Biostatistics (2017) **00**, 00, *pp.* 1–18
doi:10.1093/biostatistics/kxx065

Time-varying proportional odds model for mega-analysis of clustered event times

TANYA P. GARCIA*

Texas A&M University, Department of Epidemiology and Biostatistics, TAMU 1266, College Station, TX 77843-1266 USA

tpgarcia@sph.tamhsc.edu

KAREN MARDER

Columbia University Medical Center, Department of Neurology and Psychiatricy, Sergievsky Center and Taub Institute, 630 West 168th Street, New York, NY 10032, USA

YUANJIA WANG

Columbia University, Department of Biostatistics, Mailman School of Public Health, New York, NY 10032, USA

SUMMARY

Mega-analysis, or the meta-analysis of individual data, enables pooling and comparing multiple studies to enhance estimation and power. A challenge in mega-analysis is estimating the distribution for clustered, potentially censored event times where the dependency structure can introduce bias if ignored. We propose a new proportional odds model with unknown, time-varying coefficients, and random effects. The model directly captures event dependencies, handles censoring using pseudo-values, and permits a simple estimation by transforming the model into an easily estimable additive logistic mixed effect model. Our method consistently estimates the distribution for clustered event times even under covariate-dependent censoring. Applied to three observational studies of Huntington's disease, our method provides, for the first time in the literature, evidence of similar conclusions about motor and cognitive impairments in all studies despite different recruitment criteria.

Keywords: Logistic mixed model; Mega-analysis; Proportional odds; Pseudo-values; Varying coefficients.

1. INTRODUCTION

Mega-analysis, or the meta-analysis of individual patient data, is the gold standard for jointly analyzing and synthesizing individual observations from multiple studies. It allows for direct between-study comparisons and more precise estimates as information is borrowed across studies. In the mega-analysis of time-to-event data, an important objective is estimating and comparing the distribution functions for multiple clinical events. A key challenge, not handled by standard approaches, is modeling clustered time-to-event data such as the ages of multiple, life-impacting events (e.g., cognitive and motor impairment). The multiple

*To whom correspondence should be addressed.

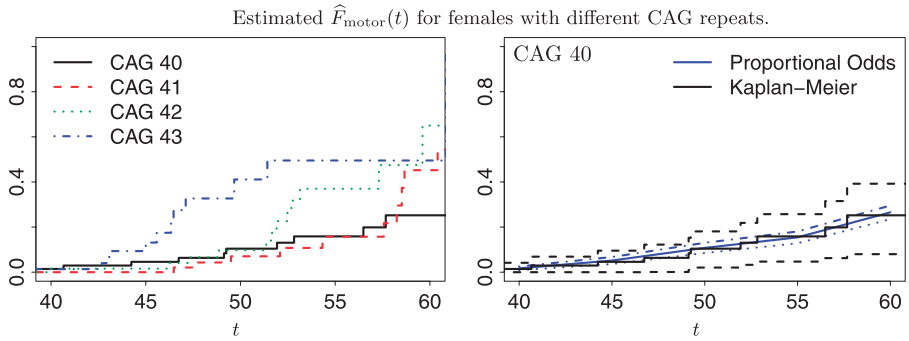

Fig. 1. Estimates of the distribution function for motor-diagnosis in the PREDICT study. Results shown for females with different CAG repeats; results were similar for COHORT and PHAROS. Left: The crossing of the Kaplan–Meier curves at different CAG repeats (e.g., curves corresponding to 40 CAG repeats and 41 CAG repeats) suggest that the proportional hazards assumption is violated. Right: Estimates from our time-varying proportional odds model are compared with the Kaplan–Meier curve at 40 CAG repeats. The proportional odds model overlaps well the Kaplan–Meier estimates indicating a good model fit.

events measured on each participant lead to clusters of correlated event times, and correct regression inferences depend on the proper modeling of the intraclass dependency and potential right-censoring of the event times.

Our work is motivated by three studies of Huntington's disease (HD). HD is a neurodegenerative disease caused by a repeat expansion of cytosine–adenine–guanine (CAG) in the huntingtin gene (Langbehn *and others*, 2010). Persons with more CAG repeats have earlier disease onset (Langbehn *and others*, 2010). The HD studies we analyzed are the Cooperative Huntington's Observational Research Trial (COHORT; Dorsey *and others*, 2012), the Prospective Huntington At Risk Observational Study (PHAROS; Huntington Study Group, 2006), and the Neurobiological Predictors of Huntington's Disease (PREDICT; Paulsen *and others*, 2014). A key interest is examining clinical events representative of HD progression such as age of first cognitive impairment and age of HD motor-diagnosis (see Section 4). Current analyses of these events are done for each study separately making between study comparisons intractable and use separate models for each event that ignore any dependency between them. We propose a new mega-analysis methodology that comprehensively analyzes the events and reveals that all studies indicate a similar relationship between cognitive and motor impairments despite different recruitment criteria.

Current methods for analyzing clustered event times use a Cox proportional hazards model. One method models the event times marginally as a Cox model, but this may not capture the intra-class dependency well and is inefficient (Chen *and others*, 2010). A second method is a proportional hazards frailty model (Clayton and Cuzick, 1985) which models the within-cluster correlation with a frailty in the hazard function. However, the proportional hazards assumption may not hold for clustered event time data nor for a mega-analysis when the studies have different follow-up lengths and censoring patterns (Zeng and Lin, 2007). The proportional hazards assumption is violated for the HD studies as seen from the crossing Kaplan–Meier curves in Figure 1.

Violations of the proportional hazards assumption are common. Adjusting the non-proportionality can be achieved with time-varying covariates, but constructing the appropriate time-function for the covariates may not be evident. When time-varying covariates are unavailable (as in our case), artificially constructing time-varying covariates can be misleading (Zeng and Lin, 2007).

An alternative is the proportional odds model which assumes the hazard ratio between two sets of covariate values converges to one. This is convenient when covariate effects diminish over time (e.g., treatment effects) and better fits our data (see Section 4.2.1). The proportional odds model has been

studied for clustered event times with scalar (Cai *and others*, 2002) and multivariate random effects (Zeng *and others*, 2005). But the methods are either computationally demanding in the variance estimation (Cai *and others*, 2002) or under heavy censoring (Zeng *and others*, 2005).

The aim of this article is to extend the proportional odds model with random effects to the mega-analysis of multiple event times prone to censoring. We contribute to the literature in three ways. First, our model incorporates time-varying, functional coefficients which assists in assessing the temporal impact of covariates on the right-censored event times. This is important in HD when evaluating genetic factors (e.g., CAG repeats) that are fixed values but may have a time-varying impact. Knowing when these genetic factors are more likely to impact clinical events can lead to adaptive treatment designs to deter HD progression. Time-varying coefficients in a proportional odds model have been studied, but not for clustered event time data, and the extension is non-trivial. Second, we propose a simpler estimation procedure that handles multiple data hierarchy and avoids unnecessary model assumptions. We handle censoring with pseudo-values (Logan *and others*, 2011) and functional covariate effects with splines, and ultimately cast the problem into an additive logistic mixed effect model. Within this simpler framework, we consistently estimate the functional parameters and distribution functions. Third, our approach helps to improve power, precision, and reliability of parameter estimates. Our approach is available in R software.

2. Methods

2.1. *Proposed time-varying, proportional odds model with random effects*

Clustered time-to-event data from multiple studies is multilevel data. The levels are event types (level one) observed for each subject (level two) in different studies (level three). Let e, i, s index levels one, two, and three, respectively, where $e = 1, \ldots, m$, $i = 1, \ldots, n_s$, and $s = 1, \ldots, S$. Let $(O_{eis}, \Delta_{eis}, \boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis})$ denote the observed data where $O_{eis} = \min(T_{eis}, C_{eis})$ with T_{eis} the event time, C_{eis} the censoring time, and $\Delta_{eis} = I(T_{eis} \leq C_{eis})$ the event indicator. The terms $\boldsymbol{X}_{eis} \in \mathbb{R}^{p_{\boldsymbol{x}}}$ and $\boldsymbol{Z}_{eis} \in \mathbb{R}^{p_{\boldsymbol{z}}}$ are model covariates.

Our objective is to model the distribution of the events accounting for data hierarchy and right-censoring. We propose a time-varying, proportional odds model with random effects:

$$\begin{aligned} &\text{logit}[\text{pr}\{T_{eis} \leq t | \boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}, \boldsymbol{W}_{eis}, \boldsymbol{R}_{is}(t)\}] \\ &= \beta_0(t) + \gamma_e(t) + \omega_s(t) + \boldsymbol{\beta}_{es}^{\mathrm{T}}(t)\boldsymbol{Z}_{eis} + \sum_{k=1}^{p_{\boldsymbol{x}}} \alpha_{esk}(X_{eisk}, t) + \boldsymbol{W}_{eis}^{\mathrm{T}}\boldsymbol{R}_{is}(t), \end{aligned} \tag{2.1}$$

where $\text{logit}(p) = \log\{p/(1-p)\}$. The motivation for this model is three fold. First, the logit link allows us to cast it into a simpler logistic regression framework similar to Efron (1988). However, our work differs from Efron (1988) in that we present a new, easy approach to handle censoring for clustered time-to-event data (Section 2.3). Second, the time-varying coefficients facilitate assessing the impact of covariates over time. Third, the subject-study random effects, $\boldsymbol{R}_{is}(t)$, capture dependencies between events measured on the same subject.

In model (2.1), the time-varying fixed effects include the intercept $\beta_0(t)$, the event-specific effect $\gamma_e(t)$, the study-specific effect $\omega_s(t)$, and the $p_{\boldsymbol{z}}$-dimensional coefficient $\boldsymbol{\beta}_{es}(t)$ associated with $\boldsymbol{Z}_{eis}$. For identifiability, we set $\gamma_1(t) = \omega_1(t) = 0$. Thus, $\beta_0(t)$ denotes the effect of the baseline study and event type; $\gamma_e(t)$, $e \neq 1$, denotes the difference between event type e and baseline event type; and $\omega_s(t)$, $s \neq 1$, denotes the difference between study s and baseline study. The parameters $\alpha_{esk}(X, t)$ are time-varying and smooth functionals in X to capture unknown, possibly non-linear covariate effects. The $\alpha_{esk}(X, t)$ are assumed to be twice differentiable in X to ensure smoothness and are centered for identifiability. We represent the effects of $\boldsymbol{X}_{eis}$ and $\boldsymbol{Z}_{eis}$ differently because we make non-parametric assumptions about $\boldsymbol{X}_{eis}$ and parametric assumptions about $\boldsymbol{Z}_{eis}$. In our application, X is the CAG-repeat length, an important

genetic marker associated with HD onset (Langbehn *and others*, 2010). Modeling CAG-repeats non-parametrically via $\alpha_{esk}(X,t)$ reduces the chance of misspecifying its impact on multiple HD events. In general, choosing between non-parametric and parametric assumptions is driven by the application. One could consider only non-parametric or only parametric effects, but our method does not fundamentally change. Thus, we proceed under the general scenario of both effect types. Lastly, model (2.1) captures the different dependencies in the data with the $p_{\boldsymbol{w}}$-dimensional random effect $\boldsymbol{R}_{is}(t)$. The associated covariates $\boldsymbol{W}_{eis}$ may consist of elements from $\boldsymbol{X}_{eis}$ and $\boldsymbol{Z}_{eis}$, and when $\boldsymbol{W}_{eis}$ is a vector of ones, $\boldsymbol{R}_{is}(t)$ is a random intercept. We assume that $\boldsymbol{R}_{is}(t) \sim \text{Normal}\{\boldsymbol{0}, \boldsymbol{\Omega}(t)\}$ with $\boldsymbol{\Omega}(t)$ unknown.

A variant of model (2.1) is one with shared study parameters:

$$\begin{aligned}&\text{logit}[\text{pr}\{T_{eis} \le t | \boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}, \boldsymbol{W}_{eis}, \boldsymbol{R}_{is}(t)\}] \\ &= \beta_0(t) + \gamma_e(t) + \boldsymbol{\beta}_e^{\mathrm{T}}(t)\boldsymbol{Z}_{eis} + \sum_{k=1}^{p_{\boldsymbol{x}}} \alpha_{ek}(X_{eisk}, t) + \boldsymbol{W}_{eis}^{\mathrm{T}}\boldsymbol{R}_{is}(t). \end{aligned} \tag{2.2}$$

Compared to model (2.1), model (2.2) assumes that the logit distribution function has a similar shape for all studies and thus combines all study information to more precisely estimate the parameter effects. In this case, the study-specific effect $\omega_s(t)$ is suppressed. In Section 2.4, we provide a formal procedure to help determine when to use model (2.1) or (2.2).

After estimating the parameters in model (2.1) (or its variant model (2.2)), we can integrate over the unobserved random effects to obtain the marginal distribution functions for specific events in each study. The marginal distribution, denoted as $F_{es}(t|\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) = \text{pr}(T_{es} \le t | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{W} = \boldsymbol{w})$, serves to compare the likeliness of events occurring. For HD, the marginal distributions help assess when cognitive and motor impairments are likely to occur for subpopulations defined by $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}$. We use the conditional model (2.1) as a starting point for our analysis rather than a marginal model to avoid incorrect inferences (Lee and Nelder, 2004).

A sufficient condition to ensure $F_{es}(t|\boldsymbol{x},\boldsymbol{z},\boldsymbol{w})$ is monotonically non-decreasing in t is to require $\beta_0(t)$, $\gamma_e(t)$, $\omega_s(t)$, $\boldsymbol{\beta}_{es}(t)$, and $\alpha_{esk}(x,t)$ be non-decreasing in t. Instead, we apply a post-estimation procedure to ensure $F_{es}(t|\boldsymbol{x},\boldsymbol{z},\boldsymbol{w})$ is proper (Section S.1 of supplementary material available at *Biostatistics* online).

2.2. *Estimation without censoring*

We first present the estimation procedure when censoring is ignored, and show how the method adjusts under censoring in Section 2.3. The estimation is presented for the general model (2.1); adjustments for the variant model (2.2) are noted. In general, we cast the model into an additive logistic mixed effects model for different $t = t_0$. At each t_0 (see Section 2.3.3 for choice of t_0's), we define a binary event status $Y_{eis}(t_0) = I(T_{eis} \le t_0)$ and approximate $\alpha_{esk}(x,t)$ with splines:

$$\alpha_{esk}(x,t_0) = \boldsymbol{B}_{esk}^{\mathrm{T}}(x)\boldsymbol{a}_{esk}(t_0), \quad e = 1,\ldots,m; s = 1,\ldots,S; k = 1,\ldots,p_{\boldsymbol{x}}. \tag{2.3}$$

Here, $\boldsymbol{a}_{esk}(t_0)$ is a $p_{\boldsymbol{a}}$-dimensional spline coefficient vector, and $\boldsymbol{B}_{esk}(\cdot)$ are known basis functions. The choice of regression splines and its knots has been extensively studied (de Boor, 2001), and we use penalized cubic regression splines with knots at equally spaced sample quantiles of $\boldsymbol{X}$.

Under these transformations, the model in (2.1) satisfies

$$E\{Y_{eis}(t_0) | \boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}, \boldsymbol{W}_{eis}, \boldsymbol{R}_{is}(t_0)\} = \frac{\exp[\eta_{eis}\{\boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}; \boldsymbol{\xi}(t_0)\} + \boldsymbol{R}_{is}^{\mathrm{T}}(t_0)\boldsymbol{W}_{eis}]}{1 + \exp[\eta_{eis}\{\boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}; \boldsymbol{\xi}(t_0)\} + \boldsymbol{R}_{is}^{\mathrm{T}}(t_0)\boldsymbol{W}_{eis}]}, \tag{2.4}$$

where $\eta_{eis}\{\boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}; \boldsymbol{\xi}(t_0)\} = \beta_0(t_0) + \gamma_e(t_0) + \omega_s(t_0) + \boldsymbol{\beta}_{es}^{\mathrm{T}}(t_0)\boldsymbol{Z}_{eis} + \sum_{k=1}^{p_{\boldsymbol{x}}} \boldsymbol{B}_{esk}{}^{\mathrm{T}}(X_{eisk})\boldsymbol{a}_{esk}(t_0)$. Here, $\boldsymbol{\xi}(t_0) = \{\beta_0(t_0), \gamma_1(t_0), \ldots, \gamma_m(t_0), \boldsymbol{\xi}_{11}^{\mathrm{T}}(t_0), \ldots, \boldsymbol{\xi}_{mS}^{\mathrm{T}}(t_0)\}^{\mathrm{T}}$ is all fixed-effect parameters where each $\boldsymbol{\xi}_{es}(t_0) = \{\omega_s, \boldsymbol{\beta}_{es}^{\mathrm{T}}(t_0), \boldsymbol{a}_{es1}^{\mathrm{T}}(t_0), \ldots, \boldsymbol{a}_{esp_{\boldsymbol{x}}}^{\mathrm{T}}(t_0)\}^{\mathrm{T}}$ is a vector of length $q = 1 + p_z + \sum_{k=1}^{p_{\boldsymbol{x}}} p_{\boldsymbol{a}}$ containing parameters in study s. Our transformations leading to (2.4) thus result in an additive logistic mixed effect model. Such a model belongs to the family of generalized additive mixed models and is estimated by maximizing the double penalized quasi-likelihood (DPQL).

In our case, the DPQL we maximize at each t_0 is

$$-\frac{1}{2}\sum_{s,i,e} d_{eis}\{y_{eis}(t_0), \theta_{eis}(t_0)\} - \frac{1}{2}\sum_{s,i} \boldsymbol{r}_{is}^{\mathrm{T}}\boldsymbol{\Omega}(t_0)^{-1}\boldsymbol{r}_{is} - \frac{1}{2}\sum_{s,k,e} \lambda_{esk}\boldsymbol{a}_{esk}{}^{\mathrm{T}}(t_0)\boldsymbol{\mathcal{K}}_{esk}\boldsymbol{a}_{esk}(t_0) \tag{2.5}$$

where $\theta_{eis}(t_0) = \mathrm{pr}\{T_{eis} \leq t_0 | \boldsymbol{x}_{eis}, \boldsymbol{z}_{eis}, \boldsymbol{w}_{eis}, \boldsymbol{r}_{is}; \boldsymbol{\xi}(t_0)\}$ denotes the expected value of $Y_{eis}(t_0)$ given $\boldsymbol{x}_{eis}, \boldsymbol{z}_{eis}, \boldsymbol{w}_{eis}, \boldsymbol{r}_{is}$. The first term in (2.5) is the likelihood part of the model with

$$d_{eis}\{y_{eis}(t_0), \theta_{eis}(t_0)\} \propto -2\int_{y_{eis}(t_0)}^{\theta_{eis}(t_0)} \frac{y_{eis}(t_0) - u}{h(u)}\mathrm{d}u$$

denoting the conditional deviance function of the parameters $\boldsymbol{\xi}(t_0)$ given $\boldsymbol{r}_{is}$ and $h(u) = u(1-u)$. The second and third terms in (2.5) are penalty terms. The second term is a penalty for the Laplace approximation of the log quasi-likelihood (Breslow *and others*, 1993) used to simplify the integration over the random effects. The third term is a penalty that determines the smoothness of $\alpha_{esk}(\cdot, t_0)$: we penalize the "wiggliness" of the spline approximation measured through $\boldsymbol{\mathcal{K}}_{esk} = \int \boldsymbol{B}''_{esk}(x)\left\{\boldsymbol{B}''_{esk}(x)\right\}^{\mathrm{T}}\mathrm{d}x$ and control the penalization with the smoothing parameter λ_{esk}.

Model parameters $\boldsymbol{\xi}(t_0)$, $\boldsymbol{\Omega}(t_0)$, and the smoothing parameters λ_{esk} are then estimated as follows. First, for fixed λ_{esk}, estimates for $\boldsymbol{\xi}(t_0)$ and $\boldsymbol{\Omega}(t_0)$ are derived from maximizing the DPQL in (2.5) using penalized iterative re-weighted least squares that maintain the identifiability constraints on $\alpha_{esk}(\cdot, t_0)$. Second, λ_{esk} are estimated using generalized approximate cross-validation. The approach is detailed in Wood (2006) and implemented in the `mgcv` package in R.

The estimation process leads to consistent estimates of $\boldsymbol{\beta}_0(t_0)$, $\gamma_e(t_0)$, $\omega_s(t_0)$, $\boldsymbol{\beta}_{es}(t_0)$, and $\boldsymbol{\Omega}(t_0)$ when the smoothing parameters $\lambda_{esk} \to 0$ and sample sizes $n_s \to \infty$, $s = 1, \ldots, S$ (Wood, 2006). This is because the penalized iterative re-weighted least squares maximizes the DPQL, and the likelihood part (i.e., the first term in (2.5)) dominates as the sample size increases. Therefore, the consistency of the maximum likelihood estimation ensures consistency of the model parameters.

For the spline approximations, when the number of knots is fixed, the penalties in the estimation may induce approximation bias in $\widehat{\boldsymbol{a}}_{esk}(t_0)$ (Marra and Radice, 2010). But the approximation bias will converge to zero asymptotically as the number of knots in the spline expansion increases with the sample size. Thus, consistency of the smoothing coefficients is ensured asymptotically.

2.3. *Estimation with censoring*

When event times are censored, the key change is the construction of $Y_{eis}(t_0) = I(T_{eis} \leq t_0)$. In two instances, the value of $Y_{eis}(t_0)$ is exactly known. When there is no censoring (i.e., $\Delta_{eis} = 1$), and when $\Delta_{eis} = 0$ and $C_{eis} \geq t_0$. In the latter case, we have $t_0 \leq C_{eis} < T_{eis}$ implying that $Y_{eis}(t_0) = 0$. Otherwise, when $\Delta_{eis} = 0$ and $C_{eis} < t_0$, $Y_{eis}(t_0)$ is unobservable. We propose to replace this unobservable $Y_{eis}(t_0)$ with a pseudo-value (Logan *and others*, 2011) and show that doing so still yields consistent estimates when maximizing the DPQL.

2.3.1. *Pseudo-values under independent censoring* We first discuss computing pseudo-values when censoring is independent of covariates; see Section 2.3.2 when the assumption is relaxed.

When we assume distinct study parameters as in model (2.1), the unobservable $Y_{eis}(t_0)$ is replaced with the jack-knife pseudo-value

$$Y^D_{eis}(t_0) = n_s\widehat{F}_{es}(t_0) - (n_s - 1)\widehat{F}^{-(i)}_{es}(t_0),$$

where the superscript D denotes *distinct*. Here, $\widehat{F}_{es}(t_0) = 1 - \widehat{S}_{es}(t_0)$ where $\widehat{S}_{es}(t_0)$ is the Kaplan–Meier estimator using information about event type e from all n_s subjects in study s. Likewise, $\widehat{F}^{-(i)}_{es}(t_0)$ is the Kaplan–Meier estimator after removing observation i from study s. The Kaplan–Meier estimators are study-specific because model (2.1) assumes distinct study parameters and are event-specific because the event types are different and non-competing.

The pseudo-value $Y^D_{eis}(t_0)$ satisfies the following properties:

(P1) $E\{Y^D_{eis}(t_0)\} = \text{pr}(T_{eis} \leq t_0)$ and implies that the pseudo-value estimates are unbiased.
(P2) When there is no censoring prior to t_0, $Y^D_{eis}(t_0)$ simplifies to $Y_{eis}(t_0) = I(T_{eis} \leq t_0)$.
(P3) For $i \neq i'$, $Y^D_{eis}(t_0)$ and $Y^D_{e'i's}(t_0)$ are approximately independent as n_s tends to infinity.
(P4) $\lim_{n_s\to\infty} E\{Y^D_{eis}(t_0)|\boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}, \boldsymbol{W}_{eis}, \boldsymbol{R}_{is}(t)\} = \text{pr}\{T_{eis} \leq t_0|\boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}, \boldsymbol{W}_{eis}, \boldsymbol{R}_{is}(t)\}$.

Property (P1) holds because the Kaplan–Meier is an (approximately) unbiased estimator for dependent data (Ying and Wei, 1994). The resulting unbiasedness shows that $Y^D_{eis}(t_0)$ satisfies the bias-correction property of jack-knife estimates in general. Property (P2) holds because when there is no censoring, the Kaplan–Meier estimator reduces to $\widehat{S}_{es}(t_0) = (n_{is} - d_{is})/n_s$ for $t_i \leq t_0 \leq t_{i+1}$ where t_i denotes the time of event for person i, d_{is} denotes the number of events at t_i in study s, and n_{is} denotes the number in the risk set in study s at time t_i. Plugging in this simplification to $Y^D_{eis}(t_0)$ leads to (P2). Under no censoring, we may thus use the pseudo-value $Y^D_{eis}(t_0)$ even when the binary indicator $Y_{eis}(t_0)$ is observable because the two are equivalent. Properties (P3) and (P4) are proven in Section S.2 of supplementary material available at *Biostatistics* online. They imply that (asymptotically) the relationship between pseudo-values and the covariates and random effects is exactly the conditional distribution $\text{pr}\{T_{eis} \leq t_0|\boldsymbol{X}_{eis}, \boldsymbol{Z}_{eis}, \boldsymbol{W}_{eis}, \boldsymbol{R}_{is}(t)\}$ in (2.1). This means we may replace $Y_{eis}(t_0)$ with $Y^D_{eis}(t_0)$ in the DPQL (2.5), treating the latter as the raw data. The terms $\{Y^D_{eis}(t_0) - \theta_{eis}(t_0)\}$ in the DPQL are approximately independent by (P3) and have mean zero by (P4). The DPQL estimating equations thus are unbiased even when using pseudo-values.

When we have shared study parameters as in model (2.2), the jack-knife pseudo-values are

$$Y^I_{eis}(t_0) = N\widehat{F}_e(t_0) - (N - 1)\widehat{F}^{-(i)}_e(t_0),$$

where $N = \sum_{s=1}^{S} n_s$ and superscript I emphasizes *indistinct* or shared parameters. To form $Y^I_{eis}(t_0)$, we combine study information. Here, $\widehat{F}_e(t_0) = 1 - \widehat{S}_e(t_0)$ where $\widehat{S}_e(t_0)$ is the Kaplan–Meier estimator using information about event type e from the N subjects over all studies, and $\widehat{F}^{-(i)}_e(t_0)$ is a similar estimator after removing observation i. Analogous properties to those in (P1)–(P4) hold for $Y^I_{eis}(t_0)$; proofs involve summing over all studies (e.g., n_s replaced with N).

2.3.2. *Pseudo-values under covariate-dependent censoring* When censoring depends on a discrete covariate U with values $1, 2, \ldots, m_u$, we adjust the pseudo-values to be covariate-specific. For example, when subject i has covariate value $U = u$, the pseudo-value $Y^D_{eis}(t_0)$ would change as follows. First, we would replace $\widehat{F}_{es}(t_0)$ with $\widehat{F}_{esu}(t_0) = 1 - \widehat{S}_{esu}(t_0)$ where $\widehat{S}_{esu}(t_0)$ is the Kaplan–Meier estimator for event type e based on all subjects in study s with covariate $U = u$. Second, we would replace the sample size n_s with n_{su}, the number of subjects in study s with covariate $U = u$. Similarly, pseudo-value $Y^I_{eis}(t_0)$

would now involve $\widehat{F}_{eu}(t_0) = 1 - \widehat{S}_{eu}$ where $\widehat{S}_{eu}$ is the Kaplan–Meier estimator for event type e based on subjects from all studies with $U = u$, and sample size N_u, the total number of subjects from all studies with $U = u$.

Our idea for covariate-specific pseudo-values is motivated by Andersen and Perme (2010). For a non-mega-analysis setting, they showed that this approach corrects the bias introduced when covariate-dependent censoring is ignored, but induces higher variability. We show similar features in our numerical study (Section 3). When there is covariate-dependent censoring, Properties (P1) to (P4) in Section 2.3.1 remain valid. Properties (P1) and (P2) hold because even when the Kaplan–Meier estimator is stratified by discrete covariates, it remains unbiased and reduces to a simple form when there is no censoring. Properties (P3) and (P4) hold because their proofs extend to the covariate-dependent censoring case; see end of Section S.2 of supplementary material available at *Biostatistics* online.

When censoring depends on a continuous U, we may construct pseudo-values as above using the kernel-weighted Kaplan–Meier estimator. But this procedure becomes onerous when U is multivariate for discrete and continuous cases. An alternative is to fit a Cox regression for censoring times. These require a separate investigation, and we explore discrete covariates in this article.

2.3.3. *Main algorithm* We now summarize our algorithm when the event times are prone to censoring. Implementation details are in Section S.3 of supplementary material available at Biostatistics online. We cast our model into an additive logistic mixed effects model at different $t = t_0$. Our marginal approach leads to random effects $R_{is}(t_0)$ which induces association between events at each t_0. Because we apply our procedure to a range of t_0 values, the association between events is maintained over this entire range. In theory, the exact number of t_0's influences how smoothly $\boldsymbol{\beta}_{es}(t)$, $\alpha_{esk}(x,t)$ is approximated over t, where more t_0's generally give more wiggly estimates. Because estimating at all t_0 values is infeasible, we recommend choosing at least five t_0 values evenly spaced across the range of $O_{eis} = \min(T_{eis}, C_{eis})$, in addition to those t_0 of key interest for a specific application.

At each t_0, $\alpha_{esk}(x,t_0)$ is approximated using regression splines in equation (2.3), and we define $Y_{eis}(t_0) = I(T_{eis} \leq t_0)$. When $Y_{eis}(t_0)$ is unobservable, it is replaced by a pseudo-value that accounts for independent censoring (Section 2.3.1) or covariate-dependent censoring (Section 2.3.2), and if the model assumes distinct study parameters ($Y^D_{eis}(t_0)$ or $Y^I_{eis}(t_0)$). Under these transformations, model (2.1) becomes an additive logistic mixed effect model whose parameters are estimated by maximizing the DPQL in (2.5). Estimated parameters are then plugged in, and the marginal distributions $F_{es}(t|\boldsymbol{x},\boldsymbol{z},\boldsymbol{w})$ is obtained after integrating out the random effects.

Consistent estimation by maximizing the DPQL holds even when $Y_{eis}(t_0)$ are pseudo-values. First, maximizing the DPQL leads to consistent estimates of $\boldsymbol{\beta}_0(t_0)$, $\gamma_e(t_0)$, $\omega_s(t_0)$, $\boldsymbol{\beta}_{es}(t_0)$, and $\boldsymbol{\Omega}(t_0)$ because the likelihood part in equation (2.5) (i.e., the first term) dominates as the sample size increases. Even when $Y_{eis}(t_0)$ are pseudo-values, the likelihood part remains unbiased because it involves $\{Y_{eis}(t_0) - \theta_{eis}(t_0)\}$ which is approximately independent by property (P3) and has mean zero by property (P4) in Section 2.3.1. Second, consistency of the spline approximation of $\alpha_{esk}(x,t_0)$ holds with pseudo-values because the pseudo-value estimation converges at a rate faster than the convergence rate of a non-parametric regression (Logan *and others*, 2011). Thus, heuristically, uncertainty from pseudo-value estimation does not affect the asymptotic results of $\alpha_{esk}(x,t_0)$.

2.4. *Performing inference over a range of time points*

In Section S.4 of supplementary material available at *Biostatistics* online, we discuss inference of the parameters at a fixed $t = t_0$. To learn the overall functional behavior of the parameters, a comprehensive

test of interest is

$$H_0 : \mu_{es}(t) = \mu_{es'}(t) \text{ for all } t \quad \text{vs.} \quad H_A : \mu_{es}(t) \neq \mu_{es'}(t) \text{ for at least one } t, s \neq s'. \tag{2.6}$$

Here, $\mu_{es}(t)$ is a general parameter that can be: (i) $F_{es}(t|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w})$ for fixed $\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}$ to test if the marginal distribution for event e differs between study s and s'; (ii) $\boldsymbol{\beta}_{es}(t)$ to test if the effect associated with $\boldsymbol{Z}_{eis}$ for event e differs between study s and s'; and (iii) $\alpha_{esk}(x, t)$ for a fixed x to test if the effect of x for event e differs between study s and s'.

This test assesses if functional terms between studies *differ over a range of time points*, and if not, we could use shared study parameters. For example, if one fails to reject the null hypothesis that $\boldsymbol{\beta}_{es}(t) = \boldsymbol{\beta}_{es'}(t)$ for all t, then $\boldsymbol{\beta}_{es}^{\mathrm{T}}(t)\boldsymbol{Z}_{eis}$ in (2.1) could be replaced by $\boldsymbol{\beta}_{e}^{\mathrm{T}}(t)\boldsymbol{Z}_{eis}$. Doing so would reduce the number of parameters and the combined study information would improve the precision and efficiency of parameter effects. If we conclude that all distinct study parameters in (2.1) are common over all studies, then we would have the shared study parameter model (2.2).

A challenge in testing the hypothesis in (2.6) is comparing the functional processes $\mu_{es}(t)$ and $\mu_{es'}(t)$ over all t. We address this by modifying the bootstrap method of Crainiceanu *and others* (2012) for our problem. The idea is to construct a joint confidence interval, $CI_J(t)$, for $\mu_{es}(t) - \mu_{es'}(t)$ over a range of t-values that corrects for multiple comparisons and accounts for correlations at nearby time points. If at any t, $CI_J(t)$ includes the null value, we fail to reject H_0 in (2.6) and otherwise, we reject. The form of $CI_J(t)$ is in Section S.5 of supplementary material available at *Biostatistics* online.

3. SIMULATION STUDIES

3.1. *Simulation study design*

We empirically assessed four aspects of our method. First, we evaluated our algorithm (Section 2.3.3) by evaluating the consistency, efficiency, and pointwise 95% confidence intervals of the functional parameters in models (2.1) and (2.2). Second, we evaluated the performance of assessing study differences between functional parameters using the bootstrap-based joint confidence intervals (Section 2.4). Third, we evaluated how parameter estimates were affected if study data were analyzed separately rather than our proposed joint method. Fourth, we assessed the robustness of our method to covariate-dependent censoring.

We generated data similar to the HD data (Section 4) except we considered $S = 2$ studies and $m = 2$ events for ease in presentation. We set the sample sizes as $n_1 = 400$ for Study 1, and $n_2 = 500$ for Study 2. We set $p_x = p_z = 1$ so that X and Z were one-dimensional covariates. We generated X from a uniform [0,1] distribution and Z from a binary distribution to mimic the HD data with standardized CAG repeat length (X) and a binary gender variable (Z). We used a random, level-two intercept $\boldsymbol{R}_{is}(t) \sim \text{Normal}(0, 1)$.

The functional coefficients in (2.1) were set in two different ways. Let $g(t) = \log(t/50)$.

1. *Parameter setting A: True model has distinct study parameters.* We set $\beta_0(t) = 0.25g(t)$, $\beta_{11}(t) = g(t), \beta_{21}(t) = 0.25g(t), \beta_{12}(t) = 0.25g(t), \beta_{22}(t) = g(t)$; $\alpha_{11}(x, t) = 2xg(t)$, $\alpha_{21}(x, t) = 2\sin(\pi x)g(t)$, $\alpha_{12}(x, t) = \sin(5\pi x)g(t)$, $\alpha_{22}(x, t) = xg(t)$; $\gamma_2(t) = 0.25g(t)$, $\omega_2(t) = g(t)$.
2. *Parameter setting B: True model has shared study parameters.* We set $\beta_0(t) = 0.25g(t)$, $\beta_{11}(t) = \beta_{12}(t) = g(t), \beta_{21}(t) = \beta_{22}(t) = 0.25g(t)$. Also, $\alpha_{11}(x, t) = \alpha_{12}(x, t) = 2xg(t)$, $\alpha_{21}(x, t) = \alpha_{22}(x, t) = \sin(5\pi x)g(t)$. Lastly, $\gamma_2(t) = 0.25g(t)$ and $\omega_2(t) = 0$.

For both settings, we generated independent and covariate-dependent censoring. We generated independent censoring times from uniform distributions that led to 40% censoring for the first event type and 50% for the second. We generated covariate-dependent censoring as follows. When $Z = 1$, uniformly distributed

censoring times were generated that induced 40% censoring, and when $Z = 0$, uniformly distributed censoring times were generated that induced 50% censoring. See Section S.6 of supplementary material available at *Biostatistics* online for results at higher censoring.

3.1.1. *Analysis models evaluated* We evaluated the simulated data with three models. Each had random intercept $\boldsymbol{R}_{is}(t)$ to capture dependencies between events measured on the same subject.

1. *Model A: Joint model with distinct study parameters*. Event information from all studies was analyzed jointly with (2.1) using distinct study parameters and pseudo-values $Y^D_{eis}(t_0)$.
2. *Model B: Shared study parameter model*. Event information from all studies was analyzed jointly with model (2.2) using shared study parameters and pseudo-values $Y^I_{eis}(t_0)$.
3. *Model C: Separate study model*. Event information from each study s^* is analyzed separately:

$$\text{logit}[\text{pr}\{T_{eis^*} \le t | X_{eis^*}, Z_{eis^*}, R_{is^*}(t)\}] = \beta_{s^*0}(t) + \gamma_{es^*}(t) + \sum_{k=1}^{p_z} \beta_{es^*k}(t) Z_{eis^*k} + \alpha_{es^*}(X_{eis^*}, t) + R_{is^*}(t).$$

This model is similar to Model A except that the intercept $\beta_{s^*0}(t)$ and failure type main effect $\gamma_{es^*}(t)$ are different for each study s^*, the study main effect is absorbed into the intercept, and $\gamma_{1s^*}(t) \equiv 0$ for identifiability. Estimation with this model uses pseudo-values $Y^D_{eis}(t_0)$.

Comparing Models A and B assesses differences in assuming distinct study parameters (Model A) vs. shared (Model B). We show that the bootstrap-based joint confidence interval (Section 2.4) performs well in determining when to choose Model B over A. Comparing Models A and C assesses efficiency gains when analyzing studies simultaneously (Model A) or not (Model C).

3.1.2. *Performance evaluation* For all models, we estimated the functional parameters and the marginal distribution function $F_{es}(t|\boldsymbol{x}, \boldsymbol{z}) = \text{pr}(T_{es} \le t | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z})$ with the algorithm in Section 2.3.3. Parameters were estimated at 21 equally spaced time points t_0 in [40,60], 100 equally spaced x-values in [0,1], and for $z = 0, 1$. We evaluated the bias, variability, and estimated 95% pointwise coverages of the estimates based on 250 simulations. Estimated variances were obtained using the asymptotic results in Section 2.4. Bias is reported as average absolute bias, calculated via $\sum_{\ell=1}^{L} |\widehat{\beta(t_{0\ell})} - \beta_0(t_{0\ell})|/L$, and similarly for $\alpha(x, t)$. Here, $\beta_0(t_{0\ell})$ is the truth and $\widehat{\beta(t_{0\ell})}$ is the average estimate based on 250 Monte Carlo simulations.

We also evaluated how well the bootstrap-based joint confidence intervals, $CI_J(t)$, in Section 2.4 could detect if studies had distinct study parameters (Parameter Setting A) or shared study parameters (Parameter Setting B). For data generated under each setting, we obtained parameter estimates using Model A and computed the differences $d_e(t) = \mu_{es}(t) - \mu_{es'}(t)$ for studies $s = 1$ and $s' = 2$ with $\mu_{es}(t)$ being (i) $\omega_s(t)$, (ii) $\beta_{es}(t)$, (iii) $\alpha_{es}(x, t)$ at $x = 0.5$, and (iv) $F_{es}(t|x, z)$ at $x = 0.5$, $z = 0, 1$. When the true model has distinct study parameters (Parameter Setting A), we expect $d_e(t)$ to be non-null so that H_0 in (2.6) would be rejected for all t. When the true model has shared study parameters (Parameter Setting B), we expect $d_e(t)$ to be null so that we would fail to reject H_0 in (2.6) for at least one t. To assess how often these ideal differences were obtained, we computed the integrated absolute coverages (IAC) of the joint confidence intervals: $\widehat{\text{IAC}} = \sum_{k=1}^{K} 1\{d_e(t) \in CI_J^{(k)}(t) \text{ for every } t\}/K$, where $1(\cdot)$ is an indicator function and $K = 250$ simulations. Each $CI_J(t)$ was computed using the procedure in Section S.5 of supplementary material available at *Biostatistics* online based on estimates from Model A and 100 bootstrap replicates. Large IAC values means the true differences $d_e(t)$ are detected; an IAC of 95% is ideal.

3.2. *Results*

We first evaluated the performance of our algorithm (Section 2.3.3) when applying Models A, B, and C to data generated with distinct study parameters (Parameter Setting A) and shared study parameters (Parameter Setting B). We first discuss results for independent censoring. When the true model has distinct study parameters and the correct model (Model A) is applied, our estimation procedure had negligible bias, the estimated variances closely matched the observed variances, and the 95% coverage probabilities were at the nominal levels (Table 1 top panel). When we used Model B which assumes shared study parameters rather than distinct, we had increased bias especially in the $\widehat{\alpha}(x,t)$ terms, and coverage probabilities less than the nominal level (bold-faced entries in middle panel of Table 1). This poor performance is not surprising given that the true model has more parameters than Model B can estimate.

When the true model has shared study parameters (Parameter Setting B), Model B performs well with negligible bias, similar estimated and empirical variances, and 95% coverage probabilities matching the nominal levels (Table S.1 middle panel of supplementary material available at *Biostatistics* online). Model A also performs well even though it falsely assumes distinct study parameters (Table S.1 top panel of supplementary material available at *Biostatistics* online). The results suggest that Model A is a more reasonable choice when the true model is unknown.

The uncertainty between using Models A and B can be avoided with the bootstrap-based joint confidence interval. Table 2 shows the estimated IAC associated with testing parameter differences between studies. The IAC closely matched the 95% nominal level for all study differences assessed. This suggests that the algorithm in Section 2.4 detected the true differences well and can help to choose between Models A and B. If, for any parameter, the joint confidence interval contains a non-null value at any t, use Model A; else, use Model B.

When analyzing the study data separately (Model C) rather than jointly (Model A), we found that for both parameter settings, Model C led to unbiased but more variable estimates. The variability for all parameter estimates was nearly double the variability observed from Model A (compare the first and third panel of variance estimates in Table 1; similarly in Table S.1 of supplementary material available at *Biostatistics* online). The lower variability with Model A suggests efficiency gains of jointly analyzing multiple study data which is advantageous for improving power and precision of hypothesis testing.

Lastly, we evaluated our method's robustness to covariate-dependent censoring using three scenarios: (i) The true model has covariate-dependent censoring, and the estimation procedure correctly assumes this dependence. (ii) The true model has covariate-dependent censoring, and the estimation procedure incorrectly assumes independence. (iii) The true model has independent censoring, and the estimation procedure incorrectly assumes covariate-dependent censoring. For each scenario, we applied Model A and computed covariate-dependent pseudo-values (Section 2.3.2) for scenarios (i) and (iii), and standard pseudo-values (Section 2.3.1) for scenario (ii).

Overall, regardless of the true censoring mechanism, when the estimation assumes covariate-dependent censoring, our method performs satisfactorily (Table 3 top and bottom panel). When there is covariate-dependent censoring and this assumption is ignored, our method yields bias because the pseudo-values should be covariate-specific and are not (bold-faced entries in middle panel of Table 3). We thus recommend that if the censoring mechanism is unknown, apply our method assuming covariate-dependent censoring as it is robust to the true censoring mechanism.

In summary, our proposed model (Model A) is flexible in estimating parameters whether they are distinct across studies or not, and is robust to covariate-dependent censoring. The joint confidence intervals also provides guidance in when to choose between Models A and B.

Table 1. *Results when the true model has distinct study parameters and we apply three different models. We report average absolute bias* (*abs bias*)*, empirical variance* (*emp var*)*, estimated variance* (*est var*)*, and 95% coverage probabilities.* $\widehat{\gamma}(\cdot)$*,* $\widehat{\beta}(\cdot)$ *denotes results averaged over* $t \in [40, 60]$*;* $\widehat{\alpha}(0.50, \cdot)$ *is results at* $x = 0.50$ *averaged over* $t \in [40, 60]$*, and* $\widehat{\alpha}(\cdot, 45)$*,* $\widehat{\alpha}(\cdot, 55)$ *are results at* $t = 45$*,* $t = 55$*, respectively, averaged over* $x \in [0, 1]$

	Study 1				Study 2			
	abs bias	emp var	est var	95% cov	abs bias	emp var	est var	95% cov
	Model A: Joint model with distinct study parameters							
$\widehat{\beta}_0(\cdot)$	0.037	0.167	0.145	0.932	0.037	0.167	0.145	0.932
$\widehat{\beta}_{1s1}(\cdot)$	0.047	0.040	0.040	0.944	0.012	0.029	0.032	0.958
$\widehat{\beta}_{2s1}(\cdot)$	0.012	0.042	0.040	0.934	0.061	0.030	0.032	0.934
$\widehat{\gamma}_2(\cdot)$	0.005	0.149	0.137	0.934	0.005	0.149	0.137	0.934
$\widehat{\omega}_2(\cdot)$	—	—	—	—	0.031	0.209	0.198	0.939
$\widehat{\alpha}_{1s}(0.5, \cdot)$	0.038	0.193	0.165	0.940	0.054	0.137	0.137	0.950
$\widehat{\alpha}_{1s}(\cdot, 45)$	0.019	0.194	0.183	0.940	0.030	0.152	0.151	0.945
$\widehat{\alpha}_{1s}(\cdot, 55)$	0.047	0.191	0.164	0.929	0.073	0.134	0.136	0.943
$\widehat{\alpha}_{2s}(0.5, \cdot)$	0.086	0.169	0.165	0.938	0.022	0.126	0.137	0.968
$\widehat{\alpha}_{2s}(\cdot, 45)$	0.017	0.185	0.182	0.949	0.008	0.141	0.151	0.958
$\widehat{\alpha}_{2s}(\cdot, 55)$	0.081	0.176	0.164	0.933	0.034	0.124	0.136	0.964
$\widehat{F}_{1s}(\cdot \mid x = 0.5, z = 0)$	0.004	0.002	0.003	0.968	0.010	0.002	0.002	0.938
$\widehat{F}_{1s}(\cdot \mid x = 0.5, z = 1)$	0.009	0.003	0.003	0.953	0.012	0.002	0.002	0.945
$\widehat{F}_{2s}(\cdot \mid x = 0.5, z = 0)$	0.011	0.003	0.003	0.943	0.007	0.002	0.002	0.937
$\widehat{F}_{2s}(\cdot \mid x = 0.5, z = 1)$	0.013	0.003	0.003	0.942	0.014	0.002	0.002	0.915
	Model B: Joint model with shared study parameters							
$\widehat{\beta}_0(\cdot)$	0.032	0.096	0.083	0.933	0.032	0.096	0.083	0.933
$\widehat{\beta}_{11}(\cdot)$	0.067	0.018	0.018	0.917	0.017	0.018	0.018	0.947
$\widehat{\beta}_{21}(\cdot)$	0.040	0.019	0.018	0.938	0.070	0.019	0.018	0.901
$\widehat{\gamma}_2(\cdot)$	0.038	0.154	0.137	0.919	0.038	0.154	0.137	0.919
$\widehat{\alpha}_1(0.5, \cdot)$	0.052	0.088	0.087	0.943	0.052	0.088	0.087	0.943
$\widehat{\alpha}_1(\cdot, 45)$	0.057	0.104	0.096	0.937	0.034	0.104	0.096	0.943
$\widehat{\alpha}_1(\cdot, 55)$	0.063	0.097	0.089	0.931	0.034	0.097	0.089	0.937
$\widehat{\alpha}_2(0.5, \cdot)$	**0.137**	0.092	0.087	0.913	0.057	0.092	0.087	0.941
$\widehat{\alpha}_2(\cdot, 45)$	0.054	0.097	0.096	0.937	0.042	0.097	0.096	0.941
$\widehat{\alpha}_2(\cdot, 55)$	**0.105**	0.088	0.089	0.928	0.032	0.088	0.089	0.949
$\widehat{F}_1(\cdot \mid x = 0.5, z = 0)$	0.004	0.001	0.001	0.960	0.004	0.001	0.001	0.960
$\widehat{F}_1(\cdot \mid x = 0.5, z = 1)$	0.017	0.001	0.001	0.938	0.006	0.001	0.001	0.967
$\widehat{F}_2(\cdot \mid x = 0.5, z = 0)$	0.022	0.001	0.001	**0.879**	0.010	0.001	0.001	0.931
$\widehat{F}_2(\cdot \mid x = 0.5, z = 1)$	0.020	0.001	0.001	0.901	0.008	0.001	0.001	0.943
	Model C: Separate study model							
$\widehat{\beta}_{s0}(\cdot)$	0.036	0.216	0.194	0.930	0.108	0.165	0.149	0.926
$\widehat{\beta}_{1s1}(\cdot)$	0.038	0.041	0.042	0.950	0.012	0.034	0.032	0.932
$\widehat{\beta}_{2s1}(\cdot)$	0.030	0.048	0.042	0.930	0.051	0.034	0.032	0.930
$\widehat{\gamma}_{2s}(\cdot)$	0.027	0.336	0.317	0.934	0.055	0.284	0.246	0.923
$\widehat{\alpha}_{1s}(0.5, \cdot)$	0.041	0.204	0.204	0.942	0.052	0.159	0.156	0.941
$\widehat{\alpha}_{1s}(\cdot, 45)$	0.027	0.245	0.224	0.940	0.028	0.180	0.175	0.942
$\widehat{\alpha}_{1s}(\cdot, 55)$	0.048	0.229	0.207	0.926	0.035	0.170	0.153	0.929
$\widehat{\alpha}_{2s}(0.5, \cdot)$	0.087	0.202	0.204	0.937	0.037	0.176	0.156	0.940
$\widehat{\alpha}_{2s}(\cdot, 45)$	0.007	0.210	0.223	0.952	0.015	0.207	0.176	0.933
$\widehat{\alpha}_{2s}(\cdot, 55)$	0.074	0.203	0.207	0.945	0.059	0.163	0.154	0.945
$\widehat{F}_{1s}(\cdot \mid x = 0.5, z = 0)$	0.002	0.002	0.003	0.966	0.010	0.002	0.002	0.930
$\widehat{F}_{1s}(\cdot \mid x = 0.5, z = 1)$	0.008	0.002	0.003	0.968	0.013	0.002	0.002	0.944
$\widehat{F}_{2s}(\cdot \mid x = 0.5, z = 0)$	0.010	0.003	0.003	0.941	0.004	0.002	0.002	0.937
$\widehat{F}_{2s}(\cdot \mid x = 0.5, z = 1)$	0.014	0.003	0.003	0.930	0.013	0.002	0.002	0.923

Table 2. *IAC associated with testing the specified null hypotheses at* $s = 1$ *and* $s' = 2$. *Near 95% nominal coverages show the method performs well in distinguishing when the true model has distinct study parameters (i.e., null hypotheses should be rejected) and when the true model has shared study parameters (i.e., null hypotheses should not be rejected)*

	True model has distinct study parameters	True model has shared study parameters
$H_0 : \beta_{1s1}(t) = \beta_{1s'1}(t)$	0.936	0.940
$H_0 : \beta_{2s1}(t) = \beta_{2s'1}(t)$	0.916	0.940
$H_0 : \omega_s(t) = \omega_{s'}(t)$	0.944	0.944
$H_0 : \alpha_{1s}(0.5, t) = \alpha_{1s'}(0.5, t)$	0.928	0.924
$H_0 : \alpha_{2s}(0.5, t) = \alpha_{2s'}(0.5, t)$	0.948	0.944
$H_0 : F_{1s}(t\|x = 0.5, z = 0) = F_{1s'}(t\|x = 0.5, z = 0)$	0.932	0.932
$H_0 : F_{1s}(t\|x = 0.5, z = 1) = F_{1s'}(t\|x = 0.5, z = 1)$	0.932	0.932
$H_0 : F_{2s}(t\|x = 0.5, z = 0) = F_{2s'}(t\|x = 0.5, z = 0)$	0.944	0.932
$H_0 : F_{2s}(t\|x = 0.5, z = 1) = F_{2s'}(t\|x = 0.5, z = 1)$	0.944	0.932

4. MEGA-ANALYSIS OF HD STUDIES

4.1. *Clinical research problem*

We applied our method to de-identified data from three HD studies: COHORT ($n = 430$), PHAROS ($n = 346$), and PREDICT ($n = 909$) introduced in Section 1. Our analysis focused on individuals at risk for HD ($\geq$36 CAG repeats) who were at least 18 years old at study entry. This restriction only affected COHORT (ages were 15–89 years) but minimally. On average, participants were mostly female ($\geq$60%), had 42 CAG repeats (range 36–50), were 40 years old at baseline (range 18–65 years), and had at least 15 years of education ($\geq$42%). Table S.3 of supplementary material available at *Biostatistics* online reports the distribution summaries of these measures.

We analyzed ages of first cognitive impairment and HD motor-diagnosis as now defined.

Age of first cognitive impairment as determined by a series of cognitive exams. Cognitive performance was evaluated using three neuropsychological tests on the Unified Huntington's Disease Rating Scale (Huntington Study Group, 1996), but we used two as results from all three tests were not provided. The Symbolic Digit Modality Test measured psychomotor speed, attention and working memory via the number of correct responses on a timed task of symbol to digit transcription. The Stroop color naming, Stroop word reading, and Stroop interference tests measured selective attention, cognitive flexibility and processing speed. The three tests involved participants naming colors, reading words, and naming ink colors in a word; the correct number of responses in 45 seconds was recorded. Low scores indicate poor cognitive function.

We defined age of first cognitive impairment as the first time when a subject's test score on any of these exams was 1.5 standard deviations below the respective average score of controls. Control subjects were individuals not at risk for HD from COHORT, PHAROS, and PREDICT.

Age of HD motor-diagnosis as determined by a trained examiner. A motor-diagnosis occurred when the examiner was 99% confident the person's extrapyramidal signs were unequivocally associated with HD. We focused on individuals who were diagnosed after their baseline visit.

Censoring for both clinical events was largely administrative with censoring rates being $\geq$ 61% for cognitive impairment and $\geq$ 81% for HD motor-diagnosis. Key research questions about these clinical events are: (*Q1*) Do the distributions for the time to first cognitive impairment and time to HD motor-diagnosis differ by study? (*Q2*) If the distributions do not differ by study, what do the studies collectively suggest about the likeliness of the events' occurrences? (*Q3*) Is one event more likely to occur before

Table 3. *Robustness of proposed method to different dependencies between covariates and censoring. We report average absolute bias* (*abs bias*)*, empirical variance* (*emp var*)*, estimated variance* (*est var*) *and 95% coverage probabilities.* $\widehat{\gamma}(\cdot)$*,* $\widehat{\beta}(\cdot)$ *denotes results averaged over* $t \in [40, 60]$*;* $\widehat{\alpha}(0.50, \cdot)$ *is results at* $x = 0.50$ *averaged over* $t \in [40, 60]$*, and* $\widehat{\alpha}(\cdot, 45)$*,* $\widehat{\alpha}(\cdot, 55)$ *are results at* $t = 45$*,* $t = 55$*, respectively, averaged over* $x \in [0, 1]$

	Study 1				Study 2			
	abs bias	emp var	est var	95% cov	abs bias	emp var	est var	95% cov
	True model has covariate-dependent censoring. Estimation assumes dependence.							
$\widehat{\beta}_0(\cdot)$	0.033	0.167	0.145	0.933	0.033	0.167	0.145	0.933
$\widehat{\beta}_{1s1}(\cdot)$	0.013	0.046	0.040	0.933	0.018	0.036	0.032	0.938
$\widehat{\beta}_{2s1}(\cdot)$	0.008	0.048	0.040	0.920	0.009	0.038	0.033	0.928
$\widehat{\gamma}_2(\cdot)$	0.009	0.151	0.138	0.936	0.009	0.151	0.138	0.936
$\widehat{\omega}_2(\cdot)$	—	—	—	—	0.040	0.207	0.200	0.941
$\widehat{\alpha}_{1s}(0.5, \cdot)$	0.050	0.191	0.166	0.941	0.043	0.138	0.138	0.952
$\widehat{\alpha}_{1s}(\cdot, 45)$	0.015	0.194	0.183	0.940	0.030	0.155	0.152	0.943
$\widehat{\alpha}_{1s}(\cdot, 55)$	0.071	0.187	0.166	0.930	0.059	0.137	0.137	0.949
$\widehat{\alpha}_{2s}(0.5, \cdot)$	0.097	0.166	0.166	0.939	0.022	0.127	0.138	0.970
$\widehat{\alpha}_{2s}(\cdot, 45)$	0.017	0.185	0.183	0.949	0.009	0.141	0.152	0.959
$\widehat{\alpha}_{2s}(\cdot, 55)$	0.090	0.169	0.166	0.936	0.038	0.125	0.137	0.967
$\widehat{F}_{1s}(\cdot \vert x = 0.5, z = 0)$	0.004	0.002	0.003	0.967	0.010	0.002	0.002	0.935
$\widehat{F}_{1s}(\cdot \vert x = 0.5, z = 1)$	0.006	0.003	0.003	0.956	0.009	0.002	0.002	0.947
$\widehat{F}_{2s}(\cdot \vert x = 0.5, z = 0)$	0.014	0.003	0.003	0.931	0.008	0.002	0.002	0.933
$\widehat{F}_{2s}(\cdot \vert x = 0.5, z = 1)$	0.014	0.003	0.003	0.932	0.009	0.002	0.002	0.926
	True model has covariate-dependent censoring. Estimation assumes independence.							
$\widehat{\beta}_0(\cdot)$	0.049	0.164	0.145	0.935	0.049	0.164	0.145	0.935
$\widehat{\beta}_{1s1}(\cdot)$	0.053	0.039	0.040	0.942	0.008	0.030	0.032	0.959
$\widehat{\beta}_{2s1}(\cdot)$	0.009	0.040	0.040	0.939	0.047	0.031	0.032	0.940
$\widehat{\gamma}_2(\cdot)$	0.007	0.146	0.138	0.937	0.007	0.146	0.138	0.937
$\widehat{\omega}_2(\cdot)$	—	—	—	—	0.038	0.203	0.199	0.940
$\widehat{\alpha}_{1s}(0.5, \cdot)$	0.042	0.190	0.166	0.940	0.050	0.138	0.138	0.950
$\widehat{\alpha}_{1s}(\cdot, 45)$	0.017	0.193	0.183	0.940	0.032	0.154	0.151	0.943
$\widehat{\alpha}_{1s}(\cdot, 55)$	0.056	0.185	0.165	0.930	0.071	0.138	0.136	0.945
$\widehat{\alpha}_{2s}(0.5, \cdot)$	**0.106**	0.165	0.166	0.937	0.017	0.126	0.138	0.969
$\widehat{\alpha}_{2s}(\cdot, 45)$	0.014	0.185	0.183	0.950	0.008	0.141	0.152	0.958
$\widehat{\alpha}_{2s}(\cdot, 55)$	**0.105**	0.168	0.165	0.934	0.026	0.124	0.137	0.968
$\widehat{F}_{1s}(\cdot \vert x = 0.5, z = 0)$	0.006	0.002	0.003	0.969	0.009	0.002	0.002	0.941
$\widehat{F}_{1s}(\cdot \vert x = 0.5, z = 1)$	0.009	0.003	0.003	0.955	0.010	0.002	0.002	0.950
$\widehat{F}_{2s}(\cdot \vert x = 0.5, z = 0)$	0.013	0.003	0.003	0.940	0.007	0.002	0.002	0.939
$\widehat{F}_{2s}(\cdot \vert x = 0.5, z = 1)$	0.015	0.003	0.003	0.936	0.011	0.002	0.002	0.928
	True model has independent censoring. Estimation assumes dependence.							
$\widehat{\beta}_0(\cdot)$	0.030	0.169	0.145	0.932	0.030	0.169	0.145	0.932
$\widehat{\beta}_{1s1}(\cdot)$	0.017	0.045	0.040	0.936	0.009	0.037	0.032	0.937
$\widehat{\beta}_{2s1}(\cdot)$	0.008	0.048	0.040	0.918	0.021	0.039	0.032	0.923
$\widehat{\gamma}_2(\cdot)$	0.009	0.152	0.137	0.932	0.009	0.152	0.137	0.932
$\widehat{\omega}_2(\cdot)$	—	—	—	—	0.034	0.210	0.199	0.939
$\widehat{\alpha}_{1s}(0.5, \cdot)$	0.043	0.193	0.165	0.940	0.049	0.137	0.137	0.952
$\widehat{\alpha}_{1s}(\cdot, 45)$	0.016	0.195	0.183	0.940	0.029	0.153	0.151	0.945
$\widehat{\alpha}_{1s}(\cdot, 55)$	0.058	0.192	0.164	0.927	0.065	0.133	0.136	0.946
$\widehat{\alpha}_{2s}(0.5, \cdot)$	0.080	0.169	0.166	0.941	0.026	0.126	0.138	0.968
$\widehat{\alpha}_{2s}(\cdot, 45)$	0.019	0.186	0.182	0.948	0.008	0.141	0.151	0.959
$\widehat{\alpha}_{2s}(\cdot, 55)$	0.070	0.177	0.164	0.935	0.043	0.123	0.136	0.963
$\widehat{F}_{1s}(\cdot \vert x = 0.5, z = 0)$	0.004	0.002	0.003	0.967	0.011	0.002	0.002	0.934
$\widehat{F}_{1s}(\cdot \vert x = 0.5, z = 1)$	0.006	0.003	0.003	0.954	0.011	0.002	0.002	0.938
$\widehat{F}_{2s}(\cdot \vert x = 0.5, z = 0)$	0.011	0.003	0.003	0.938	0.008	0.002	0.002	0.931
$\widehat{F}_{2s}(\cdot \vert x = 0.5, z = 1)$	0.012	0.003	0.003	0.939	0.011	0.003	0.002	0.914

another by age t? Answering these questions help clinicians better understand the relationship between cognitive impairment and motor-diagnosis and evaluate if the relationship differs by study. The latter is important since study recruitment differed by source (registry in COHORT vs. longitudinal in PHAROS and PREDICT), and participant knowledge of genetic mutation status.

To address these questions, we applied our proposed methodology. We defined T_1 as the age of first cognitive impairment and T_2 as the age of HD motor-diagnosis. In relation to these event times, we evaluated baseline covariates: the number of CAG repeats (X), age at baseline (Z_1), gender (Z_2) with 1 denoting female, and a binary education covariate (Z_3) with 1 denoting at least 15 years of education. We treated age at baseline parametrically similar to earlier analyses of HD progression (Tabrizi *and others*, 2013). We treated education as binary rather than continuous to handle coding differences: COHORT recorded education categorically, PHAROS and PREDICT recorded it in years. Most participants received a higher education, so the cut-off of 15 years of education evenly divided the participants. The dimensions of our covariates were thus $p_x = 1, p_z = 3$. In terms of random effects, we considered $\boldsymbol{R}_{is}(t)$ to be a random intercept.

4.2. *Analysis*

We report results at the following covariate values: 40 and 46 CAG repeats (X); age at baseline (Z_1) is 40 years; gender (Z_2) is female; and education (Z_3) is either $\geq$ 15 years of education or $<$15 years. We chose these values because age of motor-diagnosis and number of CAG repeats are inversely correlated (Langbehn *and others*, 2010), so 40 CAG repeats represent late motor-diagnosis and CAG repeats represent early motor-diagnosis. The average age at baseline was 40, and although most participants were female, results were similar for males and females. We displayed results for both values of the education variable to evaluate its effect on cognition.

4.2.1. *Evaluation of proportional odds model* We first evaluated how well our proposed time-varying proportional odds model fit the data as compared to a time-varying proportional hazards model. Estimates for a proportional hazards model can be derived by replacing the inverse logit link in model (2.1) with the cumulative distribution function for the extreme value distribution. All estimation steps in Section 2.3.3 remain the same except that the DPQL in (2.5) would involve the deviance for the generalized extreme value model. We used this modification to compare our proposed time-varying proportional odds model and a time-varying proportional hazards model. Both models were compared against a non-parametric Kaplan–Meier estimator, so the ideal model will better match the non-parametric estimates.

For this comparison, we only considered gender and CAG repeats as covariates because including others would limit the sample size and deteriorate the Kaplan–Meier estimates. For all HD studies, the proportional odds model overlapped well the Kaplan–Meier as seen in Figure 1. The proportional hazards model did not fit well. When estimating the likeliness of motor diagnosis by age 59 years, the Kaplan–Meier estimated 0.332 (95% CI 0.089, 0.51), the proportional odds model 0.33 (95% CI 0.29, 0.36), and the proportional hazards model 0.29 (95% CI 0.26, 0.31). Thus, a proportional hazards assumption leads to a 4% estimated difference with the Kaplan–Meier compared to less than 1% difference under a proportional odds assumption. The differences are even larger when evaluating the likeliness of a cognitive impairment by age 59 years in which the Kaplan–Meier estimates 0.453 (95% CI 0.186, 0.633), the proportional odds model estimates 0.42 (95% CI 0.39, 0.46) and the proportional hazards model estimates 0.37 (95% CI 0.34, 0.39). The proportional odds model leads to a mere 3% difference with the Kaplan–Meier estimate compared to the 8% difference for the proportional hazards model. Similar observations at other ages indicate that a proportional odds model is a better fit than a proportional hazards model.

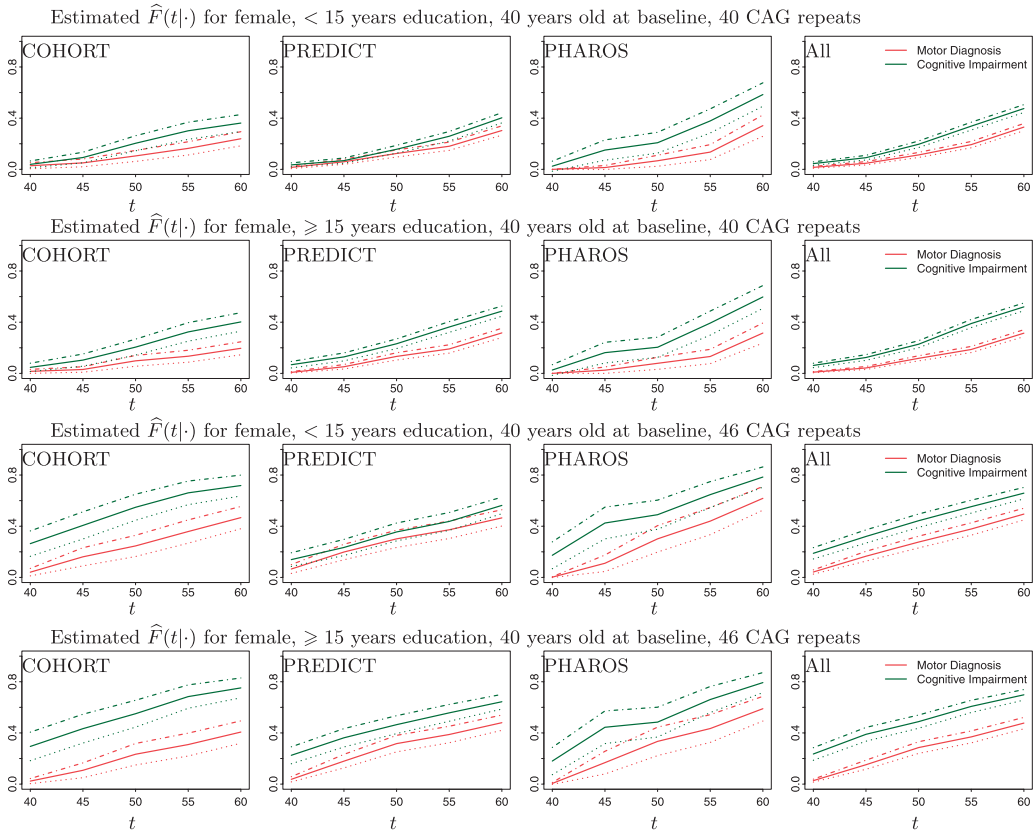

Fig. 2. Estimated marginal distribution functions (solid line) and pointwise 95% confidence band (dashed and dotted lines) for HD motor-diagnosis and first cognitive impairment. Results shown for COHORT, PREDICT, PHAROS from Model A (distinct study parameter model) in first three columns and from Model B (shared study parameter model that combines information from all studies) in the last column (denoted by All). Results for men were similar.

4.2.2. *Evaluation of how distributions for the clinical events differ by study* Having confirmed the appropriateness of our proportional odds model, we then addressed question *(Q1)*. This question is important for model formulation in that we determine if we should assume distinct study parameters (Model A with $p_x = 1, p_z = 3$) or shared ones (Model B with $p_x = 1, p_z = 3$).

Differences in study parameters were assessed using the bootstrap-based joint confidence interval (Section 2.4) applied to estimates from Model A. Overall, no statistically significant differences were noted (results not shown) and the similarity of the marginal distribution estimates for each study is evident in Figure 2 (first three columns). Having similar estimated marginal distributions across all studies reflects the nature of how HD studies are conducted. First, Hogarth *and others* (2005) showed that trained clinicians have a high inter-rater agreement in the assessment of motor manifestations of HD. Given that the HD investigators in all studies received similar training, it is not surprising that the marginal distribution of motor-diagnosis is not study dependent. Second, all cognitive tests had clear quantitative results not influenced by impartial judgments. Thus, because all HD studies used the same quantitative measures for the cognitive exams explains why we observed similar marginal distributions for the cognitive impairment.

4.2.3. *Evaluation of how distributions for the clinical events compare* The observed similarities in the marginal distributions in each study suggest that we could combine the information from COHORT, PHAROS, and PREDICT using the shared study parameter model (Model B with $p_x = 1$ and $p_z = 3$). Results from Model B were used to address questions (*Q2*) and (*Q3*) concerning how the marginal distributions for the clinical events compare.

The far-right column in Figure 2 shows the estimated marginal distributions obtained from Model B. An immediate benefit of combining study information is efficiency gains: confidence intervals in the far-right column in Figure 2 are narrowest. The figure suggests that cognitive impairment has a higher probability of occurring before HD motor-diagnosis for $t \in [40, 60]$. This observed relationship supports results from a simple analysis in which we evaluated the percentages of subjects who experienced motor-diagnosis and cognitive impairment by age t, $t = 40, 45, \ldots, 60$ among those who experienced both. Table S.4 of supplementary material available at *Biostatistics* online shows larger percentages for cognitive impairment than motor-diagnosis. This simple analysis, however, ignores censoring and correlations between the event types which our approach accommodates.

Our results also support the recent finding that cognitive impairments emerge years before a motor diagnosis (Paulsen *and others*, 2014). Compared to these earlier findings which based their conclusions on only one clinical study and evaluated each failure type separately, our approach jointly and comprehensively analyzes three large studies. For comparison, Figure S.1 of supplementary material available at *Biostatistics* online shows the results from Model B compared to Model C in Section 3.1.2 with $p_x = 1$ and $p_z = 3$ which analyzes each study separately. The wider confidence bands in Models C indicate efficiency loss if the studies are analyzed separately. Our proposed model can thus significantly reduce variability compared to the models in Paulsen *and others* (2014).

5. DISCUSSION

We proposed a novel analysis of clustered failure time data based on a time-varying, proportional odds model with random effects. The method handles different correlation structures resulting from the clustering nature of events measured on the same individual and between-study heterogeneity and maintains model flexibility in its non-parametric modeling of covariate effects. In clinical studies with a small number of participants, our method can increase the sample size by meaningfully combining the individual data into a model that adjusts for the study from which the data were drawn. The result leads to improved power and precision.

Applying our method to three HD studies led us to observe that marginal distributions for age of motor-diagnosis is similar across all three studies despite different study recruitment criteria; a similar result holds for the distribution of age of first cognitive impairment. (Results for the estimated functional parameters are in Section S.7 of supplementary material available at *Biostatistics* online.) The similarity implies that the assessment of motor-diagnosis and cognitive impairment is consistent regardless of the clinical study. We did not assess for differences between study sites which can have varied sample sizes. However, we could not perform this analysis since site information was unavailable. Also, our definition of cognitive impairment is based on results from two neurological tests; a future analysis could evaluate each cognitive exam separately to evaluate differences.

We assumed a time-varying proportional odds model as motivated by our HD study. We can modify our method to a proportional hazards model by replacing the inverse logit link in (2.1) with the cumulative distribution function for the extreme value distribution. This is in the family of generalized additive mixed models, so we may use the mgcv package in R with complementary log–log link function for estimation. Lastly, a limitation of our model is that we assume time-varying coefficients, not time-varying covariates. As suggested by a reviewer, a proportional odds model with time-varying covariates and random effects

could be derived from extending the work of Yang and Prentice (1999). This new model will be considered in future work.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

This study used data from the COHORT study which received support from HP Therapeutics, Inc.; PREDICT HD Study which received support from NINDS, and PHAROS Study which received support from the National Human Genome Research Institute and NINDS. We thank the COHORT, PREDICT-HD, and PHAROS investigators and respective coordinators who collected data and/or samples, as well as participants and their families who made this work possible. *Conflict of Interest*: None declared.

Funding

Huntington's Disease Society of America Human Biology Project Fellowship; National Institute of Neurological Disease and Stroke (NINDS; K01NS099343, NS073671, NS082062, 5R01NS40068); National Center for Advancing Translational Sciences (2UL1RR024156-06); CHDI Foundation; Chamber's Family Fund; and Texas A&M School of Public Health Research Enhancement and Development Initiative (REDI-23-202059-36000).

References

Andersen, P. K. and Perme, M. P. (2010). Pseudo-observations in survival analysis. *Statistical Methods in Medical Research* **19**, 71–99.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association* **88**, 9–25.

Cai, L., Cheng, T. and Wei, S. C. (2002). Semiparametric mixed-effects models for clustered failure time data. *Journal of the American Statistical Association* **97**, 514–522.

Chen, Y., Chen, K. and Ying, Z. (2010). Analysis of multivariate failure time data using marginal proportional hazards model. *Statistica Sinica* **20**, 1025–1041.

Clayton, D. and Cuzick J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A* **148**, 82–117.

Crainiceanu, C. M., Staicu, A. M., Ray, S. and Punjabi, N. (2012). Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statistics in Medicine* **31**, 3223–3240.

de Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer.

Dorsey, E. R. and The Huntington Study Group COHORT Investigators. (2012) Characterization of a large group of individuals with Huntington disease and their relatives enrolled in the COHORT Study. *PLOS One* **7**, e29522.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association* **83**, 414–425.

Hogarth, P., Kayson, E., Kieburtz, K., Marder, K., Oakes, D., Rosas, D., Shoulson, I., Wexler, N.S., Young, A.B., Zhao, H. and others. (2005). Interrater agreement in the assessment of motor manifestations of Huntington's disease. *Movement Disorders* **20**, 293–297.

Huntington Study Group. (1996). Unified Huntington's disease rating scale: reliability and consistency. *Movement Disorder* **11**, 136–142.

HUNTINGTON STUDY GROUP PHAROS INVESTIGATORS. (2006). At risk for HD: the Prospective Huntington At Risk Observational Study cohort enrolled. *Archives of Neurology* **63**, 991–996.

LANGBEHN, D. R., HAYDEN, M. R., PAULSEN, J. S. (2010). CAG-repeat length and the age of onset in HD. *American Journal of Medical Genetics* **153**, 397–408.

LEE, Y. AND NELDER, J. (2004). Conditional and marginal models: another view. *Statistical Science* **19**, 219–238.

LOGAN, B., ZHANG, M. AND KLEIN, J. (2011). Marginal models for clustered time to event data with competing risks using pseudo-values. *Biometrics* **67**, 1–7.

MARRA, G. AND RADICE, R. (2010). Penalised regression splines: theory and application to medical research. *Statistical Methods in Medical Research* **19**, 107–125.

PAULSEN, J., LONG, J. D., JOHNSON, H. J., AYLWARD, E. H., and others. (2014). Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: a decade of the PREDICT-HD study. *Frontiers in Aging Neuroscience* **6**, Article 78.

TABRIZI, S. J., SCAHILL, R. I., OWEN, G., and others. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *The Lancet Neurology* **12**, 637–649.

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

YANG, S. AND PRENTICE, R. (1999). Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association* **94**, 125–136.

YING, Z. AND WEI, L. J. (1994). The Kaplan-Meier estimate for dependent failure time observations. *Journal of Multivariate Analysis* **50**, 17–29.

ZENG, D. AND LIN, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society, Series B* **69**, 507–564.

ZENG, D., LIN, D. AND YIN, G. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association* **100**, 470–483.

[*Received July 12, 2016; revised October 31, 2017; accepted for publication November 3, 2017*]