Biostatistics (2018), **0**, 0, *pp.* i–xi
doi:10.1093/biostatistics/BIOSTS-16245-R1-web-supplement

Web Supplementary Materials for "Evaluating longitudinal markers under two-phase study designs"

MARLENA MAZIARZ, TIANXI CAI, LI QI, ANNA LOK & YINGYE ZHENG

APPENDIX

Appendix A. Design-based true sampling weight from two-phase studies

We consider here two-phase studies where all longitudinal samples from individuals selected into the subcohort are measured, therefore $\pi_{ij} \equiv \pi_i$. The form of π_i can be obtained explicitly for both stratified case cohort (sCCH) (Gray, 2009; Liu *and others*, 2012) and nested case-contro (NCC) studies (Samuelsen, 1997; Cai and Zheng, 2011) designs with a discrete stratification/matching variable $\mathbf{Z}$ taking S unique values, $\mathbf{z}_1, ..., \mathbf{z}_S$. Specifically, for sCCH design,

$$\pi_i = \sum_{d=0}^{1} \sum_{s=1}^{S} \frac{n_{ds}\ I(\Delta_i = d, \mathbf{Z}_i = \mathbf{z}_s)}{\sum_{j=1}^{n} I(\Delta_j = d, \mathbf{Z}_j = \mathbf{z}_s)}$$

where n_{ds} is the number of subjects sampled from the set $\{i : \Delta_i = d, \mathbf{Z}_i = \mathbf{z}_s\}$, typically specified by design. For a matched NCC (mNCC) design with m controls matched to each case on the matching variable $\mathbf{Z}$, π_i can be calculated as $\widetilde{\pi}_i = \Delta_i + (1 - \Delta_i)\{1 - \widetilde{G}(\mathbf{W}_i)\}$, with

$$\widetilde{G}(\mathbf{W}_i) = \prod_{j:X_j \leqslant X_i} \left\{1 - \frac{m\Delta_j I(\mathbf{Z}_j = \mathbf{Z}_i)}{\sum_{k=1}^{n} I(X_k \geqslant X_j, \mathbf{Z}_k = \mathbf{Z}_i) - 1}\right\}. \tag{A.1}$$

*To whom correspondence should be addressed.

Appendix B. Asymptotic linear expansion of $\widehat{R}_i(\tau_0|s_{ij})$

We first show that $\widehat{\boldsymbol{\beta}}_{\tau_0}$ in Equation (3) is consistent for $\boldsymbol{\beta}_{\tau_0}$. For simplicity, we drop the subscript of $\widehat{\boldsymbol{\beta}}_{\tau_0}$ and denote the true value of $\boldsymbol{\beta}_{\tau_0}$ by $\boldsymbol{\beta}_0$. Denote

$$\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0) = \mathcal{H}_i(s_{ij})\left\{I(X_i \leqslant s_{ij} + \tau_0) - g\{\boldsymbol{\beta}^{\intercal}\mathcal{H}_i(s_{ij})\}\right\}$$

where $\mathcal{H}_i(s_{ij})$ is a vector of partial longitudinal information collected up to time s_{ij} and may include some flexible functionals of components in $\mathbf{H}_i(s_{ij})$, the observed covariate information up to time s_{ij} for subject i. Assume that $\widehat{\pi}^{\mathbb{S}}_{ij} = P(\xi_{ij} = 1|\mathbf{w}_{ij})$ is modeled correctly by (2) with parameters denoted by $\boldsymbol{\gamma}$, where $\mathbf{w}_{ij} = (X_i, \Delta_i, \mathbb{A}_{ij}^{\intercal})^{\intercal}$ (See section 3.1 of the main paper). The estimator using Nadaraya-Watson weights can be derived under similar conditions and procedures in Appendix A of Qi, Wang and Prentice (2005), and is thus omitted here. The estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be obtained simultaneously be solving estimation equation

$$U(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \\ \\ U^{\gamma}(\boldsymbol{\gamma}) \end{pmatrix} = 0,$$

where

$$U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})}\widehat{w}^{\mathbb{C}}_{ij}(\tau_0)\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0),$$

and

$$U^{\gamma}(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\xi_{ij} - \pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})\}\frac{\partial \pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}.$$

Note that $U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\gamma)}w^{\mathbb{C}}_{ij}(\tau_0)\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\gamma)}(\widehat{w}^{\mathbb{C}}_{ij}(\tau_0) - w^{\mathbb{C}}_{ij}(\tau_0))\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0).$$

The second term is bounded by

$$\max_{i,j}|\widehat{w}^{\mathbb{C}}_{ij}(\tau_0) - w^{\mathbb{C}}_{ij}(\tau_0)| \times \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})}|\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0)|.$$

The empirical sum converges by the law of large numbers and the term $\max_{i,j} |\widehat{w}^{\mathbb{C}}_{ij}(\tau_0) - w^{\mathbb{C}}_{ij}(\tau_0)| \to 0$ in probability since the Kaplan-Meier estimator is uniformly consistent. Thus, it turns out that

$$U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})} w^{\mathbb{C}}_{ij}(\tau_0)\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0) + o_p(1).$$

Take partial derivative of $U^{\beta}(\boldsymbol{\beta}, \gamma)$ with respect to $\boldsymbol{\beta}$, we have

$$\begin{aligned}\left.\frac{\partial U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})} \widehat{w}^{\mathbb{C}}_{ij}(\tau_0)\mathcal{H}_i(s_{ij})\mathcal{H}_i(s_{ij})^T \dot{g}\{\boldsymbol{\beta}_0^{\mathsf{T}}\mathcal{H}_i(s_{ij})\} \\ &\xrightarrow{\mathcal{P}} \sum_{j=1}^{m_i} E\left[w^{\mathbb{C}}_{ij}(\tau_0) w^{\mathbb{S}}_{ij} \mathcal{H}_i(s_{ij})\mathcal{H}_i(s_{ij})^T \dot{g}\{\boldsymbol{\beta}_0^{\mathsf{T}}\mathcal{H}_i(s_{ij})\}\right] \equiv I^{\beta}\end{aligned}$$

uniformly in a neighborhood of $\boldsymbol{\beta}_0$. Clearly, $\frac{\partial U^{\gamma}(\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} = 0$. It can be shown that $\frac{\partial U^{\gamma}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$ converges to its limit I^{γ} and $\frac{\partial U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$ converges to its limit $I^{\beta\gamma}$ and denote

$$I = \begin{pmatrix} I^{\beta} & I^{\beta\gamma} \\ 0 & I^{\gamma} \end{pmatrix}.$$

Because $det(I^{\beta}) \neq 0$, $det(I^{\gamma}) \neq 0$, $det(I^{\beta\gamma}) \neq 0$, and $det(I^{\gamma\beta}) = 0$, I is invertible. Following the same argument as in the proof of Theorem 2 of Xu *and others* (2009), we have the existence and consistency of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}})$.

Now we show the asymptotic expansion of $n^{1/2}U^{\beta}(\boldsymbol{\beta}, \boldsymbol{\gamma})$. It can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})} w^{\mathbb{C}}_{ij}(\tau_0)\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})} (\widehat{w}^{\mathbb{C}}_{ij}(\tau_0) - w^{\mathbb{C}}_{ij}(\tau_0))\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0).$$

Following from the asymptotic expansion of Kaplan-Meier estimator,

$$n^{\frac{1}{2}}\left\{G(t)/\widehat{G}(t) - 1\right\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\mathcal{U}_{Gi}(t) + o_p(1),$$

we have

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{R_i} \frac{\xi_{ij}}{\pi^{\mathbb{S}}_{ij}(\boldsymbol{\gamma})} (\widehat{w}^{\mathbb{C}}_{ij}(\tau_0) - w^{\mathbb{C}}_{ij}(\tau_0))\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0) = n^{-1/2}\sum_{i=1}^{n}\int \mathcal{U}_{Gi}(u)d\Phi(u) \tag{B.1}$$

where $\mathcal{U}_{Gi}(t) = \int_0^t P(X_i > u)^{-1}[I(X_i \leqslant u)(1 - \Delta_i) + I(X_i \geqslant u)d\log\{G(u)\}]$, and

$$\Phi(u) = \sum_{j}^{m_i} E\{I(X_i \wedge s_{ij} + \tau_0 \leqslant u) w^{\mathbb{C}}_{ij}(\tau_0)\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0)\}. \tag{B.2}$$

Hence,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^{n} \zeta_i(\tau_0|s),$$

where

$$\zeta_i(\tau_0|s) = \mathcal{T} I^{-1} \left(\begin{array}{c} \sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})} w_{ij}^{\mathbb{C}}(\tau_0)\mathcal{R}_i(\boldsymbol{\beta}, s_{ij}, \tau_0) + \int \mathcal{U}_{Gi}(u) d\Phi(u) \\ \sum_{j=1}^{m_i} \{\xi_{ij} - \pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})\} \frac{\partial \pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \end{array} \right),$$

and $\mathcal{T} = \{\mathcal{T}_{jk}\}$ be a $p_1 \times (p_1 + p_2)$ matrix with elements $\mathcal{I}_{jk} = 1$ for $j = 1, \ldots, p_1$, $k = j$, and $\mathcal{T}_{jk} = 0$ otherwise, where p_1 is the length of $\boldsymbol{\beta}$ and p_2 is the length of $\boldsymbol{\gamma}$. Then $\widehat{\mathcal{U}}_{\mathcal{R}}(\tau_0 \mid s) = \sqrt{n}\left[\widehat{R}_i(\tau_0|s) - R_i(\tau_0|s)\right]$ is asymptotically equivalent to a sum of i terms, $n^{-1/2}\sum_{i=1}^{n}\{\zeta_i(\tau_0|s)\frac{\partial R_i(\tau_0|s)}{\partial \boldsymbol{\beta}}\} \equiv n^{-1/2}\sum_{i=1}^{n} \zeta_{Ri}(\tau_0|s)$.

Appendix C. Asymptotic linear expansion of longitudinal summary measures $\mathcal{A}_{s,\tau_0}$

Let $\mathcal{U}_{\mathcal{A}_{s,\tau_0}} = \sqrt{n}(\widehat{\mathcal{A}}_{s,\tau_0} - \mathcal{A}_{s,\tau_0})$. As an example, we consider $\mathcal{A}_{s,\tau_0}$ in the case of $\text{TPF}_{s,\tau_0}(\psi)$. Below we use $\text{TPF}_{s,\tau_0}(\psi)$ to denote a simple setting where $R_i(\tau_0 \mid s_{ij})$ is known with parameters estimated with another cohort, i.e.,

$$\widehat{\text{TPF}}_{s,\tau_0}(\psi) = \frac{\sum_{ij} \text{I}(|s_{ij} - s| \leqslant \epsilon)\text{I}(R_i(\tau_0 \mid s_{ij}) > \psi)\, \text{I}(s < X_i \leqslant s + \tau_0)\, \widehat{w}_{ij}^{\mathbb{C}}(\tau_0)\widehat{w}_{ij}^{\mathbb{S}}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} \text{I}(s < X_i \leqslant s + \tau_0)\, \widehat{w}_{ij}^{\mathbb{C}}(\tau_0)\widehat{w}_{ij}^{\mathbb{S}}}.$$

Let $\mathcal{U}_{\text{TPF}_{s,\tau_0}}(\psi) = n^{1/2}\{\widehat{\text{TPF}}_{s,\tau_0}(\psi) - \text{TPF}_{s,\tau_0}(\psi)\}$, we have

$$\mathcal{U}_{\text{TPF}_{s,\tau_0}}(\psi) = n^{-1/2} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\hat{\boldsymbol{\gamma}})} \widehat{w}_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\text{TPF},ij}(\psi), \tag{C.1}$$

where $\zeta_{\text{TPF},ij}(\psi) = \frac{I(s<X_i\leqslant s+\tau_0)}{1-P(T>s+\tau_0)}\{\text{I}(R_i(\tau_0 \mid s_{ij}) > \psi)\text{I}(|s_{ij} - s| \leqslant \epsilon) - \text{TPF}_{s,\tau_0}(\psi)\}$. Further, C.1 is equivalent to

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})} \widehat{w}_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\text{TPF},ij}(\psi) \\ & + n^{-1/2} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \left\{ \frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\hat{\boldsymbol{\gamma}})} - \frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})} \right\} \widehat{w}_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\text{TPF},ij}(\psi) \end{aligned} \tag{C.2}$$

Following the same argument in Appendix B, the first term of (C.2) can be written as

$$n^{-1/2} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m_i} \frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})} w_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\text{TPF},ij}(\psi) + \int \mathcal{U}_{Gi}(u) d\Phi_{\mathcal{A}}(u) \right\} \equiv n^{-1/2} \sum_{i=1}^{n} \eta_{1,\mathcal{A}i}(\tau_0|s),$$

where $\Phi_{\mathcal{A}}(u) = E\{\sum_j^{m_i} I(X_i \wedge s_{ij} + \tau_0 \leqslant u) w_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\mathrm{TPF},ij}(\psi)\}$, and the second term of (C.2) can be written as

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})^2}\left\{\pi_{ij}^{\mathbb{S}}(\hat{\boldsymbol{\gamma}}) - \pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})\right\} w_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\mathrm{TPF},ij}(\psi)$$
$$+\, n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi_{ij}}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})^2}\left\{\pi_{ij}^{\mathbb{S}}(\hat{\boldsymbol{\gamma}}) - \pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})\right\}\{\widehat{w}_{ij}^{\mathbb{C}}(\tau_0) - w_{ij}^{\mathbb{C}}(\tau_0)\}\zeta_{\mathrm{TPF},ij}(\psi)$$

By the Taylor expansion and by expanding the form of $\pi_{ij}^{\mathbb{S}}$, the first term above is asymptotically equivalent to

$$n^{-1/2} I^{\gamma}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\xi_{ij} - \pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})\}\frac{\partial \pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} E\left\{\sum_{j=1}^{m_i}\frac{1}{\pi_{ij}^{\mathbb{S}}(\boldsymbol{\gamma})} w_{ij}^{\mathbb{C}}(\tau_0)\zeta_{\mathrm{TPF},ij}(\psi)\right\} \equiv n^{-1/2}\sum_{i=1}^{n}\eta_{2,\mathcal{A}i}(\tau_0|s)$$

and the second term is negligible. Therefore, we have

$$\mathcal{U}_{\mathrm{TPF}_{s,\tau_0}} = n^{-1/2}\sum_{i=1}^{n}(\eta_{1,\mathcal{A}i} + \eta_{2,\mathcal{A}i}) \equiv n^{-1/2}\sum_{i}\eta_{\mathcal{A}i}(\tau_0|s).$$

Appendix D. Nested case-control sample of the HALT-C clinical trial

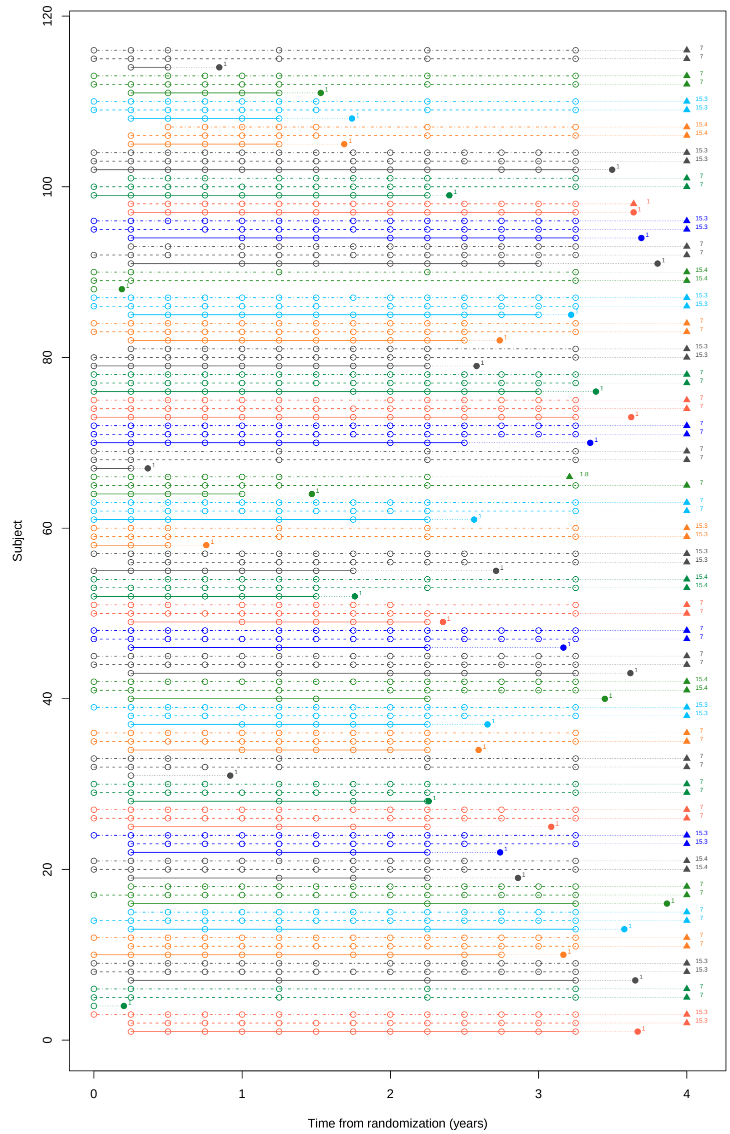

Fig. 1. An overview of the nested case-control sample of the HALT-C clinical trial. The visits where DCP marker was measured are shown as empty circles, the event times are shown as filled circles, and non-events (censoring times) as filled triangles. The subjects are grouped according to their matching in the nested case-control study, with subjects with the same color belonging to the same risk set. The estimated nested case-control inverse probability sampling weights are shown to the right of each event time for the cases, and censoring time for the non-events.

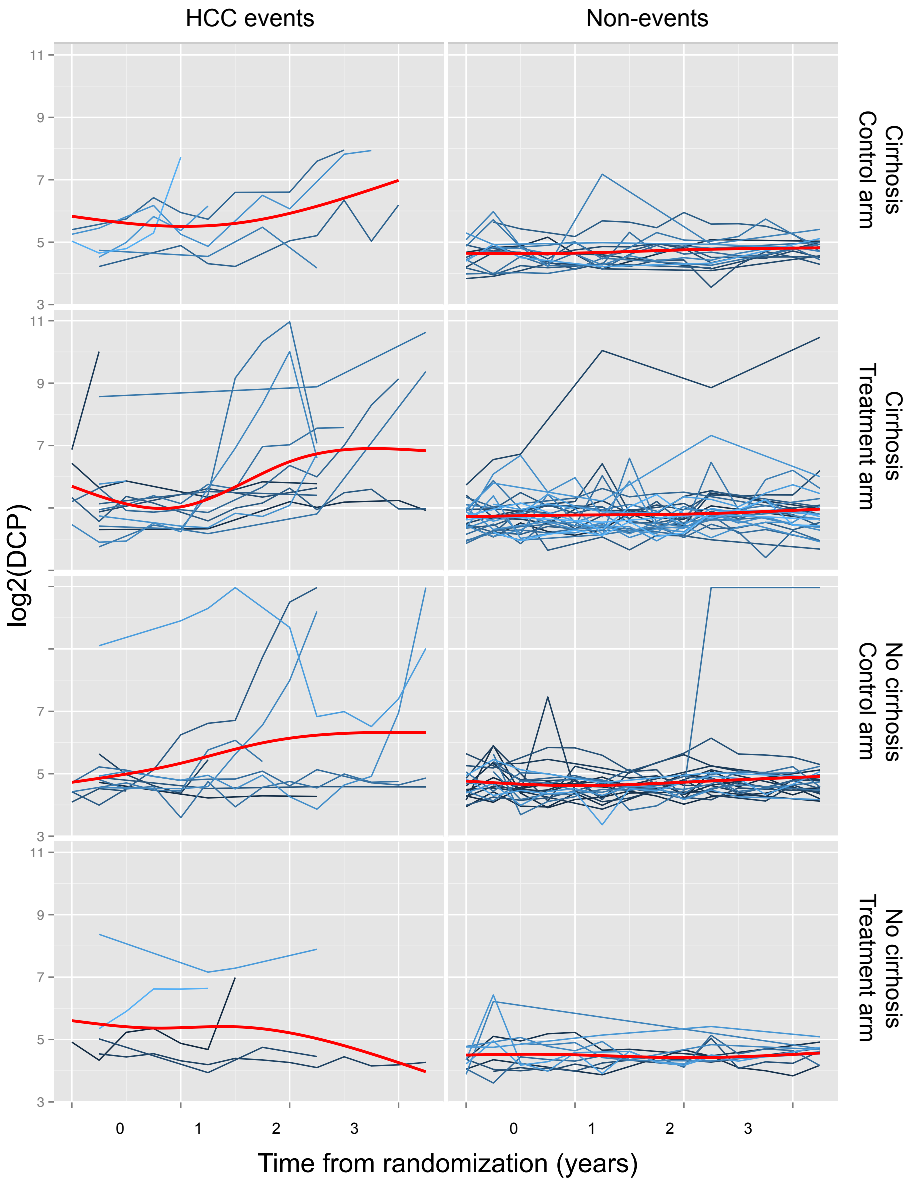

Fig. 2. The des-γ-carboxyprothrombin (DCP) marker observations in nested case-control subset of the HALT-C clinical trial, stratified by cirrhosis and treatment assignment. The DCP values were truncated at 2000 units and $\log_2$ transformed. The thick red lines indicate the conditional linear fit with time modeled as a spline with 3 degrees of freedom.

Appendix E. Results of analysis of example simulated datasets

A tutorial where we show examples on how to obtain risk predictions based on longitudinal marker in a case-cohort or a nested case-control sample, using methods described the manuscript available at `http://rpubs.com/marlenamaziarz/risk-prediction-in-longitudinal-two-phase-studies`. The code for estimating the weights in CCH and NCC samples, obtaining risk predictions and evaluating those predictions, as well as example datasets for CCH and NCC are available in `https://github.com/marlenamaziarz/longitudinal-two-phase`. Specifically, the examples in the tutorial illustrate the use of our methods to predict risk using PC_{GLM} based on a longitudinal biomarker with a survival outcome, as well as all evaluation measures of prediction and their standard error discussed in the paper. Namely, the evaluation measures we considered are the prediction error (PE), true and false positive fractions (TPF and FPF), the area under the Receiver Operating Characteristics curve (AUC), the proportion of cases followed (PCF) and the proportion of cases needed to be followed (PNF). Below are results of the analysis using example (simulated) datasets.

Nested case-control

Table 1. Results of the nested case-control (NCC) analysis for different combinations of conditioning time (s) and prediction time (t = s + τ) (s from 6 to 36 in increments of 6, and τ = 6). Prediction timeframe = τ, Est = estimate, SE = standard error, n.case = number of cases in the sample at time s, n.ctrl = number of controls at s, and n.cens = number of subjects censored up to time s. SE was estimated based on 500 perturbations. PE = prediction error, TPF/FPF = true/false positive fraction, AUC = the area under the Receiver Operating Characteristics curve, PCF = the proportion of cases followed, and PNF = the proportion of cases needed to be followed. A tutorial on how to run this analysis can be found on Rpubs, the code and example datasets are on GitHub.

Measure	s	t	τ	Est	SE	n_{case}	n_{ctrl}	n_{cens}
PE	6	12	6	0.03	0.00	53	512	8
TPF(0.4)	6	12	6	0.13	0.05	53	512	8
FPF(0.3)	6	12	6	0.01	0.00	53	512	8
AUC	6	12	6	0.79	0.03	53	512	8
PCF(0.2)	6	12	6	0.55	0.07	53	512	8
PNF(0.8)	6	12	6	0.44	0.07	53	512	8
PE	12	18	6	0.06	0.01	74	423	15
TPF(0.4)	12	18	6	0.24	0.06	74	423	15
FPF(0.3)	12	18	6	0.04	0.01	74	423	15
AUC	12	18	6	0.83	0.02	74	423	15
PCF(0.2)	12	18	6	0.61	0.06	74	423	15
PNF(0.8)	12	18	6	0.35	0.06	74	423	15
PE	18	24	6	0.10	0.01	89	310	24
TPF(0.4)	18	24	6	0.24	0.05	89	310	24
FPF(0.3)	18	24	6	0.09	0.02	89	310	24
AUC	18	24	6	0.80	0.03	89	310	24
PCF(0.2)	18	24	6	0.56	0.05	89	310	24
PNF(0.8)	18	24	6	0.40	0.06	89	310	24
PE	24	30	6	0.11	0.01	57	226	27
TPF(0.4)	24	30	6	0.22	0.06	57	226	27
FPF(0.3)	24	30	6	0.12	0.02	57	226	27
AUC	24	30	6	0.78	0.03	57	226	27
PCF(0.2)	24	30	6	0.47	0.07	57	226	27
PNF(0.8)	24	30	6	0.46	0.05	57	226	27
PE	30	36	6	0.12	0.01	50	145	31
TPF(0.4)	30	36	6	0.31	0.07	50	145	31
FPF(0.3)	30	36	6	0.12	0.03	50	145	31
AUC	30	36	6	0.81	0.03	50	145	31
PCF(0.2)	30	36	6	0.47	0.06	50	145	31
PNF(0.8)	30	36	6	0.41	0.05	50	145	31
PE	36	42	6	0.11	0.02	19	104	22
TPF(0.4)	36	42	6	0.37	0.11	19	104	22
FPF(0.3)	36	42	6	0.20	0.04	19	104	22
AUC	36	42	6	0.72	0.06	19	104	22
PCF(0.2)	36	42	6	0.42	0.11	19	104	22
PNF(0.8)	36	42	6	0.53	0.07	19	104	22

Table 2. Results of the case-cohort (CCH) analysis for different combinations of conditioning time (s) and prediction time (t = s + τ) (s from 6 to 36 in increments of 6, and τ = 6). Prediction timeframe = τ, Est = estimate, SE = standard error, n.case = number of cases in the sample at time s, n.ctrl = number of controls at s, and n.cens = number of subjects censored up to time s. SE was estimated based on 500 perturbations. PE = prediction error, TPF/FPF = true/false positive fraction, AUC = the area under the Receiver Operating Characteristics curve, PCF = the proportion of cases followed, and PNF = the proportion of cases needed to be followed. A tutorial on how to run this analysis can be found on Rpubs, the code and example datasets are on GitHub

Measure	s	t	τ	Est	SE	n_{case}	n_{ctrl}	n_{cens}
PE	6	12	6	0.04	0.00	74	639	124
TPF(0.4)	6	12	6	0.05	0.04	74	639	124
FPF(0.3)	6	12	6	0.01	0.00	74	639	124
AUC	6	12	6	0.81	0.02	74	639	124
PCF(0.2)	6	12	6	0.61	0.06	74	639	124
PNF(0.8)	6	12	6	0.41	0.05	74	639	124
PE	12	18	6	0.06	0.01	89	475	75
TPF(0.4)	12	18	6	0.14	0.06	89	475	75
FPF(0.3)	12	18	6	0.02	0.01	89	475	75
AUC	12	18	6	0.83	0.02	89	475	75
PCF(0.2)	12	18	6	0.65	0.06	89	475	75
PNF(0.8)	12	18	6	0.36	0.07	89	475	75
PE	18	24	6	0.10	0.01	104	326	45
TPF(0.4)	18	24	6	0.20	0.06	104	326	45
FPF(0.3)	18	24	6	0.05	0.01	104	326	45
AUC	18	24	6	0.77	0.03	104	326	45
PCF(0.2)	18	24	6	0.45	0.05	104	326	45
PNF(0.8)	18	24	6	0.42	0.06	104	326	45
PE	24	30	6	0.10	0.01	85	208	33
TPF(0.4)	24	30	6	0.28	0.07	85	208	33
FPF(0.3)	24	30	6	0.09	0.02	85	208	33
AUC	24	30	6	0.83	0.03	85	208	33
PCF(0.2)	24	30	6	0.58	0.06	85	208	33
PNF(0.8)	24	30	6	0.33	0.07	85	208	33
PE	30	36	6	0.11	0.01	54	128	26
TPF(0.4)	30	36	6	0.21	0.07	54	128	26
FPF(0.3)	30	36	6	0.09	0.03	54	128	26
AUC	30	36	6	0.76	0.04	54	128	26
PCF(0.2)	30	36	6	0.53	0.07	54	128	26
PNF(0.8)	30	36	6	0.56	0.07	54	128	26
PE	36	42	6	0.10	0.02	26	84	18
TPF(0.4)	36	42	6	0.22	0.10	26	84	18
FPF(0.3)	36	42	6	0.09	0.04	26	84	18
AUC	36	42	6	0.70	0.07	26	84	18
PCF(0.2)	36	42	6	0.53	0.11	26	84	18
PNF(0.8)	36	42	6	0.69	0.16	26	84	18

References

Cai, T. and Zheng, Y. (2011). Nonparametric evaluation of biomarker accuracy under nested case-control studies. *J. Am. Stat. Assoc. Journal of the American Statistical Association* **106**(494), 569–580.

Gray, Robert J. (2009). Weighted analyses for cohort sampling designs. *Lifetime Data Analysis* **15**(1), 24–40.

Liu, Dandan, Cai, Tianxi and Zheng, Yingye. (2012). Evaluating the predictive value of biomarkers with stratified case-cohort design. *Biometrics* **68**(4), 1219–1227.

Samuelsen, S. O. (1997). A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* **84**(2), 379–394.

Xu, Qiang, Paik, Myunghee Cho, Luo, Xiaodong and Tsai, Wei-Yann. (2009). Reweighting estimators for cox regression with missing covariates. *Journal of the American Statistical Association* **104**(487), 1155–1167.