test presented above. Let $\mathbf{f}_{q_1}$ the vector of cumulative effects associated with exposure history $\mathcal{Z}^1(t)$ and $\mathbf{f}_{q_2}$ the respective vector for exposure history $\mathcal{Z}^2(t)$.

We want to test $H_0 : \mathbf{f}_{q_1} = \mathbf{f}_{q_2}$ vs. $H_1 : \mathbf{f}_{q_1} \neq \mathbf{f}_{q_2}$ or alternatively

$$H_0 : \mathbf{f}_{q_2} - \mathbf{f}_{q_1} = 0 \text{ vs. } H_1 : \mathbf{f}_{q_2} - \mathbf{f}_{q_1} \neq 0 \tag{14}$$

and using the representation via basis functions and coefficients we get

$$\tilde{\mathbf{f}}_q = \mathbf{f}_{q_2} - \mathbf{f}_{q_1} = \mathbf{X}_{q_2}\boldsymbol{\gamma}_q - \mathbf{X}_{q_1}\boldsymbol{\gamma}_q = (\mathbf{X}_{q_2} - \mathbf{X}_{q_1})\boldsymbol{\gamma}_q = \tilde{\mathbf{X}}_q\boldsymbol{\gamma}_q.$$

Note that when $\mathbf{f}_{q_1} = \mathbf{0}$, the testing procedure is analogous to the overall test (12). Furthermore, using $\tilde{\mathbf{f}}_q$ and $\tilde{\mathbf{X}}_q$ we could obtain expressions for the test statistic T_r analogously to (13). However, it is not obvious what the effective degrees of freedom (edf) would be for the difference of two splines, that are needed to obtain the degrees of freedom of the χ^2 distribution. Obviously,

when $\mathbf{f}_{q_1} = \mathbf{0}$ the edf of the difference spline reduce to the edf of $\mathbf{f}_{q_2}$ and in the other extreme when $\mathbf{f}_{q_2} = \mathbf{f}_{q_1}$ exactly, the edf for the difference spline would be 1. The edf for the general case therefore should be somewhere in between, but it's not straight forward how to obtain them. To the authors' knowledge, at the moment there are no procedures available to perform such a generalized test in the context of penalized additive models, thus more methodological research is needed in this area.

3 Association between caloric intake and mortality in ICU patients

3.1 Data and Objective

We apply our method in a retrospective analysis of a large international multi-center study with $n = 9661$ critically ill patients (after preprocessing and application of exclusion criteria) with a maximal follow up of 60 days or until release from hospital. Starting with the day of admission (day 0), goal calories have been determined for each patient by a nutritionist or physician and the actual caloric intake provided by the hospital staff has been recorded for a maximum of 11 *calendar* days after the date of ICU admission, which we denote by $t_e \in \{1, \ldots, 11\}$.

We are interested in the relationship of caloric adequacy and acute mortality, that is, mortality within 30 days after ICU admission. In total, 1974 (20.4%) patients died within this period. As patients released from hospital before $t = 30$ were presumably healthier than patients that remained in hospital, the interpretation of the nutritional effects may be hindered by this informative censoring. For our main analysis, we assume that patients released from the hospital have survived at least until $t = 30$.

For the application of the piece-wise exponential model the time line was partitioned in one-day intervals until $t_{\max} = 30$. We only included patients that survived at least 96 hours, consequently we began evaluation in interval $(4, 5]$. For the purposes of the following analysis we set the status of all patients still alive after $t = 30$ to "censored".

It is important to emphasize, that our analysis considers two different time scales, the time scale of the follow up, denoted by t and the time scale of exposure, denoted by t_e. Survival

times were calculated as differences $t_{death/release} - t_{admission}$. For example, when a patient was admitted to the hospital on Monday, 3pm and died on Sunday 2pm, his survival time was ≈ 5.96 days. Thus, when we discretize the time-line in one day intervals, this patient's event would fall in the interval $(5, 6]$.

Calories on the other hand were recorded at the end of a calendar day. Therefore the first *full* calendar day at the ICU, which we denote by t_e=1, does not necessarily coincide with the first "day" $(0, 1]$ on the scale of the follow-up, but rather spans between the first $(0, 1]$ and the second $(1, 2]$ "day" (or rather 24 hour period of the follow-up). Consequently, when we want to include a lag time of 4 days (96 hours) between exposure at time t_e=1 (2nd calendar day at ICU) and survival, we need to count $((t_e + 1) + 4) \in (5, 6]$ to obtain the correct interval at which at least 4 "days" (96 hours) since exposure have passed on the time scale of the follow up.

3.2 Modeling approach

We adjusted for various potential confounders, including subject specific covariates age, BMI, sex, diagnosis at admission and admission category, the Apache II Score (an overall measure of the patients' health status at admission) as well as patient unrelated covariates like year of admission and a random effect (Gaussian frailty) for the ICUs. Since we model the mortality risk beginning in interval $(4, 5]$ (due to application of exclusion criteria), we also included variables that describe the patients' ICU stay up to that point, namely number of days under mechanical ventilation (MV) and number of days with additional oral intake (OI), number of days with parenteral nutrition (PN), and number of days receiving Propofol (PF) on the first 3 full calendar days of the ICU stay, respectively.

To be able to compare different caloric intakes independently of a patient's weight and caloric requirements, we define a patient's caloric adequacy (CA) as

$$CA = \text{actual daily caloric intake(in } kcal)/\text{goal calories(in } kcal). \tag{15}$$

In many cases, however, we could not quantify the caloric intake exactly, as some patients received additional oral intake on some days, for which the amount of calories had not been recorded. Patients receiving additional oral intake (OI) must have been extubated at some point

and were therefore presumably healthier. At the same time, the *recorded* amount of calories received is, on average, lower when a portion of the calories has been received orally. Ignoring additional OI could therefore result in spurious correlations. Thus we discretized the daily CA, depending on whether or not additional OI occurred (see Table 2). Furthermore, some patients, again presumably healthier, were released from the ICU before the end of the eleven day nutrition protocol phase (incomplete protocols). In these cases we assumed that patients were able to sate themselves after their release from the ICU and assigned nutrition category C_{III} for the remainder of the protocol days. Both assumptions have been investigated in terms of sensitivity analyses (not shown) and did not yield substantial changes to the estimated nutritional effects.

Discretized Caloric Adequacy (dCA)	Description
C_I:	$0\% \leq CA < 30\%$ and no additional oral intake
C_{II}:	$30\% \leq CA < 70\%$ and no additional oral intake, or $0\% \leq CA < 30\%$ and additional oral intake
C_{III}:	$CA \geq 70\%$ and no additional oral intake, or $30\% \leq CA < 70\%$ and additional oral intake, as well as days with incomplete protocols.

Table 2: Discretization of relative caloric intake in categories C_I (lower), C_{II} (mid) and C_{III} (upper)category nutrition.

The effect of nutrition is represented in the model by two terms, $g_{C_{II}}(\cdot,\cdot)$ and $g_{C_{III}}(\cdot,\cdot)$, while category C_I was considered the "reference" category, thus direct interpretation of $g_{C_{II}}$ and $g_{C_{III}}$ is only possible with respect to a (hypothetical) patient that received C_I on all 11 days. Each of the terms has the structure defined in section 2.2.3, except that $z_i(t_e)$ now is represented by dummy variables $z_i^{C_{II}}(t_e)$, $z_i^{C_{III}}(t_e)$, indicating if nutrition at time t_e has been in category C_{II} and C_{III} respectively. Equation (16) shows the final model specification:

$$
\begin{aligned}
\log(\lambda_i(t|\mathbf{X}_i, \mathbf{z}_i, \ell_i)) = f_0(\tilde{t}) &+ \mathbf{X}\boldsymbol{\beta} \\
&+ \beta_{Apache} \cdot x_{i,Apache} + \beta_{Apache:\tilde{t}} \cdot (x_{i,Apache} \cdot \tilde{t}) \\
&+ f_{age}(x_{i,age}) \cdot \tilde{t} + f_{BMI}(x_{i,BMI}) \cdot \tilde{t} \\
&+ \sum_{t_e \in \mathcal{T}(j)} g_{C_{II}}(t_e, \tilde{t}) + \sum_{t_e \in \mathcal{T}(j)} g_{C_{III}}(t_e, \tilde{t}) \\
&+ b_{\ell_i}, \text{ where}
\end{aligned}
\tag{16}
$$

- $f_0(\tilde{t})$ represents the baseline hazard rate, estimated over the interval mid-points $\tilde{t}$,
- $\mathbf{X}\boldsymbol{\beta}$ incorporates all linear time-constant effects of time-constant covariates.
- $\beta_{Apache} \cdot x_{i,Apache} + \beta_{Apache:\tilde{t}} \cdot (x_{i,Apache} \cdot \tilde{t})$ is a linear, linearly time-varying effect of the Apache II Score measured at initial admission to the ICU,
- $f_{age}(x_{i,age}) \cdot \tilde{t}$ and $f_{BMI}(x_{i,BMI}) \cdot \tilde{t}$ are smooth, linearly time-varying effects of age and BMI, modeled as so-called *varying coefficients* (see section 2.2.2 for details),
- $g_{C_{II}}$ and $g_{C_{III}}$ are smoothly time-varying cumulative effects of the nutritional intake (see section 2.2.3 for details) and
- b_{ℓ_i} is an independent identically distributed Gaussian random intercept term attributed to different ICUs in the data set.

In this application all non-linear functions of time-constant covariates $f_p(\mathbf{x}_{,p})$ have been estimated using P-Splines (Eilers and Marx, 1996) with penalties based on second order differences and $M = 10$ (cf. sec. 2.2.2) basis functions spanned over equidistant knots. For the $f(t_e, t)$ terms associated with the ELRA, $M = K = 5$ (cf. eq. (10)) basis functions were used for each dimension and first order differences were used for the dimension of exposure time t_e.

The lag-lead window $\mathcal{T}_e(j)$ in (16) was defined based on substantive considerations with $t_{\text{lag}} = 4$ (see discussion in section 5 on the choice of lag and lead times) and $t_{\text{lead}} = t_{\text{lag}} + 2 \cdot t_e$. We will refer to this specification as *dynamic* lag-lead $\mathcal{T}^{dynamic}$. However, since there is little empirical knowledge about the potential lengths of such lag and especially lead periods, we

also evaluated results for a *static* lag-lead specification $\mathcal{T}^{static}$ with $t_{\text{lag}} = 4$ and $t_{\text{lead}} = 30$, which implies that partial effects, from all relevant intervals, contribute to the cumulative effect until the end of the follow up. Both specifications are depicted in Figure 1. When viewed column-wise, the figures show the intervals at which a specific protocol day ($t_e \in \{1, ..., 11\}$) can *potentially* affect the hazard. When viewed row-wise, one can obtain protocol days t_e which contribute to the cumulative nutrition effect in interval j. In general the t_{lag} and t_{lead} can be used to control the number of exposures that enter the cumulative effect at any given time t, thus potentially controlling the magnitude of this effect.

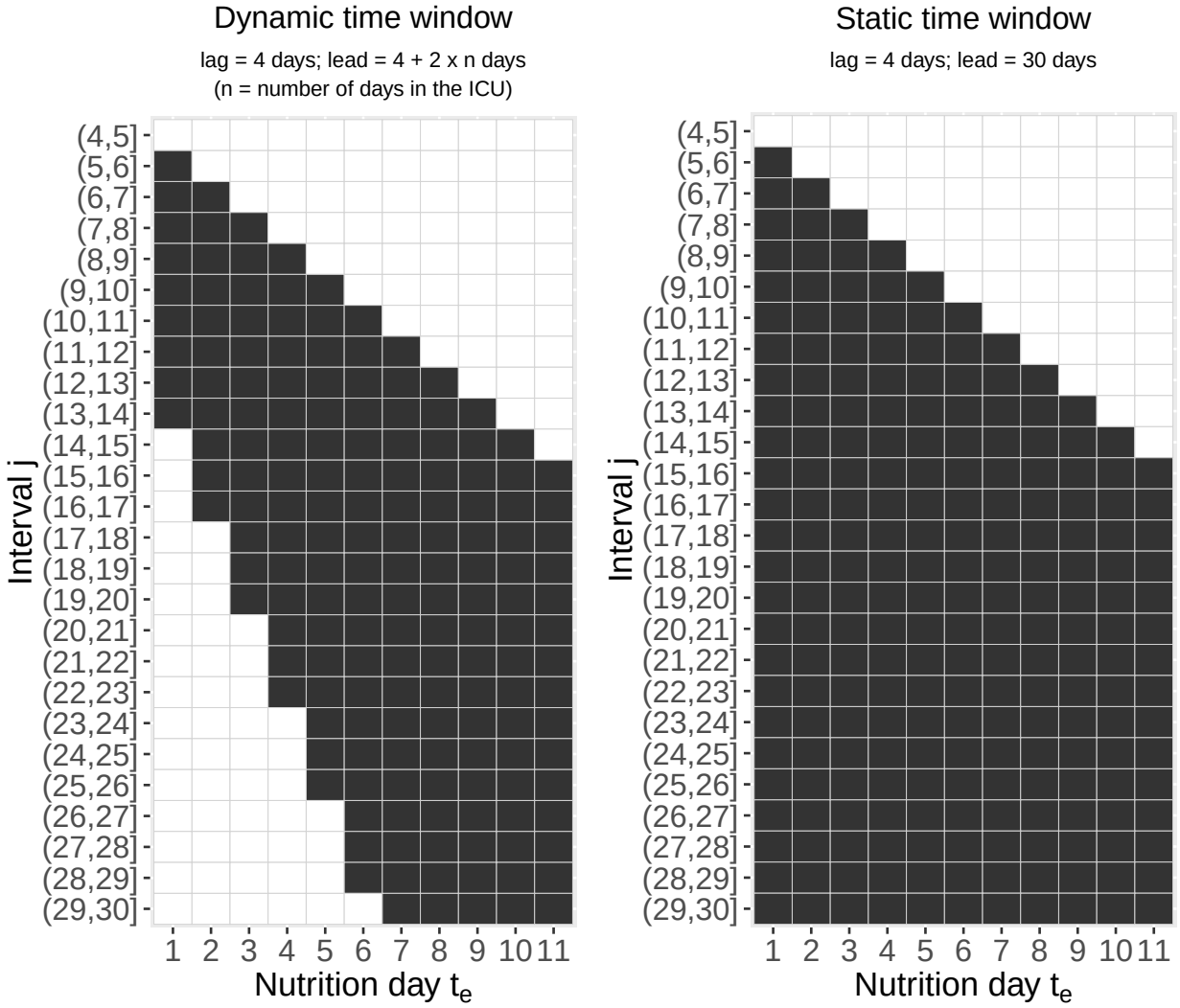

Figure 1: Two possible specifications of the lag-lead $\mathcal{T}_e(j), j = 5, \ldots, 30$. Left panel shows the *dynamic* $\mathcal{T}_e(j)$, right panel depicts the *static* $\mathcal{T}_e(j)$ with a longer and constant lead period.

3.3 Results

We are mostly interested in the relationship between caloric intake and survival, therefore here we only present the results of this association. The estimated effect surfaces $g_{C_{II}}$, $g_{C_{III}}$ (cf. section 3.2) are difficult to visualize and interpret intuitively, therefore we show the cumulative effects of nutrition at interval mid-points $\tilde{t} \in \{4.5, 5.5, \ldots, 29.5\}$ as hazard ratios

$$e_j = \lambda(j|\mathcal{Z}^2(t))/\lambda(j|\mathcal{Z}^1(t)) \tag{17}$$

for patients with different nutrition protocols $\mathcal{Z}^1(t)$ and $\mathcal{Z}^2(t)$ and identical values for all other covariates (see also suggestions by Sylvestre and Abrahamowicz (2009)).

Comparison	$\mathcal{Z}^1(t)$	$\mathcal{Z}^2(t)$
Comparison A	days 1-11: C_I	days 1-4: C_I, days 5-11: C_{II}
Comparison B	days 1-11: C_I	days 1-11: C_{II}
Comparison C	days 1-4: C_I, days 5-11: C_{II}	days 1-11: C_{II}
Comparison D	days 1-11: C_I	days 1-11: C_{III}
Comparison E	days 1-11: C_{II}	days 1-4: C_{II}, days 5-11 C_{III}
Comparison F	days 1-11: C_{II}	days 1-11: C_{III}

Table 3: Overview of evaluated comparisons with nutrition categories C_I (lower), C_{II} (mid) and C_{III} (upper) as defined in Table 2.

The six clinically relevant comparisons considered in our analysis are summarized in Table 3. The estimated cumulative effect differences and suggest that

(a) hypocaloric (category C_I) nutrition is associated with increased hazard rates throughout the follow-up period (Comparisons B, D and to a lesser extent Comparison A);

(b) based on this model, moving from constantly medium (C_{II}) to constantly full (C_{III}) nutrition is not associated with a decrease of the hazard rate (Comparisons E, F);

(c) the (small) hazard rate increases associated with hypocaloric nutrition in the first few days of the protocol phase may persist for up to 25 days after ICU admission (Comparison C).

4 Simulation Study

We performed an extensive simulation study to investigate the performance of the proposed modeling approach. We patterned the simulation study after the application example, especially with respect to data structure and the simulated effects. The main objective was to evaluate

- the ability of our approach to model the type of associations found in the application example in the presence of complex confounding
- its behavior in case of misspecification, specifically the misspecification of the lag-lead structure
- how different penalty structures influence estimation
- the properties of the confidence intervals (sec. 2.3.1) for the effect comparisons of different exposure trajectories e_j (17)
- the properties of the overall test presented in sec. 2.3.2

In the following, we briefly outline the data generation process (section 4.1) and define the settings that were considered in the simulation study (section 4.2), as well as the metrics that were used to evaluate the models performance in each of the settings (section 4.3). Finally, results are presented in section 4.4.

4.1 Data generation

To evaluate the proposed approach we simulated data from the model

$$\lambda(t) = \beta_0 + f_0(t) + f(x,t) + g_{C_{II}}(\mathcal{Z}_i(t),t) + g_{C_{III}}(\mathcal{Z}_i(t),t) \tag{18}$$

and represent this model by $\mathbf{X}^C$, the design matrix for model (18) containing complete covariate information for all subjects $i = 1, ..., n$ and intervals $j = 1, \ldots, J$, and coefficient vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_f, \boldsymbol{\gamma}_{C_{II}}, \boldsymbol{\gamma}_{C_{III}})'$ such that

$$\boldsymbol{\lambda} = \exp(\mathbf{X}^C\boldsymbol{\gamma}) = (\lambda_{1,1}, \ldots, \lambda_{1,J}, \lambda_{2,1}, \ldots, \lambda_{2,J}, \ldots \lambda_{n,1}\lambda_{n,J})' \tag{19}$$

is a length nJ vector of subject- and interval-specific hazard rates $\lambda_{i,j}$. Given the subject-specific hazard rate vector $\boldsymbol{\lambda}_{i,\cdot}$ of length J, we draw random survival times t_i from the piece-wise exponential distribution (see appendix A.2). New data sets are then constructed based on the covariates of subjects $i = 1, ..., n$ and by generating the event and offset variables according to t_i (cf. section 2.1). The procedure is summarized below:

1. Set parameters $\boldsymbol{\gamma}$, specifying the shape of $f_0(t)$, $f(x, t)$ and $g_{C_{II}}(\mathcal{Z}_i(t), t)$ and $g_{C_{III}}(\mathcal{Z}_i(t), t)$ in (18).

2. Obtain hazard rate vectors $\boldsymbol{\lambda}_{i,\cdot} = (\lambda_{i,1}, \ldots, \lambda_{i,J}) \subset \boldsymbol{\lambda} = \mathbf{X}^C\boldsymbol{\gamma}$ for each subject i, $i = 1, \ldots, n$.

3. For replications $r = 1, \ldots, R$

   (a) Draw new random survival times $t_{i,r}$ from the piece-wise exponential distribution with rates $\boldsymbol{\lambda}_i$ (as described in Appendix A.2).

   (b) Given survival times $t_{i,r}$, obtain $\mathbf{X}^r$ by subsetting the complete data $\mathbf{X}^C$ such that only observations for intervals j during which subject i is under risk remains in the data and adjust the event and offset covariates accordingly when $t_{i,r} < t_{max}$.

   (c) Return $\mathbf{X}^r$ and $(t_{i,r})_{i=1,\ldots,n}$.

4.2 Settings

Let $\boldsymbol{\gamma}_g = (\boldsymbol{\gamma}_{C_{II}}, \boldsymbol{\gamma}_{C_{III}})'$ be the coefficient vector controlling the functional shape of the ELRAs $g_{C_{II}}$ and $g_{C_{III}}$ in (18), $\mathcal{T} \in \{\mathcal{T}^{dynamic}, \mathcal{T}^{static}\}$ the lag-lead window that defines how partial effects are cumulated (cf. Figure 1) and $P \in \{P_1, P_2\}$ the P-Spline penalty type imposed on the estimates of $\boldsymbol{\gamma}_g$ ($P_1 \widehat{=}$ 1st order differences, $P_2 \widehat{=}$ 2nd order differences).

Generation of data sets $\mathbf{X}^r$ depends on $\boldsymbol{\gamma}_g$ and $\mathcal{T}$, whereas estimation of $\hat{\boldsymbol{\gamma}}_g$ from $\mathbf{X}^r$ depends on $\mathcal{T}$ and P. Accordingly, the different simulation scenarios considered in this work can be summarized by the parameters used for data generation and by the specification of the model terms associated with the ELRAs (cf. Table 4).

Settings I-III differ with respect to the functional shape of the ELRAs implied by the coefficients $\boldsymbol{\gamma}_g^1$, $\boldsymbol{\gamma}_g^2$ and $\boldsymbol{\gamma}_g^3$ (cf. supplementary appendix Figure 6). The parameters were set such

Data Generation	Estimation	Setting	
$\boldsymbol{\gamma}_g^1, \mathcal{T}^{dynamic}$	$\mathcal{T}^{dynamic}, P_1$	I	I.a
	$\mathcal{T}^{dynamic}, P_2$		I.b
	$\mathcal{T}^{static}, P_1$		I.c
	$\mathcal{T}^{static}, P_2$		I.d
$\boldsymbol{\gamma}_g^2, \mathcal{T}^{dynamic}$	$\mathcal{T}^{dynamic}, P_1$	II	II.a
	$\mathcal{T}^{dynamic}, P_2$		II.b
	$\mathcal{T}^{static}, P_1$		II.c
	$\mathcal{T}^{static}, P_2$		II.d
$\boldsymbol{\gamma}_g^3, \mathcal{T}^{static}$	$\mathcal{T}^{dynamic}, P_1$	III	III.a
	$\mathcal{T}^{dynamic}, P_2$		III.b
	$\mathcal{T}^{static}, P_1$		III.c
	$\mathcal{T}^{static}, P_2$		III.d
—, —	$\mathcal{T}^{dynamic}, P_1$	IV	IV.a
	$\mathcal{T}^{dynamic}, P_2$		IV.b
	$\mathcal{T}^{static}, P_1$		IV.c
	$\mathcal{T}^{static}, P_2$		IV.d

Table 4: Settings considered in the simulation study. Data generation setting IV refers to a special case where coefficients associated with the ELRA where set to zero ($\boldsymbol{\gamma}_g = \mathbf{0}$).

that $g_{C_{II}}$ is similar to the effects estimated in section 3.3 and we set $g_{C_{III}} = 1.2 g_{C_{II}}$ for all settings. In Settings *I* and *II* a dynamic lag-lead was used to specify the ELRAs, in Setting *III* the static lag-lead. Setting *IV* is an important special case, where we assume that nutrition has no effect on the hazard in the data generating model, i.e., $\boldsymbol{\gamma}_g = \mathbf{0}$ and consequently $e_j = 1, \forall j$ (cf. equation (17)). Parameters of the baseline $f_0(t)$ and the confounder effect $f(x,t)$ were held constant across all settings (cf. Figure 5).

For each data generating setting we compare estimates based on all 4 combinations of $\mathcal{T}$ (dynamic/static) and P (P_1/P_2), yielding 16 simulation settings in total with $R = 500$ replications each. Depending on whether $\mathcal{T}$ equals the lag-lead window of the data generation, the model is specified correctly or misspecified. Accordingly, another way to categorize the different settings is:

- Correctly specified $\mathcal{T}$: Settings I.a, I.b, II.a, II.b, III.c, III.d
- Misspecified $\mathcal{T}$
  - Lead too long: Settings I.c, I.d, II.c, II.d
  - Lead too short: Settings III.a, III.b
- ELRA = 0: Settings IV.a - IV.d

4.3 Evaluation

In the data generation process, we included confounders in addition to variables related to the ELRA to obtain a more complex and realistic simulation. However, at the evaluation stage of the simulations we focus on the ELRA.

Let e_j denote the true cumulative effect of the nutrition in interval $j = 1, \ldots, J$ as defined in equation (17) and $\hat{e}_{j,r}$ the respective estimation obtained from simulation run $r = 1, \ldots, R = 500$. For the quantitative evaluation of the models performance we defined the *root mean squared error*

$$\text{RMSE} = \frac{1}{R} \sum_{r=1}^{R} \left(\frac{1}{J} \sum_{j=1}^{J} (e_j - \hat{e}_{j,r})^2 \right)^{1/2} \tag{20}$$

and *coverage*

$$\text{coverage}_\alpha = \frac{1}{R}\sum_{r=1}^{R}\sum_{j=1}^{J} I\left(e_j \in [\hat{e}_{j,r} \pm z_{1-\alpha/2}\widehat{SE}_{j,r}]\right). \tag{21}$$

In the following we set $\alpha = 0.05$. Additionally bias is investigated graphically by comparison of trajectories of individual simulation runs, the average trajectory over all simulation runs and the true effects (e.g. Figure 3).

4.4 Results

Summarized results of the simulation study are presented in Figure 2. Full raw results are presented in section A.3 of the supplementary appendix. The top panel of Figure 2 depicts mean coverage (21) of models with different penalty (P_1/P_2) and lag-lead (static/dynamic) specifications across all comparisons (cf. Table 3) and Settings I – IV. Setting IV can bee viewed as a baseline case, where all model specifications lead to close to nominal (95%) levels of coverage (although models with dynamic lag-lead and P_1 penalties have coverages slightly below 95% across all comparisons). In general, models based on P_2 penalties (gray lines) have better coverage properties compared to those based on P_1 (black lines) and the largest deviations from the nominal level occur when models with P_1 penalty are misspecified with respect to $\mathcal{T}$ (dashed black line Setting I, solid black line Setting III). The bottom panel presents model evaluation with respect to RMSE (20). Models based on P_1 penalties have lower RMSEs across most settings and comparisons. RMSEs are, on average, highest when models are misspecified (dashed lines Settings I and II, solid lines Setting III), and in Setting II, where the true effects e_j have more variable shapes compared to other settings.

Figure 3 shows exemplary raw results for *Setting I.a* and *Setting I.b*, where data was simulated with a dynamic lag-lead and where models were correctly specified with respect to $\mathcal{T}$, but different penalties, P_1 (top panel) and P_2 (bottom panel), were used at the estimation stage. Qualitatively the true relationship was on average (dashed gray lines) captured in both settings and for all comparisons, however, especially for models based on P_1 penalties, true effects are on average slightly underestimated, especially at the beginning of the follow up, which in turn

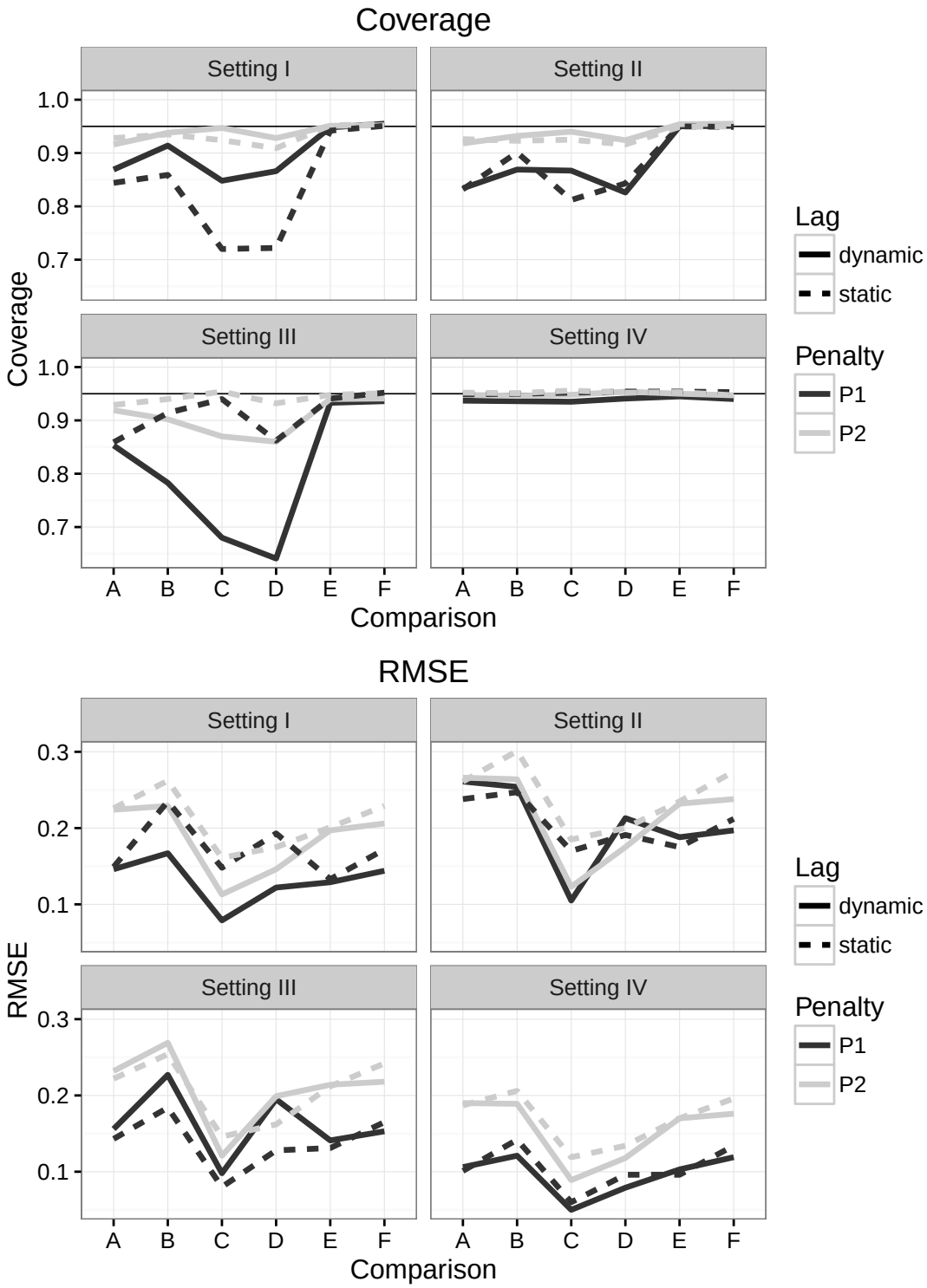

Figure 2: Simulation results summarized by settings for data generation (facets), penalty (black (P1) vs. gray (P2) lines), and lag-lead specification (dynamic (solid) vs. static (dashed) lines). RMSEs and Coverages were averaged for each comparison and subsetting.

explains the lack of coverage for models with this specification. This problem is more severe for comparisons A – D, where the true effect is more pronounced, compared to comparisons E and F. Similar findings can be found for the other Settings (cf. section A.3, see also Table 5). In general, models based on P_2 penalties appear, on average, to capture wiggly effects more closely and show better coverage properties, but tend to exhibit higher RMSEs due to a higher variance of the individual trajectories (Figure 3, supplementary Figure 9). Especially in case of a misspecified lead time (supplementary Figures 8, 10, 11), these models have much better coverage properties compared to models based on P_1 penalties.

In Figure 4 we present the results of the simulation study with respect to the overall test (12) presented in section 2.3.2. In our simulations we included two ELRA terms, therefore we perform tests for both terms, namely $H_0 : g_{C_{II}} = g_{C_{III}} = 0$. In Figure 4 column facets indicate the setting for the data generation process (Setting I-IV) and rows facets indicate the estimation settings (a-d). For each combination $R = 500$ simulation runs were performed and the overall test calculated for both effects, $g_{C_{II}}$ (black crosses) and $g_{C_{III}}$ (gray circles). The empirical rejection rate is denoted by α_e respectively. One important observation is that the distribution of p-values is uniform for data generation setting IV, when in fact $g_{C_{II}} = g_{C_{III}} = 0$, and the empirical rejection rates are close to the nominal level of $\alpha = 0.05$. The power of the test (α_e for settings I-III) is also satisfactory, especially for $g_{C_{III}}$ that was simulated to have a larger effect than $g_{C_{II}}$, although it is generally lower when data was generated from Setting III and when models where fit using penalty specification P_2.

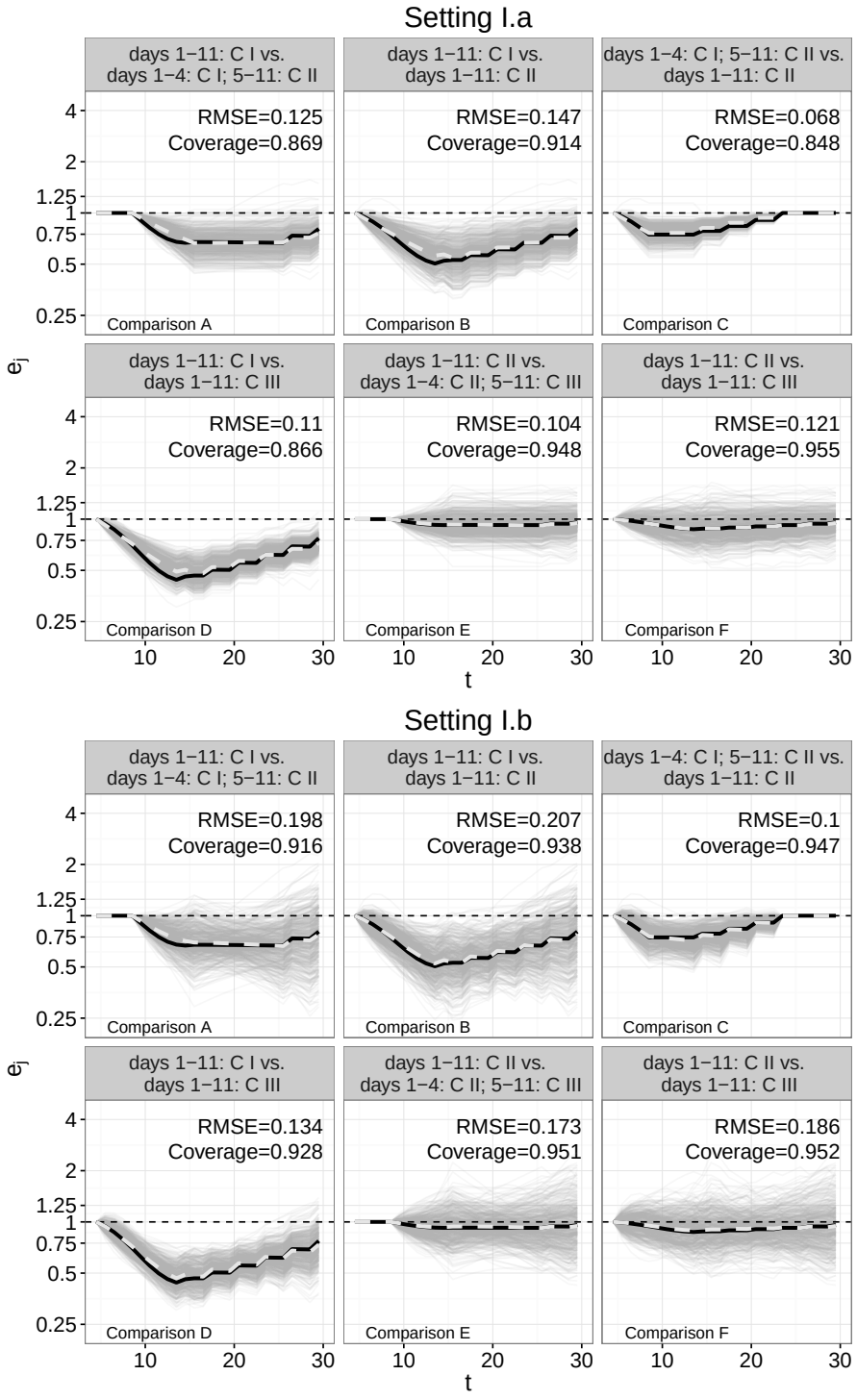

Figure 3: Settings *I.a* and *I.b*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed red line indicates the piece-wise average over all $R = 500$ trajectories.

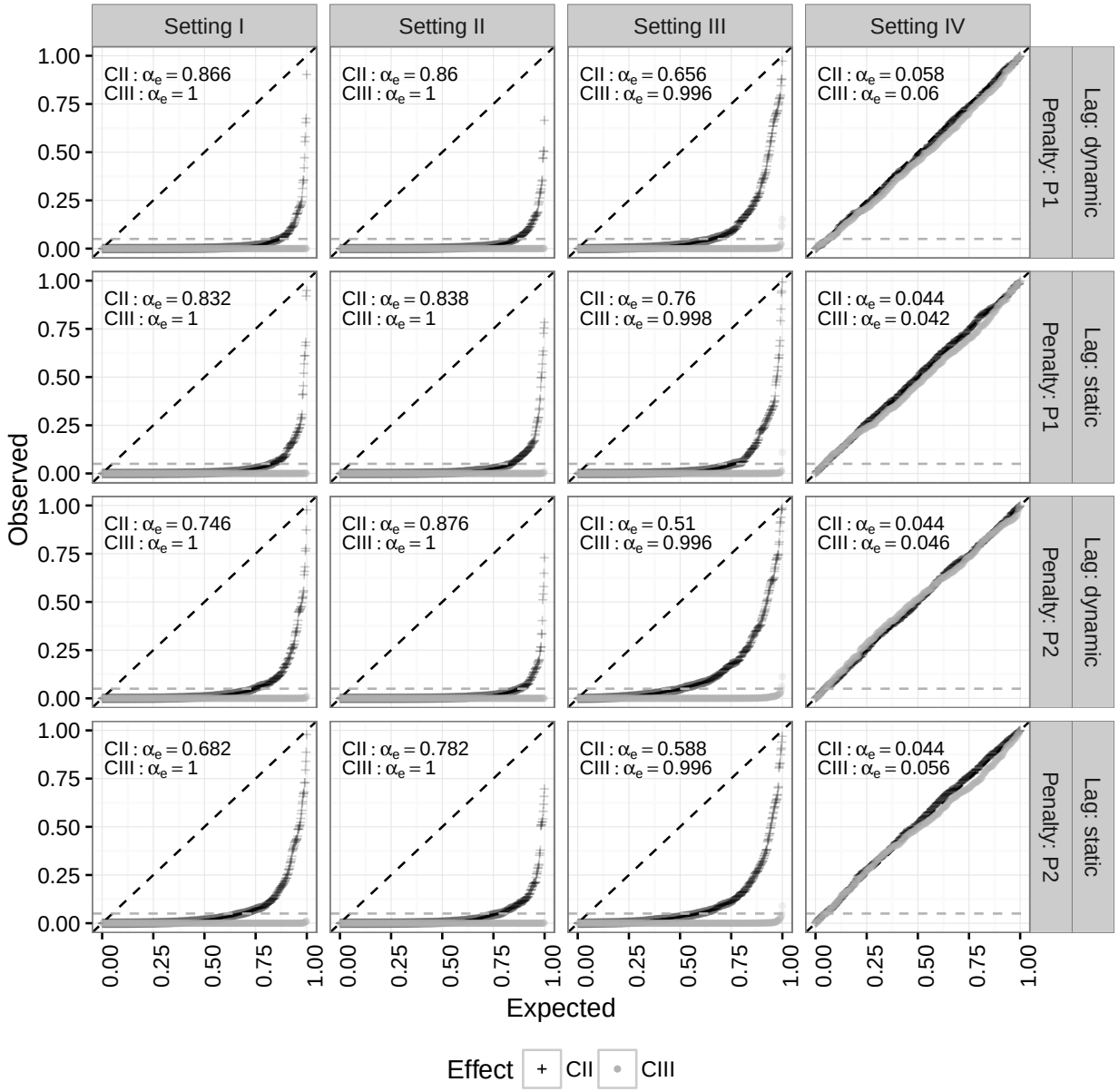

Figure 4: Quantile-Quantile plots of the p-values, empirical (y-axis) and theoretical ($U(0,1)$, x-axis) quantiles, for each combination of data generation (columns) and estimation (rows) settings. Empirical rejection rates to the nominal level $\alpha = 0.05$ are denoted by α_e. Black crosses depict p-values for the hypothesis $H_0 : g_{C_{II}} = 0$, gray circles indicate the p-values for the hypothesis $H_0 : g_{C_{III}} = 0$.

5 Discussion

By embedding the concept of PEMs into the framework of additive models, we were able to establish a very versatile model class for life-time data analysis that inherits the robust and flexible tools for modeling, estimation and validation of penalized generalized additive mixed models, as has been discussed in section 2. In contrast to traditional PEMs, the baseline and time-varying effects are represented as flexible, potentially non-linear penalized splines. We further presented a novel approach to model exposure-lag-response associations, or cumulative effects of time-dependent covariates (exposures), that takes into account timing and amount of the exposure as well as the time since exposure. The practical value and relevance of this approach was demonstrated by application to an important medical research question (cf. section 3) and is further enhanced by the readily available open source implementation of the proposed estimation routine. In addition to the discussed advantages recent algorithmic advances (Wood et al., 2016) for the underlying implementation also allow application to "giga-data" scenarios with $> 10^8$ number of patient-intervals under consideration. The proposed presentation of the results in form of hazard ratio trajectories for different pairs of exposure histories provides a more intuitive alternative compared to classical visualization techniques (e.g. contour plots). Simulation studies (section 4) confirmed that our method is suitable to estimate complex ELRAs and is relatively robust to misspecification. Moreover, when no true exposure effect was present, both the coverage of the proposed CIs for all comparisons and the Type I error rate of the hypothesis tests were maintained near nominal levels in all simulated scenarios, regardless of the specification of penalty and the lag and lead times (cf. Setting IV in Figures 2 (upper panel) and 4).

However, the simulation studies also revealed that CIs can sometimes have sub-nominal coverage (especially in the case of P1 penalties). Bootstrapped confidence intervals may provide improved coverage, but, depending on the number of covariates and the complexity of their effects, the computational cost of such approaches may be very high. It is also apparent that misspecification of lag and lead times can induce bias, potentially underestimating effects at the end of the follow up if the lead time was specified as too short and overestimating them if the lead time was specified as too long. Thus a data-driven selection of the relevant

time window would be preferable, which offers possibilities for future research. One approach for such a procedure could include an additional penalty, that, for example, would penalize partial effects depending on the time since exposure, similar to the double penalty approach by Obermeier et al. (2015). Another interesting extension would be the application of these methods to competing risks. This again would present challenges regarding the interpretation of nutritional effects, as the ELRA could have differing functional shapes for different outcomes.

In general, the interpretation of effects of time-dependent covariates is quite challenging as the "externality" of these variables is sometimes unclear, in that although nutrition is administered by the hospital staff, the amount of nutrition provided could still depend on the patients' health status – e.g., patients undergoing procedures due to life-threatening complications presumably receive less calories or feeding could be stopped due to the decision to withdraw life support. Handling such variables always demands a trade-off with respect to the recency of the covariate (Crowder, 2012, ch. 3.6.), that may result in better adjustment for confounding for more recent values of the covariate, but may also be fully indicative of the outcome and thus induce indication bias (Signorello et al., 2002; Sjoding et al., 2015). In our application we tried to address this issue by including a minimum lag time of four days for the nutritional effects.

6 Software details

All analyses and simulations presented have been implemented in the **R** statistical programming environment (R Core Team, 2016). To facilitate the parallel processing of the individual simulation runs we used the add-on packages **BatchJobs** and **BatchExperiments** (Bischl et al., 2015). Models were fit using package **mgcv** (Wood, 2011) and graphical visualizations have been implemented with **ggplot2** (Wickham, 2016).

Acknowledgements

The authors would like to thank Daren Heyland for providing the data set and useful discussion. Fabian Scheipl was supported by the German Research Foundation through the Emmy Noether Programme, grant GR 3793/1-1 to Sonja Greven.

References

Berger, M. M. and C. Pichard (2012). Best timing for energy provision during critical illness. *Critical Care 16*(2), 215.

Berhane, K., M. Hauptmann, and B. Langholz (2008). Using tensor product splines in modeling exposure-time-response relationships: Application to the colorado plateau uranium miners cohort. *Statistics in Medicine 27*(26), 5484–5496.

Bischl, B., M. Lang, O. Mersmann, J. Rahnenführer, and C. Weihs (2015). BatchJobs and BatchExperiments: Abstraction mechanisms for using R in batch environments. *Journal of Statistical Software 64*(11), 1–25.

Crowder, M. J. (2012). *Multivariate Survival Analysis and Competing Risks*. Chapman & Hall / CRC Texts in Statistical Science. Hoboken: CRC Press.

Demarqui, F. N., R. H. Loschi, and E. A. Colosimo (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis 14*(3), 333–356.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science 11*(2), 89–121.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics 10*(1), 101–113.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine 33*(5), 881–899.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological) 55*(4), 757–796.

Heyland, D. K., N. Cahill, and A. G. Day (2011). Optimal amount of calories for critically ill patients: Depends on how you slice the cake! *Critical Care Medicine* (39), 2619–2626.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software 38*(8), 1–29.

Marra, G. and S. N. Wood (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis 55*(7), 2372–2387.

Marra, G. and S. N. Wood (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics 39*(1), 53–74.

McClave, S. A., R. G. Martindale, V. W. Vanek, M. McCarthy, P. Roberts, B. Taylor, J. B. Ochoa, L. Napolitano, and G. Cresci (2009). Guidelines for the provision and assessment of nutrition support therapy in the adult critically ill patient: Society of critical care medicine (SCCM) and american society for parenteral and enteral nutrition (A.S.P.E.N.). *Journal of Parenteral and Enteral Nutrition 33*(3), 277–316.

Obermeier, V., F. Scheipl, C. Heumann, J. Wassermann, and H. Küchenhoff (2015). Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 64*(2), 395–412.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Signorello, L. B., J. K. McLaughlin, L. Lipworth, S. Friis, H. T. Sørensen, and W. J. Blot (2002). Confounding by indication in epidemiologic studies of commonly used analgesics. *American Journal of Therapeutics 9*(3), 199–205.

Singer, P., M. M. Berger, G. Van den Berghe, G. Biolo, P. Calder, A. Forbes, R. Griffiths, G. Kreyman, X. Leverve, and C. Pichard (2009). Espen guidelines on parenteral nutrition: Intensive care. *Clinical Nutrition 28*(4), 387–400.

Sjoding, M. W., K. Luo, M. A. Miller, and T. J. Iwashyna (2015). When do confounding-by-indication and inadequate risk adjustment bias critical care studies? a simulation study. *Critical Care 19*(1), 195.

Sylvestre, M.-P. and M. Abrahamowicz (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine 28*(27), 3437–3453.

Whitehead, J. (1980). Fitting Cox's Regression Model to Survival Data using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 29*(3), 268–275.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016 ed.). New York, NY: Springer.

Wood, S. N. (2006a). *Generalized additive models: An introduction with R*. CRC Press.

Wood, S. N. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics 62*(4), 1025–1036.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(1), 3–36.

Wood, S. N. (2012). On p-values for smooth components of an extended generalized additive model. *Biometrika 100*, 221–228.

Wood, S. N., Z. Li, G. Shaddick, and N. H. Augustin (2016, June). Generalized additive models for gigadata: modelling the UK black smoke network daily data. *Journal of the American Statistical Association*, 1–40.

Wood, S. N., F. Scheipl, and J. J. Faraway (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing 23*(3), 341–360.

Xiao, Y., M. Abrahamowicz, E. E. M. Moodie, R. Weber, and J. Young (2014). Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: Reassessing the cardiovascular risks of didanosine treatment in the Swiss HIV cohort study. *Journal of the American Statistical Association 109*(506), 455–464.

A Simulation

A.1 Data generation

In this section we describe the process of simulating survival times from model (18) in more detail. First we specify the functional shapes of $f_0(t)$, $f(x,t)$ and $g_{C_{II}}(\mathcal{Z}_i(t),t)$, $g_{C_{III}}(\mathcal{Z}_i(t),t)$ for the different simulation settings (cf. Table 4). Common to all settings were the specification of $\beta_0 = -4.5$ as well as $f_0(t)$ and $f(x,t)$ specified by coefficient vectors $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_f$ (see Figure 5).

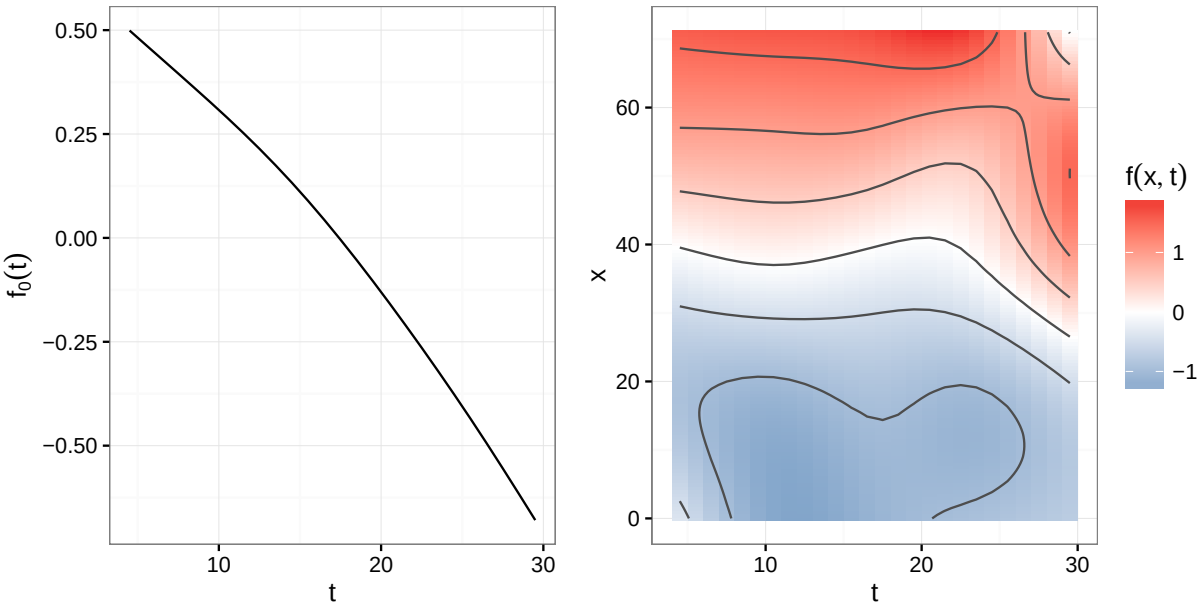

Figure 5: True (log) baseline $f_0(t)$ and confounder effect $f(x,t)$ used for data generation in the simulation studies.

The cumulative effects $g_{C_{II}}$ and $g_{C_{III}}$ are represented as hazard ratios e_j (cf. equation (17)) for the 6 comparisons defined in Table 3 and displayed in Figure 6 for settings I, II and III. The respective cumulative effects are controlled by parameters $\boldsymbol{\gamma}_g = (\gamma^k_{C_{II}}, \gamma^k_{C_{III}})'$, $k = 1,\dots,3$. Given these parameters, we define the coefficients vector $\boldsymbol{\gamma} = (\beta_0, \boldsymbol{\gamma}_0', \boldsymbol{\gamma}_f', \boldsymbol{\gamma}_g')'$ and hazards $\boldsymbol{\lambda} = \mathbf{X}^C\boldsymbol{\gamma}$. The design matrix $\mathbf{X}^C$ contains the covariate information regarding caloric adequacy, taken from the real data example discussed in section 3.1 of the main manuscript. For patients without a complete nutrition protocol (due to an event or discharge from the ICU before day 11 of the nutrition protocol), the last observation was carried forward until day 11.

The distribution of caloric adequacy categories C_I, C_{II} and C_{III} across the 11 days of nutrition protocol were taken from the real data example discussed in section 3.1 (cf. Figure 7), such that number of patients with caloric adequacy C_I is high at the beginning (~ 5000) and decreases towards the last day of nutrition protocol $t_e = 11$ (~ 1500), while C_{III} is lowest at the beginning (~ 2000) and predominant towards the end (~ 7000). Number of patients with category C_{II} nutrition is relatively small in the beginning (~ 2500) and also decreases towards the end of the nutrition protocol phase (~ 1000).

As $n = 9661$ subjects are included in the study and we consider $J = 26$ intervals, $\mathbf{X}^C$ contains $9661 \cdot 26$ rows and $\boldsymbol{\lambda}$ is a vector of the same length containing hazards for each subject and for each interval. Given these hazards, we can simulate new survival times for each subject i (see section A.2) and create simulated data sets $\mathbf{X}$ by adjusting number of observations for each subject i, their event variables and offsets according to the simulated survival times and subset $\mathbf{X}^C$, such that, for all subjects, it only includes intervals for which the respective subjects are alive at the beginning of the interval.

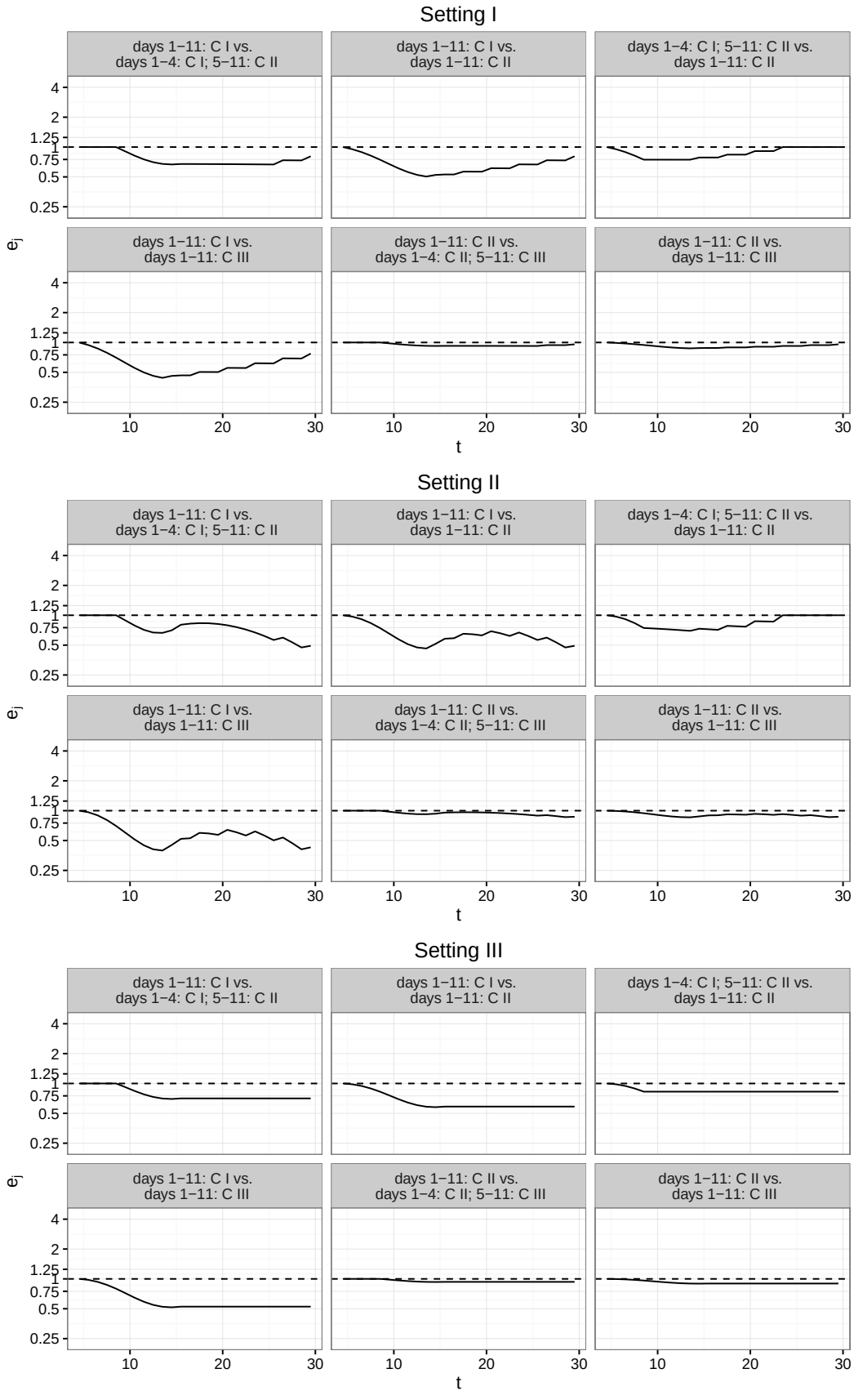

Figure 6: True effects e_j for simulation settings I, II and III (top to bottom, cf. Table 4).

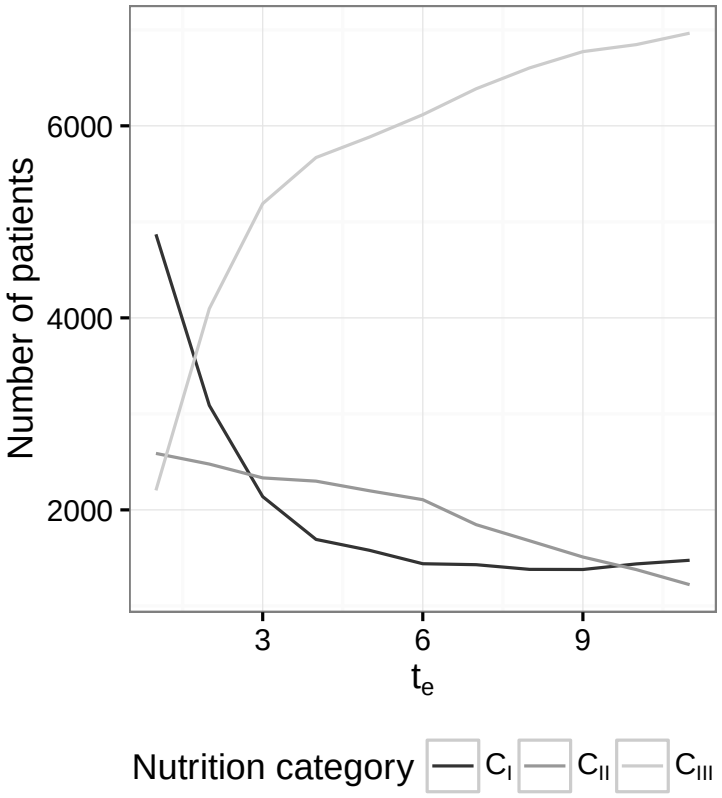

Figure 7: Number of patients with caloric adequacy C_I, C_{II} and C_{III}, respectively over the course of the 11 day nutrition protocol phase.

A.2 Drawing random survival times from piece-wise exponential distribution

Drawing survival times from piece-wise exponential distributions was performed with the `rpexp` function from the **R** package **msm** (Jackson, 2011). The procedure is described in the following:

Let κ_{j-1} the left border of the interval $(\kappa_{j-1}, \kappa_j]$, $j = 1, \ldots, J$. For each subject $i = 1, \ldots, n$,

I. Set $j = 1$.

II. While $j < J$

(i) Draw survival time t'_{ij} from the exponential distribution with rate λ_{ij}, set $t_i = t'_{ij} + \kappa_{j-1}$

(ii) if $\kappa_{j-1} < t_i \leq \kappa_j$, accept t_i

(iii) else $j = j + 1$

III. Draw t'_{iJ} with hazard λ_{iJ}, accept $t_i = t'_{iJ} + \kappa_{J-1}$.

A.3 Results

In the following raw results for the individual settings of the simulation study (cf. section 4 of the main article) are presented:

- Table 5 presents summary statistics for RMSE and Coverage$_\alpha$ of all simulation runs across all settings (cf. Table 4) and comparisons (cf. Table 3).
- Figure 8 contains individual trajectories for Settings I.c and I.d (Misspecified $\mathcal{T}^{dynamic}$, lead time to long)
- Figure 9 contains individual trajectories for Settings II.a and II.b (Correctly specified $\mathcal{T}^{dynamic}$)
- Figure 10 contains individual trajectories for Settings II.c and II.c (Misspecified $\mathcal{T}^{dynamic}$, lead time to long)
- Figure 11 contains individual trajectories for Settings III.a and III.b (Misspecified $\mathcal{T}^{static}$, lead to short)
- Figure 12 contains individual trajectories for Settings III.c and III.d (Correctly specified $\mathcal{T}^{static}$)
- Figure 13 contains individual trajectories for Settings IV.a and IV.b (Null case, no cumulative effect)
- Figure 14 contains individual trajectories for Settings IV.c and IV.d (Null case, no cumulative effect)

On average, models fit with P_1 penalties tend to underestimate the true effects at the beginning of the follow up, and overestimate, when the lead time was specified too short (Setting III.a). Models based on P_2 penalties tend, on average, to fit the true effect shape more closely, while having larger mean RMSE and better coverage properties.

			Comparison					
Setting	Model		A	B	C	D	E	F
Setting I	a	RMSE	0.125	0.147	0.068	0.110	0.104	0.121
		Coverage	0.869	0.914	0.848	0.866	0.948	0.955
	b	RMSE	0.198	0.207	0.100	0.134	0.173	0.186
		Coverage	0.916	0.938	0.947	0.928	0.951	0.952
	c	RMSE	0.133	0.219	0.141	0.184	0.106	0.144
		Coverage	0.844	0.859	0.720	0.722	0.942	0.951
	d	RMSE	0.200	0.235	0.143	0.160	0.176	0.204
		Coverage	0.928	0.935	0.924	0.909	0.952	0.954
Setting II	a	RMSE	0.248	0.241	0.094	0.200	0.164	0.178
		Coverage	0.834	0.869	0.867	0.826	0.950	0.951
	b	RMSE	0.243	0.243	0.110	0.161	0.209	0.220
		Coverage	0.918	0.932	0.940	0.924	0.954	0.955
	c	RMSE	0.225	0.226	0.161	0.180	0.151	0.185
		Coverage	0.832	0.901	0.812	0.843	0.950	0.949
	d	RMSE	0.236	0.275	0.169	0.185	0.210	0.247
		Coverage	0.926	0.923	0.925	0.916	0.949	0.950
Setting III	a	RMSE	0.133	0.202	0.129	0.181	0.116	0.131
		Coverage	0.853	0.783	0.490	0.641	0.933	0.936
	b	RMSE	0.207	0.242	0.141	0.181	0.191	0.200
		Coverage	0.919	0.902	0.626	0.860	0.940	0.941
	c	RMSE	0.121	0.151	0.065	0.112	0.107	0.138
		Coverage	0.859	0.914	0.940	0.862	0.941	0.952
	d	RMSE	0.195	0.223	0.129	0.145	0.188	0.220
		Coverage	0.929	0.940	0.954	0.932	0.948	0.951
Setting IV	a	RMSE	0.084	0.098	0.040	0.064	0.084	0.100
		Coverage	0.937	0.936	0.935	0.941	0.945	0.940
	b	RMSE	0.167	0.170	0.078	0.105	0.152	0.161
		Coverage	0.948	0.946	0.948	0.954	0.950	0.947
	c	RMSE	0.081	0.115	0.049	0.077	0.079	0.113
		Coverage	0.949	0.950	0.952	0.954	0.954	0.953
	d	RMSE	0.164	0.183	0.105	0.118	0.151	0.178
		Coverage	0.952	0.951	0.956	0.952	0.953	0.947

Table 5: Raw simulation results for all considered settings and across Comparisons A – D. Settings and comparisons are explained in the main part of the article. The nominal level of the confidence intervals was set to 95%.

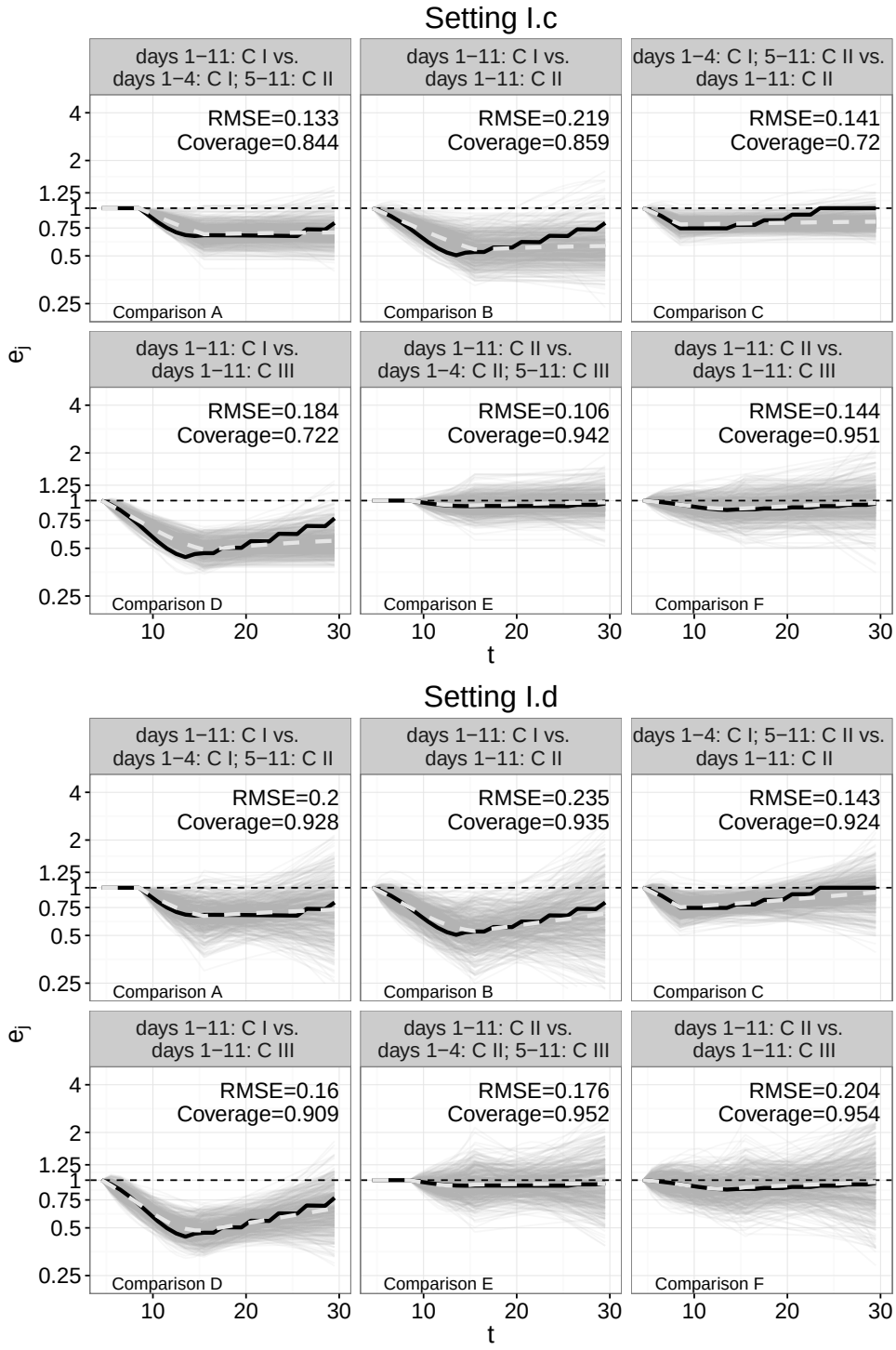

Figure 8: Settings *I.c* and *I.d*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.

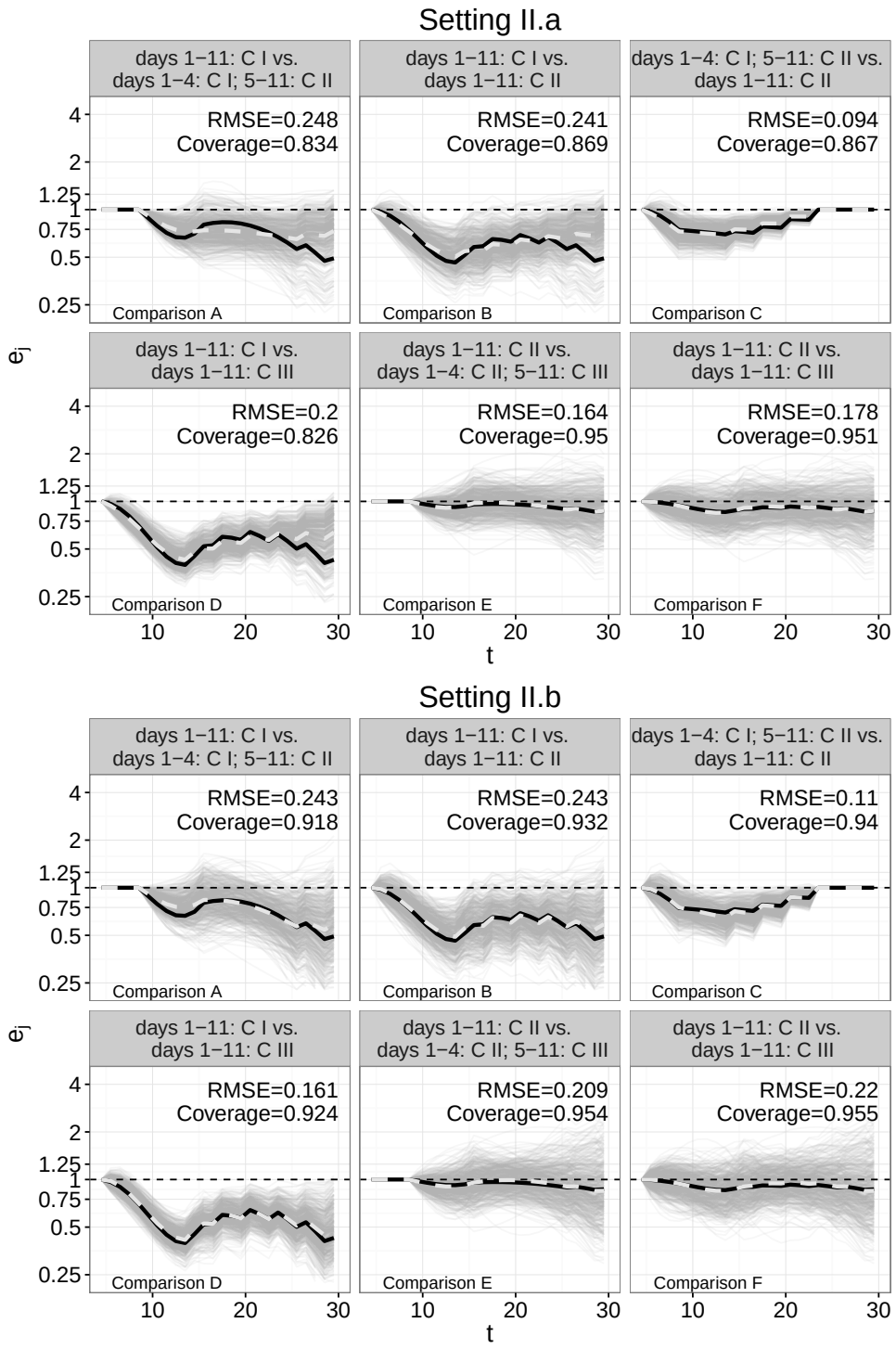

Figure 9: Settings *II.a* and *II.b*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.

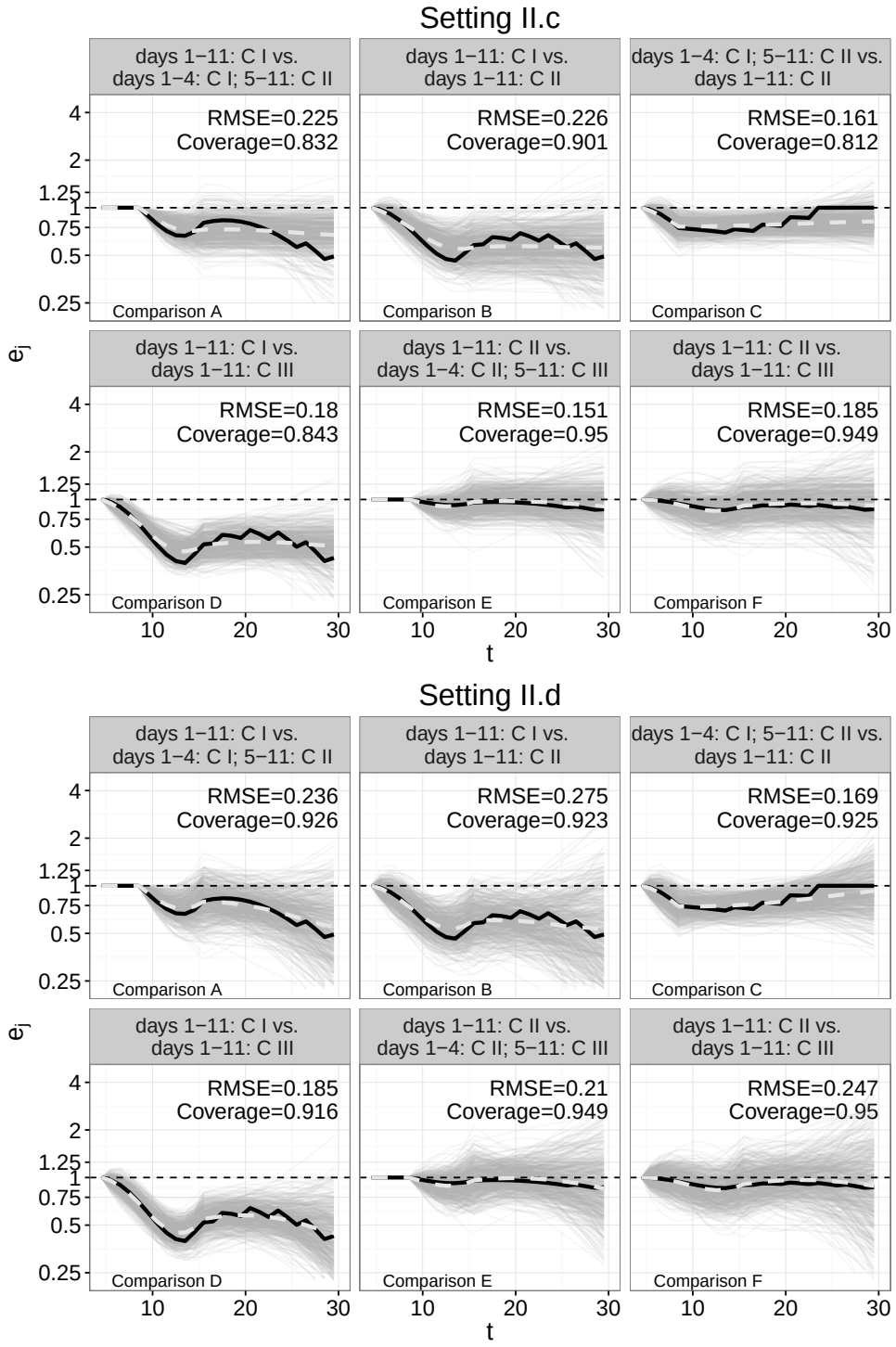

Figure 10: Settings *II.c* and *II.d*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.

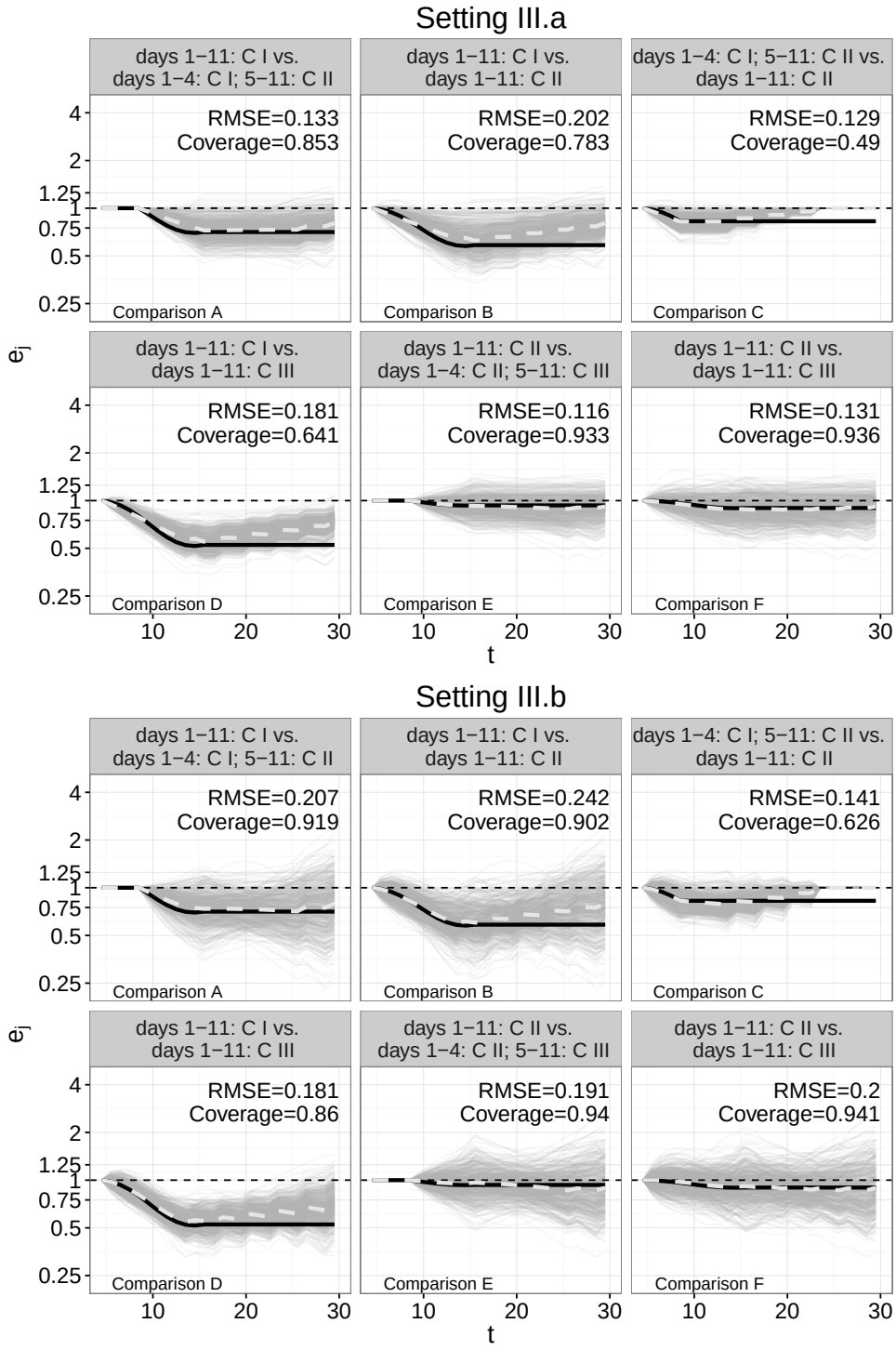

Figure 11: Settings *III.a* and *III.b*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.

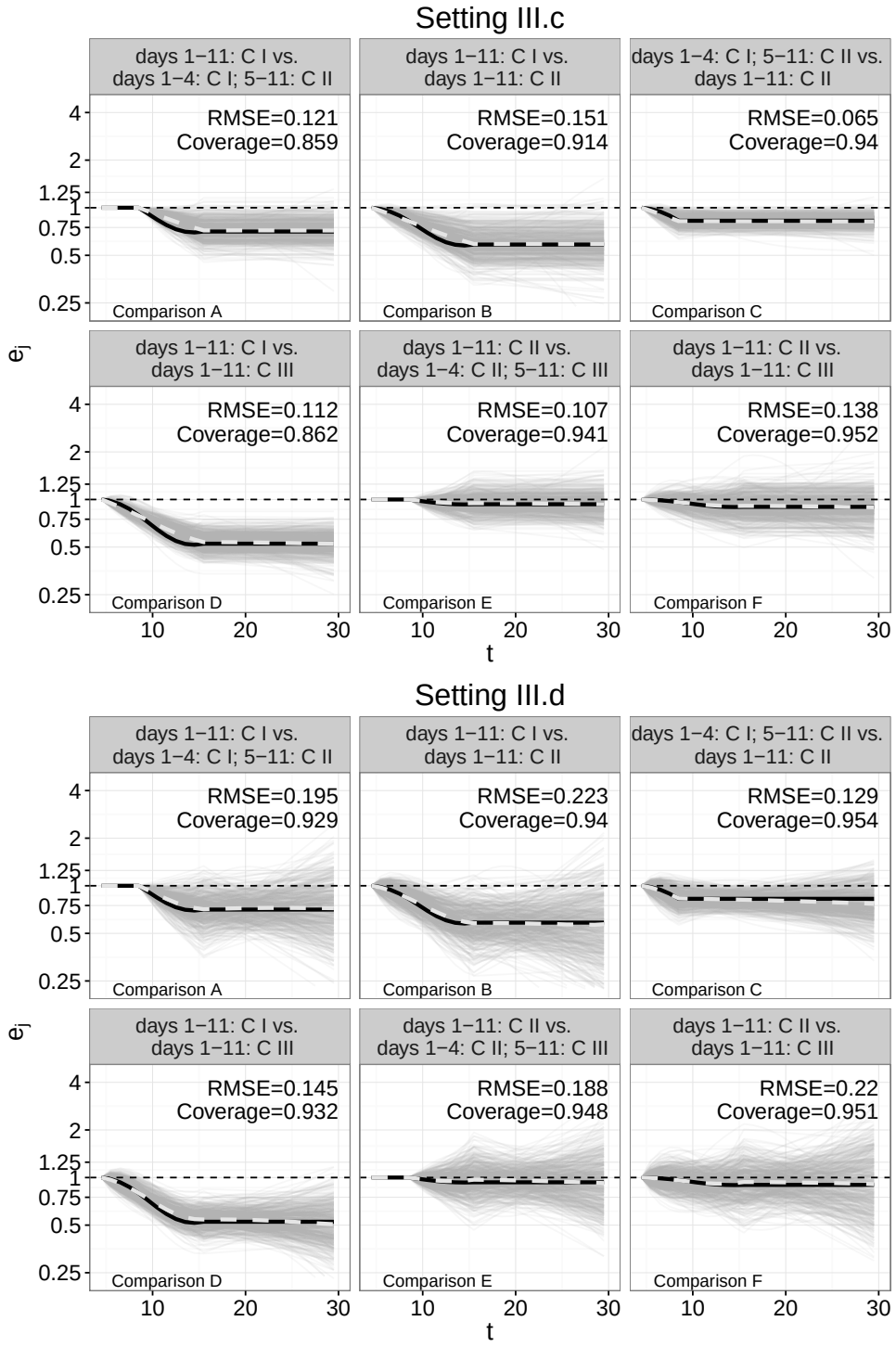

Figure 12: Settings *III.c* and *III.d*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.

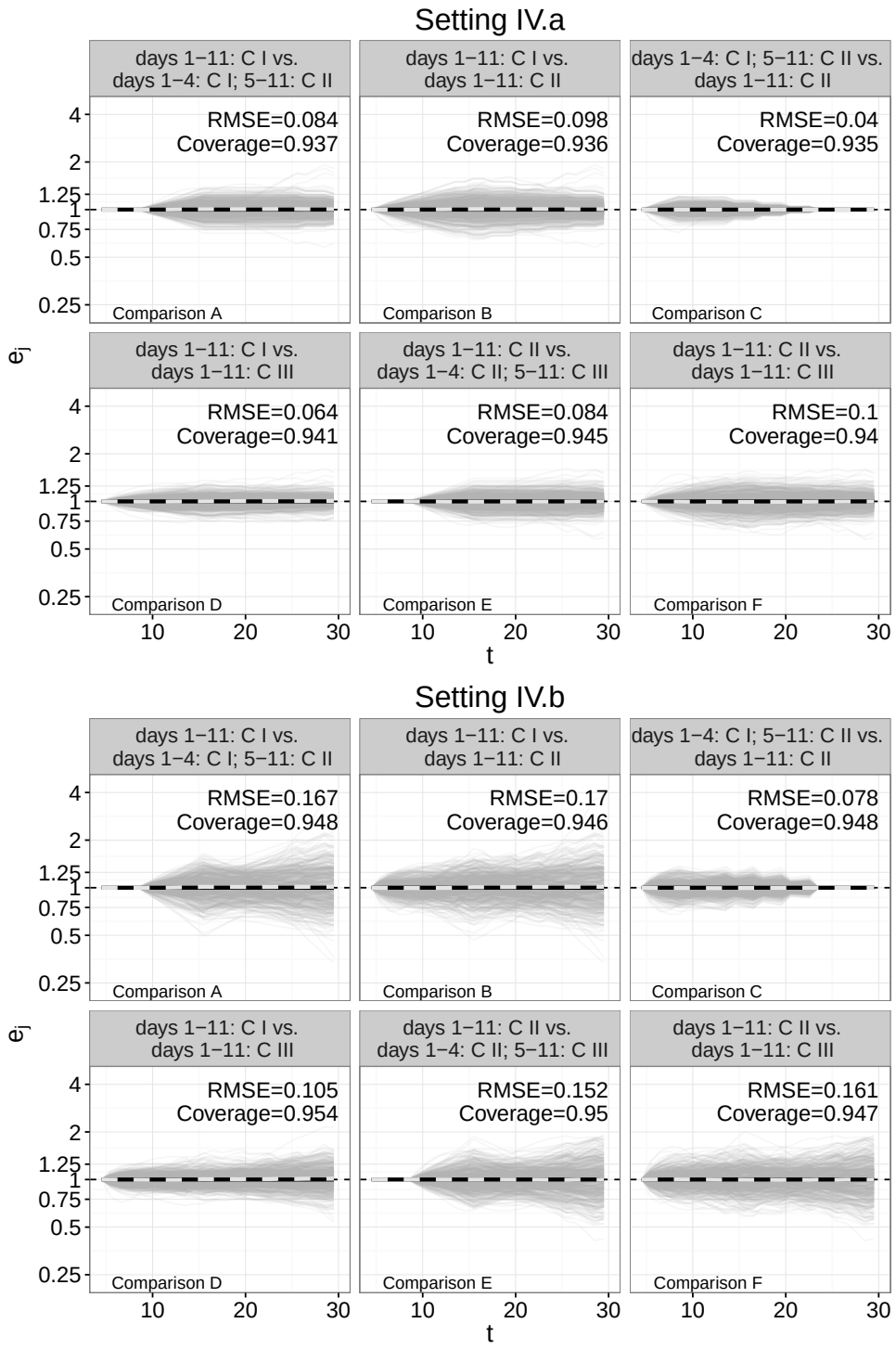

Figure 13: Settings *IV.a* and *IV.b*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.

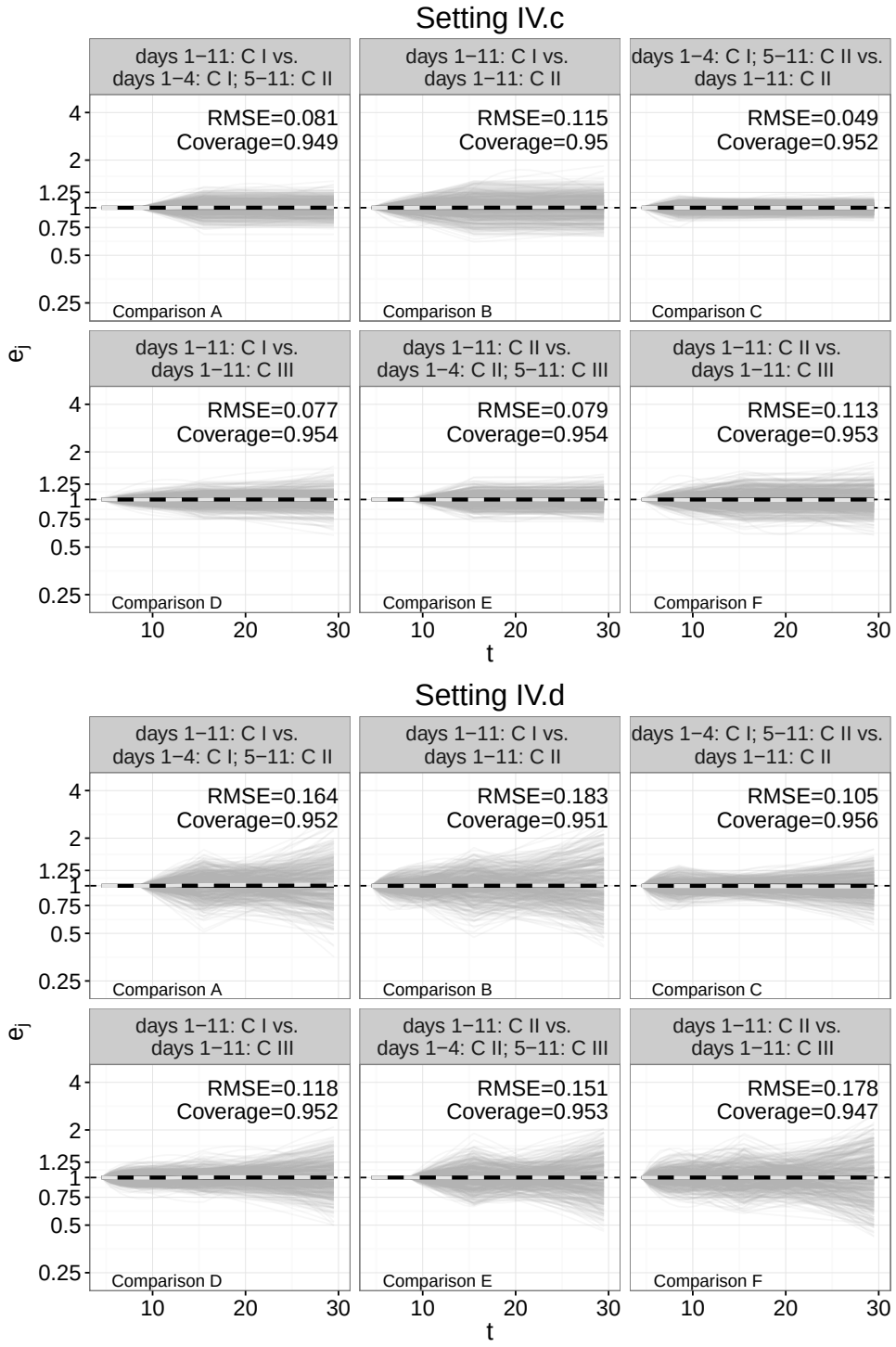

Figure 14: Settings *IV.c* and *IV.d*: Trajectories of hazard ratios $\hat{e}_j$ (gray lines) from all 500 simulation runs across six comparisons of nutrition protocols $\mathcal{Z}_i(t)$. The solid black line indicates true cumulative effects e_j. The dashed gray line indicates the average trajectories over all simulation runs.