Web-based Supplementary Materials for "Hazard Ratio Estimation for Biomarker Calibrated Dietary Exposures"

PAMELA A. SHAW

Biostatistics Research Branch,

National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA

email: shawpa@niaid.nih.gov

ROSS L. PRENTICE

Division of Public Health Sciences,

Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

rprentic@whi.org

1 Riskset Regression Calibration Parameter Estimates

This section contains further details regarding the calculation of the risk set regression calibration estimator in equation (3) of the manuscript. The risk set regression calibration estimate $\widehat{X}(t)$ of X is an estimate which is recalculated at each failure time and uses data only from individuals in the risk set $\mathcal{R}(t)$ for failure time t.

At a given time t, define

$$\widehat{X}_i(t) = \begin{cases} \widehat{E}\{X_i | Y_i(t) = 1, W_{i\cdot}, Q_{i\cdot}, Z_i\} & \text{if } R_i = 1 \\ \widehat{E}\{X_i | Y_i(t) = 1, Q_{i\cdot}, Z_i\} & \text{if } R_i = 0 \end{cases}$$

where $W_{i\cdot} = \kappa^{-1} \sum_{j=1}^{\kappa} W_{ij}$, $Q_{i\cdot} = k^{-1} \sum_{j=1}^{k} Q_{ij}$, and '$\wedge$' denotes estimate. For individuals with $Y_i(t) = 1$, denote the mean and covariance for $(X_i, W_{i\cdot}, Q_{i\cdot}, Z_i)$ as $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$. Define $\boldsymbol{\Sigma}_{22}(t) = \text{Cov}(W_{i\cdot}, Q_{i\cdot}, Z_i | Y_i(t) = 1)$ and let $\widetilde{\boldsymbol{\Sigma}}_{22}(t)$ be the matrix $\boldsymbol{\Sigma}_{22}$ with the first row and column deleted. Similarly define $\boldsymbol{\Sigma}_{12}(t) = [\Sigma_{WX}(t), \Sigma_{QX}(t), \Sigma_{ZX}(t)]$ and $\widetilde{\boldsymbol{\Sigma}}_{12}(t) = [\Sigma_{QX}(t), \Sigma_{ZX}(t)]$. Then one has

$$\widehat{X}_i(t) = \begin{cases} \widehat{\mu}_X(t) + \widehat{\widetilde{\boldsymbol{\Sigma}}}_{12}(t) \widehat{\widetilde{\boldsymbol{\Sigma}}}_{22}^{-1}(t)(\boldsymbol{\Upsilon}_i - \widehat{\boldsymbol{\mu}}_{\boldsymbol{\Upsilon}_i}) & \text{if } R_i = 1 \\ \widehat{\mu}_X(t) + \widehat{\widetilde{\boldsymbol{\Sigma}}}_{12}(t) \widehat{\widetilde{\boldsymbol{\Sigma}}}_{22}^{-1}(t)(\boldsymbol{\Upsilon}_i - \widehat{\boldsymbol{\mu}}_{\boldsymbol{\Upsilon}_i}) & \text{if } R_i = 0 \end{cases} \tag{1.1}$$

where

$$\boldsymbol{\Upsilon}_i = \begin{cases} (W_{i\cdot}, Q_{i\cdot}, Z_i)' & \text{if } R_i = 1 \\ (Q_{i\cdot}, Z_i)' & \text{if } R_i = 0, \end{cases}$$

and $\boldsymbol{\mu}_{\boldsymbol{\Upsilon}_i}(t)$ denotes the mean of $\boldsymbol{\Upsilon}_i$ at time t. For notational simplicity suppose each of X, W, Q, and Z are scalar, then the necessary moment estimates for equation (1.1) are:

$$
\begin{aligned}
\widehat{\Sigma}_\epsilon &= \frac{1}{\sum_{i=1}^n R_i} \sum_{i=1}^n \frac{1}{k-1} \sum_{j=1}^k R_i (W_{ij} - W_{i\cdot})^2 \\
\widehat{\Sigma}_W(t) &= \frac{1}{\sum_{i=1}^n R_i Y_i(t) - 1} \sum_{i=1}^n R_i Y_i(t) \{W_{i\cdot} - \widehat{\mu}_X(t)\}^2 \\
\widehat{\Sigma}_X(t) &= \widehat{\Sigma}_W(t) - \frac{\widehat{\Sigma}_\epsilon}{k} = \widehat{\Sigma}_{XW}(t) \\
\widehat{\Sigma}_{QX}(t) &= \frac{1}{\sum_{i=1}^n R_i Y_i(t) - 1} \sum_{i=1}^n R_i Y_i(t) \{Q_{i\cdot} - \widehat{\mu}_Q(t)\}\{W_{i\cdot} - \widehat{\mu}_X(t)\} = \widehat{\Sigma}_{QW}(t) \\
\widehat{\Sigma}_{ZX}(t) &= \frac{1}{\sum_{i=1}^n R_i Y_i(t) - 1} \sum_{i=1}^n R_i Y_i(t) \{Z_i - \hat{\mu}_Z(t)\}\{W_{i\cdot} - \widehat{\mu}_X(t)\} = \widehat{\Sigma}_{ZW}(t) \\
\widehat{\Sigma}_Q(t) &= \frac{1}{\sum_{i=1}^n Y_i(t) - 1} \sum_{i=1}^n Y_i(t) \{Q_{i\cdot} - \widehat{\mu}_Q(t)\}^2 \\
\widehat{\Sigma}_Z(t) &= \frac{1}{\sum_{i=1}^n Y_i(t) - 1} \sum_{i=1}^n Y_i(t) \{Z_i - \widehat{\mu}_Z(t)\}^2 \\
\widehat{\Sigma}_{QZ}(t) &= \frac{1}{\sum_{i=1}^n Y_i(t) - 1} \sum_{i=1}^n Y_i(t) \{Q_{i\cdot} - \widehat{\mu}_Q(t)\}\{Z_i - \hat{\mu}_Z(t)\}, \\
\text{where} \quad \hat{\mu}_X(t) &= \frac{1}{\sum_{i=1}^n R_i Y_i(t)} \sum_{i=1}^n R_i Y_i(t) W_{i\cdot} = \widehat{\mu}_W(t) \\
\hat{\mu}_Q(t) &= \frac{1}{\sum_{i=1}^n Y_i(t)} \sum_{i=1}^n Y_i(t) Q_{i\cdot} \\
\hat{\mu}_Z(t) &= \frac{1}{\sum_{i=1}^n Y_i(t)} \sum_{i=1}^n Y_i(t) Z_i
\end{aligned}
$$

Note that only subjects in the reliability subset contribute to the estimates of $\mu_X(t)$, $\Sigma_X(t)$, $\Sigma_W(t)$, $\Sigma_{ZX}(t)$, and $\Sigma_{QX}(t)$. The ordinary regression calibration estimate can be found by replacing $Y_i(t)$ by 1 in all of the moment estimators used for $\widehat{X}_i(t)$. This produces a common estimate $\widehat{X}_i$ for all t. For categorical Z, one can condition on Z to provide linear approximations of $E(X|W, Q, Z = z)$ and $E(X|Q, Z = z)$ for each level of Z. For speed of computation of the resulting RRC estimator of β, one may wish to approximate the RRC estimator of $\widehat{X}$ by not recalibrating at every failure time, but say according to a chosen scheme of equally spaced failure times in the range of observed failure times. This approach may be particularly useful for a large number of failure times.

2 Regularity Assumptions

We assume the general error model of Section 2.1 in the manuscript, along with its stated assumptions of independence and iid random variables, and that the random variables in the error and survival moment have finite first and second moments. We also assume the Cox model described in Section 2 of the manuscript holds. Sufficient conditions for asymptotic normal distributions for the regression parameter estimators, and for the validity of the corresponding bootstrap variance estimators, are as follows.

2.1 Regression Calibration Estimator

A. $(N_i, Y_i, X_i, \boldsymbol{Z}_i, R_i, \gamma_i, k_i, \kappa_i, \epsilon_{i1}, \ldots, \epsilon_{ik_i}, \xi_{i1}, \ldots, \xi_{i\kappa_i})$ are iid random vectors for $i = 1, \ldots, n$.

B1. $P(k > 1) > 0$ and $P(\kappa > 1) > 0$.

B2. $P(R = 1) > 0$ and $P(\boldsymbol{V} = \boldsymbol{v}) > 0$, for all $\boldsymbol{v} \in \{\boldsymbol{v} | \boldsymbol{V} = \boldsymbol{v}\}$, where $\boldsymbol{V}$ is the vector of categorical components of $\boldsymbol{Z}$.

C. Since the partial likelihood score equation is concave, with imposed regularity conditions on the nuisance parameters (condition G), a unique solution to the regression calibration equation will exist, namely $\beta^\star$. For $\boldsymbol{\theta}_0 = (\beta^\star, \boldsymbol{\phi}_0)$, where $\boldsymbol{\phi}_0$ is the true nuisance parameter vector, $\exists$ a compact neighborhood $\mathscr{N}(\boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0 = (\beta^\star, \boldsymbol{\phi}_0)$, such that:

$$\mathrm{E} \sup_{\boldsymbol{\theta} \in \mathscr{N}(\boldsymbol{\theta}_0), t \in [0,M]} |\widehat{X}(t; \widehat{\boldsymbol{\phi}})|^2 \, e^{\beta \widehat{X}(t; \widehat{\boldsymbol{\phi}})} < \infty.$$

D. There exists a finite constant $M > 0$ such that $\mathrm{P}(U \geq M) > 0$, where $U = T \wedge C$.

E. $\int_0^M \lambda_0(u) du < \infty$, for $0 < M < \infty$.

F. The derivative of the estimating equation $U_n(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$ exists and is continuous and bounded for $\boldsymbol{\theta} \in \mathscr{N}(\boldsymbol{\theta}_0)$, a compact neighborhood of $\boldsymbol{\theta}_0$. Furthermore, $\frac{\partial}{\partial \boldsymbol{\theta}} U_n(\boldsymbol{\theta})$ converges to its limit uniformly in $\mathscr{N}(\boldsymbol{\theta}_0)$ and this limit $A \equiv \lim_{n \to \infty} \frac{\partial}{\partial \boldsymbol{\theta}} U_n(\boldsymbol{\theta})$ is nonsingular at $\boldsymbol{\theta}_0$; where $\boldsymbol{\theta}_0$ is defined as in condition C.

G. The error model in Section 2.1 of the manuscript holds and for the error nuisance parameter vector $\boldsymbol{\phi}_0$, there exists vector valued function $\boldsymbol{\Psi}(\boldsymbol{\phi}) = \boldsymbol{\Psi}(W, Q, Z, R, k, \kappa, \boldsymbol{\phi})$ such that

$$\begin{aligned} \mathrm{E}\{\boldsymbol{\Psi}(\boldsymbol{\phi}_0)\} &= 0 \\ \mathrm{E}\{\boldsymbol{\Psi}(\boldsymbol{\phi}_0)\boldsymbol{\Psi}^t(\boldsymbol{\phi}_0)\} &< \infty \\ \sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\Psi}_i(\boldsymbol{\phi}_0) + o_p(1), \end{aligned}$$

that is, $\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ is asymptotically equivalent to a sum of iid components and hence, asymptotically normal.

Assumptions A and $C-E$ are similar to those made for the ordinary Cox model (Andersen and Gill, 1982; Tsiatis, 1981). Assumptions $B1$ - $B2$ list conditions for the numbers of replicates for the main and biomarker instruments. Note assumption $B1$ is made for the regression calibration estimator proposed in equation (3) of the manuscript, also described above in Section 1; however, a regression calibration estimator is possible without replicates of either Q or W. As noted by Neuhouser et al. (2008), since $E[X|Q,Z] = E[W|Q,Z]$, one can simply regress W on (Q, Z) to obtain an appropriate $\widehat{X}$ for regression calibration. $P(\boldsymbol{V} = \boldsymbol{v}) > 0$ is important for regularity of stratum specific estimates. Assumption F ensures uniform convergence of the estimating equation and thus, since the estimating equation for β is a complicated function of the nuisance parameters, that certain combinations of these parameters are avoided. In particular, there needs to be nonzero correlation between Q and X, and the covariance matrix for (X, W, Q, Z) needs to be nonsingular. Assumption G ensures there exists an asymptotically consistent and normal estimator, $\widehat{\boldsymbol{\phi}}$, for $\boldsymbol{\phi}_0$, as $\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ can be written as i.i.d. components asymptotically. Moment estimators are one possible choice for $\boldsymbol{\Psi}$. Wang et al. (1997) provides a more general class of $\sqrt{n}$-consistent estimators satisfying these conditions.

2.2 Conditional Score

The regularity conditions for the conditional score equation are the same as conditions $A-G$ in Section 2.1, except that condition C, specific to terms in the estimating equation for $\boldsymbol{\theta}$, is replaced by C'.

C'. $\exists \mathscr{N}(\boldsymbol{\theta}_0)$, a compact neighborhood around the true value of the parameter $\boldsymbol{\theta}_0 = (\beta_0, \boldsymbol{\phi}_0)$ such that: $\quad \mathrm{E}\{\sup_{\boldsymbol{\theta} \in \mathscr{N}(\boldsymbol{\theta}_0), t \in [0,M]} \zeta^2(\boldsymbol{\theta}, t) E_0^2(\boldsymbol{\theta}, t)\} < \infty$.

Assumption C' implies that the coefficient of X in the measurement error model, $\delta_1 + \delta_3 Z$, is bounded away from zero almost everywhere in $\mathscr{N}(\boldsymbol{\theta}_0)$. This is a reasonable requirement for Q, as $\delta_1 + \delta_3 Z = 0$ implies that Q does not depend on the true covariate X. Note the conditional score estimating equation (equation (5) of the manuscript) was developed assuming a discrete, scaler variable Z in the general measurement error model equation (1). This method can be extended to a more general vector $\boldsymbol{Z}$, so long as the scale bias in the general measurement model (defined by δ_3) is only dependent on the components of $\boldsymbol{Z}$ that are discrete.

2.3 Nonparametric Corrected Score

The regularity conditions for the nonparametric score equation are the same as conditions $A-G$ in Section 2.1, except that condition C, specific to terms in the estimating equation for $\boldsymbol{\theta}$, is replaced by C''.

C''. $\exists \mathscr{N}(\boldsymbol{\theta}_0)$, a compact neighborhood around the true value of the parameter $\boldsymbol{\theta}_0 = (\beta_0, \boldsymbol{\delta}_0)$ such that:
$\mathrm{E}[\sup_{\boldsymbol{\theta} \in \mathscr{N}(\boldsymbol{\theta}_0)} \widetilde{X}_z^2(\boldsymbol{\delta}) \exp\{2\beta \widetilde{X}_z(\boldsymbol{\delta})\}] < \infty$, for all $z \in Z$.

Assumption C'', like assumption C' above, implies that $\delta_1 + \delta_3 Z$ is bounded away from zero almost everywhere in $\mathscr{N}(\boldsymbol{\theta}_0)$. Note the nonparametric corrected score estimating equation

(equation (6) of the manuscript) was developed assuming a discrete, scaler variable Z in the general measurement error model equation (1). This method can be extended to a more general vector $\boldsymbol{Z}$, so long as the scale bias in the general measurement model (defined by δ_3) is only dependent on the components of $\boldsymbol{Z}$ that are discrete.

3 Derivation of conditional intensity

This section includes a heuristic argument for the derivation of the conditional intensity used in the conditional score equation presented in Section 2.3 of the manuscript. This argument follows similarly to one provided by Tsiatis and Davidian (2001) in their development of a conditional score estimator for the proportional hazards setting where the covariate of interest is observed with classical measurement error, i.e. with independent mean zero random error.

First consider an individual not in the biomarker subset. Because error and random effect terms for Q_{ij} are assumed to be normally distributed, one can condition on the observed $Q_{i\cdot} = \frac{1}{k}\sum_{j=1}^{k} Q_{ij}$ in place of $Q_{i1}, Q_{i2}, \ldots, Q_{ik}$. Assuming the general measurement error model, one has $Q_{i\cdot}|(X_i, Z_i) \sim N(\delta_0 + \delta_1 X_i + \delta_2 Z_i + \delta_3 X_i Z_i, \Sigma_{e_i})$, where $\Sigma_{e_i} = \Sigma_{\gamma_i} + \Sigma_{\xi/k}$. At time u, the conditional density for $\{dN_i(u), Q_{i\cdot}\}$ given individual i is at risk and the time-independent covariates X_i and Z_i is:

$$\begin{aligned}
p(dN_i(u),&Q_{i\cdot}|Y_i(u) = 1, X_i, Z_i)\\
&= p(dN_i(u)|Q_{i\cdot}, Y_i(u) = 1, X_i, Z_i) \times p(Q_{i\cdot}|Y_i(u) = 1, X_i, Z_i)\\
&= \{\lambda_0(u)du \exp(\beta_1 X_i + \beta_2 Z_i)\}^{dN_i(u)}\{1 - \lambda_0(u)du \exp(\beta_1 X_i + \beta_2 Z_i)\}^{1-dN_i(u)}\\
&\qquad \times (2\pi\Sigma_{e_i})^{-\frac{1}{2}} \exp\{-\frac{1}{2\Sigma_{e_i}}(Q_{i\cdot} - \delta_0 - \delta_1 X_i - \delta_2 Z_i - \delta_3 X_i Z_i)^2\}\\
&= \{\lambda_0(u)du \exp(\beta_1 X_i + \beta_2 Z_i)\}^{dN_i(u)}(2\pi\Sigma_{e_i})^{-\frac{1}{2}}\\
&\qquad \times \exp\{-\frac{1}{2\Sigma_{e_i}}(Q_{i\cdot} - \delta_0 - \delta_1 X_i - \delta_2 Z_i - \delta_3 X_i Z_i)^2\} + o_p(du)\\
&= \{\lambda_0(u)du \exp(\beta_2 Z_i)\}^{dN_i(u)}(2\pi\Sigma_{e_i})^{-\frac{1}{2}}\\
&\qquad \times \exp[-\frac{1}{2\Sigma_{e_i}}\{(Q_{i\cdot} - \delta_0 - \delta_2 Z_i)^2 + (\delta_1 + \delta_3 Z_i)^2 X_i^2\}]\\
&\qquad \times \exp\left[\frac{X_i}{\Sigma_{e_i}}\{\beta_1 \Sigma_{e_i} dN_i(u) + (\delta_1 + \delta_3 Z_i)(Q_{i\cdot} - \delta_0 - \delta_2 Z_i)\}\right] + o_p(du)
\end{aligned}$$

The sufficient statistic for X_i can be defined as:

$$\zeta_i = \beta_1 \Sigma_{e_i} dN_i(u) + (\delta_1 + \delta_3 Z_i)(Q_{i\cdot} - \delta_0 - \delta_2 Z_i).$$

Making a change of variables $Q_{i\cdot} \mapsto \zeta_i$, one has:

$$Q_{i\cdot} = (\delta_1 + \delta_3 Z_i)^{-1}\{\zeta_i - \beta_1 \Sigma_{e_i} dN_i(u)\} + (\delta_0 + \delta_2 Z_i)$$

Then

$$
\begin{aligned}
\mathrm{P}(dN_i(u), \zeta_i | Y_i(u) = 1, X_i, Z_i) &= \{\lambda_0(u) du \exp(\beta_2 Z_i)\}^{dN_i(u)} (2\pi\Sigma_{e_i})^{-\frac{1}{2}} (\delta_1 + \delta_3 Z_i)^{-1} \\
&\quad \times \exp\left[-\frac{1}{2\Sigma_{e_i}} \left\{ (\delta_1 + \delta_3 Z_i)^{-1} (\zeta_i - \beta_1 \Sigma_{e_i} dN(u)) \right\}^2 + (\delta_1 + \delta_3 Z_i)^2 X_i^2 \right] \\
&\quad \times \exp\left(\frac{X_i \zeta_i}{\Sigma_{e_i}} \right) + o_p(du) \\
&= \{\lambda_0(u)\, du \exp(\beta_2 Z_i)\}^{dN_i(u)} (2\pi\Sigma_{e_i})^{-\frac{1}{2}} (\delta_1 + \delta_3 Z_i)^{-1} \\
&\quad \times \exp\left\{ -(\delta_1 + \delta_3 Z_i)^{-2} \frac{\zeta_i^2 - 2\beta_1 \Sigma_{e_i} dN_i(u) \zeta_i + \beta_1^2 \Sigma_{e_i}^2 dN_i(u)}{2\Sigma_{e_i}} \right\} \\
&\quad \times \exp\left(\frac{-(\delta_1 + \delta_3 Z_i)^2 X_i^2}{2\Sigma_{e_i}} + \frac{X_i \zeta_i}{\Sigma_{e_i}} \right) + o_p(du) \\
&= \{\lambda_0(u)\, du \exp(\beta_2 Z_i)\}^{dN_i(u)} K(X_i, Z_i, \zeta_i) \\
&\quad \exp\left[(\delta_1 + \delta_3 Z_i)^{-2} \left\{ \beta_1 dN_i(u) \zeta_i - \frac{\Sigma_{e_i} \beta_1^2}{2} dN_i(u) \right\} \right] + o_p(du)
\end{aligned}
$$

The conditional probability $\mathrm{P}(dN_i(u) = 1 | \zeta_i = c, X_i, Z_i, Y_i(u) = 1)$ is

$$
\frac{\mathrm{P}(dN_i(u) = 1, \zeta_i = c | X_i, Z_i, Y_i(u) = 1)}{\mathrm{P}(dN_i(u) = 0, \zeta_i = c | X_i, Z_i, Y_i(u) = 1) + \mathrm{P}(dN_i(u) = 1, \zeta_i = c | X_i, Z_i, Y_i(u) = 1)}.
$$

The terms without $dN_i(u)$, denoted by $K(X_i, Z_i, c)$, will cancel from the numerator and denominator. Up to order du the numerator is

$$
\lambda_0(u) du \exp(\beta_2 Z_i) K(X_i, Z_i, c) \exp\left\{ \beta_1 (\delta_1 + \delta_3 Z_i)^{-2} c - \frac{\Sigma_{e_i} \beta_1^2}{2} (\delta_1 + \delta_3 Z_i)^{-2} \right\}.
$$

Up to order 1 the denominator is

$$
\mathrm{P}(dN_i(u) = 0, \zeta_i = c | X_i, Z_i, Y_i(u) = 1) = K(X_i, Z_i, c).
$$

Thus one has

$$
\begin{aligned}
&\mathrm{P}(dN_i(u) = 1 | \zeta_i = c, X_i, Z_i, Y_i(u) = 1) \\
&\quad = \lambda_0(u)\, du \exp(\beta_2 Z_i) \\
&\qquad \times \frac{K(X_i, Z_i, c) \exp\left\{ \beta_1 (\delta_1 + \delta_3 Z_i)^{-2} c - \frac{\Sigma_{e_i} \beta_1^2}{2} (\delta_1 + \delta_3 Z_i)^{-2} \right\}}{K(X_i, Z_i, c)} + o_p(du) \\
&\quad = \lambda_0(u) du \exp\left\{ \beta_1 (\delta_1 + \delta_3 Z_i)^{-2} c - \frac{\Sigma_{e_i} \beta_1^2}{2} (\delta_1 + \delta_3 Z_i)^{-2} + \beta_2 Z_i \right\} + o_p(du).
\end{aligned}
$$

From (3), one has the conditional intensity process

$$
\begin{aligned}
&\lim_{du \to 0} du^{-1} \mathrm{P}\{dN_i(u) = 1 | \zeta_i, X_i, Z_i, Y_i(u)\} \\
&\qquad = \lambda_0(u) \exp\left\{ \frac{\beta_1 \zeta_i - \frac{\beta_1^2 \Sigma_{e_i}}{2}}{(\delta_1 + \delta_3 Z_i)^2} + \beta_2 Z_i \right\} Y_i(u).
\end{aligned}
$$

A similar procedure is followed for the reference subset. In this case the sufficient statistic is

$$\zeta_i = \beta_1 \Sigma_{e_i} \Sigma_\epsilon dN_i(u) + \Sigma_\epsilon\{(\delta_1 + \delta_3 Z_i)(Q_{i\cdot} - \delta_0 - \delta_2 Z_i)\} + \Sigma_{e_i} W_{i\cdot}$$

$P(dN_i(u)|\zeta_i, Y_i(u), X_i, Z_i)$ will be derived from $\mathrm{P}(dN_i(u), Q_{i\cdot}, W_{i\cdot}|Y_i(u), X_i, Z_i)$. Rewrite

$$W_{i\cdot} \;=\; \frac{1}{\Sigma_{e_i}}\{\zeta_i - \beta_1\Sigma_{e_i}\Sigma_\epsilon dN_i(u) - \Sigma_\epsilon(\delta_1+\delta_3 Z_i)(Q_{i\cdot} - \delta_0 - \delta_2 Z_i)\}$$

Then $\mathrm{P}(dN_i(u), \zeta_i, Q_i|Y_i(u), X_i, Z_i)$

$$\begin{aligned}
&= \{\lambda_0(u)du\exp(\beta_2 Z_i)\}^{dN_i(u)}\,(2\pi\Sigma_\epsilon)^{-\frac{1}{2}}(2\pi\Sigma_{e_i})^{-\frac{1}{2}}\frac{1}{\Sigma_{e_i}}\exp\left(\frac{X_i\zeta_i}{\Sigma_\epsilon\Sigma_{e_i}}\right)\\
&\quad\times\exp\left[-\frac{1}{2\Sigma_{e_i}}\{\tilde{Q}_{i\cdot}^2 + (\delta_1+\delta_3 Z_i)^2 X_i^2\} - \frac{1}{2\Sigma_\epsilon}X_i^2\right]\\
&\quad\times\exp\left[-\frac{1}{2\Sigma_\epsilon}\frac{1}{\Sigma_{e_i}^2}\{\widetilde{\zeta_i} - \Sigma_\epsilon(\delta_1+\delta_3 Z_i)\tilde{Q}_{i\cdot}\}^2\right] + o_p(du),
\end{aligned}$$

where $\widetilde{\zeta_i} = \zeta_i - \beta_1\Sigma_{e_i}\Sigma_\epsilon dN_i(u)$ and $\tilde{Q}_{i\cdot} = Q_{i\cdot} - \delta_0 - \delta_2 Z_i$.

One can now complete the square for $Q_{i\cdot}$ and integrate out $Q_{i\cdot}$. Then

$\mathrm{P}(dN_i(u),\zeta_i|Y_i(u) = 1, X_i, Z_i)$

$$\begin{aligned}
&= \{\lambda_0(u)du\exp(\beta_2 Z_i)\}^{dN_i(u)}(2\pi\Sigma_\epsilon)^{-\frac{1}{2}}(2\pi\Sigma_{e_i})^{-\frac{1}{2}}\frac{1}{\Sigma_\epsilon}a(Z_i)\\
&\times\exp\left\{\frac{X_i\zeta_i}{\Sigma_\epsilon\Sigma_{e_i}} - \frac{1}{2\Sigma_\epsilon}X_i^2 - \frac{(\delta_1+\delta_3 Z_i)^2}{2\Sigma_{e_i}}X_i^2\right\}\\
&\times\exp\left[\frac{-\{\zeta_i - \beta_1\Sigma_{e_i}\Sigma_\epsilon dN_i(u)\}^2}{2\Sigma_\epsilon\Sigma_{e_i}^2} + \frac{\{\zeta_i - \beta_1\Sigma_{e_i}\Sigma_\epsilon dN_i(u)\}^2(\delta_1+\delta_3 Z_i)^2}{2\Sigma_{e_i}^2\{\Sigma_{e_i} + \Sigma_\epsilon(\delta_1+\delta_3 Z_i)^2\}}\right] + o_p(du)\\
&= \exp\left[-\frac{\{\zeta_i - \beta_1\Sigma_{e_i}\Sigma_\epsilon dN_i(u)\}^2}{2\Sigma_{e_i}\Sigma_\epsilon\{\Sigma_{e_i} + \Sigma_\epsilon(\delta_1+\delta_3 Z_i)^2\}}\right]\widetilde{K}(X_i, Z_i, \zeta_i) + o_p(du)\\
&\equiv g(\zeta_i, dN_i(u), Z_i)\widetilde{K}(X_i, Z_i, \zeta_i) + o_p(du)
\end{aligned}$$

$\mathrm{P}(dN_i(u) = 1|\zeta_i = c, Z_i = v, Y_i(u) = 1)$

$$\begin{aligned}
&= \frac{g(c,1,v) + o_p(du)}{g(c,0,v) + g(c,1,v)} = \frac{g(c,1,v) + o_p(du)}{g(c,0,v) + o_p(1)} = \frac{g(c,1,v)}{g(c,0,v)} + o_p(du)\\
&= \lambda_0(u)du\,\exp(\beta_2 Z_i)\\
&\quad\times\exp\left[-\frac{(c - \beta_1\Sigma_{e_i}\Sigma_\epsilon)^2}{2\Sigma_{e_i}\Sigma_\epsilon\{\Sigma_{e_i} + \Sigma_\epsilon(\delta_1+\delta_3 Z_i)^2\}} + \frac{c^2}{2\Sigma_{e_i}\Sigma_\epsilon\{\Sigma_{e_i} + \Sigma_\epsilon(\delta_1+\delta_3 Z_i)^2\}}\right] + o_p(du)\\
&= \lambda_0(u)du\,\exp\left\{\frac{\beta_1 c - \beta_1^2\Sigma_{e_i}\Sigma_\epsilon/2}{\Sigma_{e_i} + \Sigma_\epsilon(\delta_1+\delta_3 Z_i)^2} + \beta_2 Z_i\right\} + o_p(du)
\end{aligned}$$

The conditional intensity process is thus

$$\lim_{du\to 0} du^{-1}\mathrm{P}(dN_i(u)=1|\zeta_i,X_i,Z_i,Y_i(u)=1)$$
$$=\lambda_0(u)\exp\left\{\frac{\beta_1\zeta_i-\beta_1^2\Sigma_{e_i}\Sigma_\epsilon/2}{\Sigma_{e_i}+\Sigma_\epsilon(\delta_1+\delta_3 Z_i)^2}+\beta_2 Z_i\right\}Y_i(u).$$

4 Conditional and Corrected Score Nuisance Parameter Estimation

The conditional score and corrected score models require a plug-in estimate for $\boldsymbol{\delta}=(\delta_0,\delta_1,\delta_2,\delta_3)$ from equation (1) in the main text. This estimate can be found by solving the following system of linear equations for $\boldsymbol{\delta}$.

$$\begin{aligned}
\mathrm{E}Q &= \delta_0+\delta_1\mu_X+\delta_2\mu_Z+\delta_3\mu_{XZ}\\
\mathrm{Cov}(W,Q) &= \delta_1\mathrm{Var}(X)+\delta_2\mathrm{Cov}(X,Z)+\delta_3\mathrm{Cov}(X,XZ)\\
\mathrm{Cov}(Z,Q) &= \delta_1\mathrm{Cov}(X,Z)+\delta_2\mathrm{Var}(Z)+\delta_3\mathrm{Cov}(Z,XZ)\\
\mathrm{Cov}(WZ,Q) &= \delta_1\mathrm{Cov}(X,XZ)+\delta_2\mathrm{Cov}(Z,XZ)+\delta_3\mathrm{Var}(XZ).
\end{aligned}$$

Thus,

$$\widehat{\boldsymbol{\delta}}=\begin{bmatrix}1 & \widehat{\mu}_X & \widehat{\mu}_Z & \widehat{\mu}_{XZ}\\ 0 & \widehat{\Sigma}_X & \widehat{\Sigma}_{XZ} & \widehat{\Sigma}_{X,XZ}\\ 0 & \widehat{\Sigma}_{XZ} & \widehat{\Sigma}_Z & \widehat{\Sigma}_{Z,XZ}\\ 0 & \widehat{\Sigma}_{X,XZ} & \widehat{\Sigma}_{Z,XZ} & \widehat{\Sigma}_{XZ}\end{bmatrix}^{-1}\begin{bmatrix}\widehat{\mu}_Q\\ \widehat{\Sigma}_{WQ}\\ \widehat{\Sigma}_{ZQ}\\ \widehat{\Sigma}_{WZ,Q}\end{bmatrix}$$

The moment parameters on the right hand side of the above equation can be estimated from the data using the usual moment estimators. The conditional score estimator also requires estimates for the measurement error variances from equations (1) and (2), i.e. $\mathrm{Var}(\epsilon)$, $\mathrm{Var}(\gamma|Z=z)=a\exp(bz)$, and $\mathrm{Var}(\xi)$. An estimate for Σ_ϵ is provided in Section 1 above. The parameters (a,b) can be estimated by creating a system of equations from the following relationship: $\mathrm{Var}(Q|Z=z)=\mathrm{Var}(\delta_0+\delta_1X+\delta_2Z+\delta_3XZ+\gamma+\xi|Z=z)$.
For example, for binary Z one has:

$$\begin{aligned}
a &= \mathrm{Var}(Q|Z=0)-\delta_1^2\mathrm{Var}(X|Z=0)-\mathrm{Var}(\xi)\quad\text{and}\\
ae^b &= \mathrm{Var}(Q|Z=1)-(\delta_1+\delta_3)^2\mathrm{Var}(X|Z=1)-\mathrm{Var}(\xi).
\end{aligned}$$

Parameter estimates for a and b are thus found by plugging in the estimates for $\widehat{\boldsymbol{\delta}}$ from above, $\widehat{\Sigma}_\xi=n^{-1}\sum_{i=1}^n(\kappa-1)^{-1}\sum_{j=1}^{\kappa}(Q_{ij}-Q_{i\cdot})^2$, and the other moments are the usual moment estimators conditioned on the appropriate level of Z.

5 Cumulative Baseline Hazard Estimation

5.1 Consistency

The consistency of the proposed estimator $\widehat{\Lambda}$ in equation (7) can be seen from the following argument. We use the Huang and Wang (2000) representation of the Breslow estimator as a functional of two empirical process, $\widehat{\mathcal{E}}\{\Delta I(U \leq \cdot)\}$ and $\widehat{\mathcal{E}}\{\exp(bX)I(U \geq \cdot)\}$, where $\widehat{\mathcal{E}}$ denotes the sample mean. Thus, the Breslow estimator for the Cox model with covariate X can be rewritten as

$$\widehat{\Lambda}_0(t;\widehat{\beta}) = \int_0^t \frac{dN(u)}{\sum_{i=1}^n Y_j(u)\exp(\widehat{\beta}X_j)} = \int_0^t \frac{d\widehat{\mathcal{E}}\{\Delta I(U \leq u)\}}{\widehat{\mathcal{E}}\{\exp(\widehat{\beta}X)I(U \geq u)\}}.$$

Under the Cox model with parameter β, this estimator is consistent for

$$\Lambda_0(t;\beta) = \int_0^t \frac{d\mathcal{E}\{\Delta I(U \leq u)\}}{\mathcal{E}\{\exp(\beta X)I(U \geq u)\}}.$$

Recall the random variable derived from the error model,

$$\widetilde{X}_i(\boldsymbol{\delta}) = \frac{Q_{i\cdot} - \delta_0 - \delta_2 Z_i}{\delta_1 + \delta_3 Z_i},$$

where $Q_{i\cdot}$ is the average of k_i replicates of Q_{ij}. At the true nuisance parameter value $\boldsymbol{\delta}^0 = (\delta_{00}, \delta_{10}, \delta_{20}, \delta_{30})$, $\widetilde{X}_i(\boldsymbol{\delta}^0)$ is composed of X_i plus an error term that depends on the value of Z_i. That is

$$\widetilde{X}_i(\boldsymbol{\delta}^0) = X_i + \frac{\gamma_i + \xi_{i\cdot}}{\delta_{10} + \delta_{30} Z_i} = X_i + \nu_i,$$

For simplicity, assume there are $k_i = 2$ replicates of Q for all members of the cohort and $\kappa_i = 2$ replicates of W for all members of the biomarker subset. Recall R_i is the biomarker subset membership indicator for subject i. Here, we again borrow notation from Huang and Wang (2000) and denote the two independent replicates of W as $W^{(1)}$ and $W^{(2)}$. We stratify on values of Z, because the error in $\widetilde{X}_i$ depends on values of Z_i. At $\boldsymbol{\delta}^0$ and $Z_i = z$, the proposed estimator for the cumulative baseline hazard function (7) can be written as

$$\begin{aligned}&\widehat{\Lambda}_{z0}^{NP}(t;b;\boldsymbol{\delta}^0) = \\ &(\widehat{\mathcal{E}}[I(Z=z)R\exp\{b(W^{(1)} - W^{(2)})/2\}])^{-1} \\ &\times \widehat{\mathcal{E}}\left(I(Z=z)R\exp[b\{\widetilde{X}(\boldsymbol{\delta}_0) - (W^{(1)} + W^{(2)})/2\}]\right) \int_0^t \frac{d\widehat{\mathcal{E}}\{I(Z=z)\Delta I(U \leq u)\}}{\widehat{\mathcal{E}}\{I(Z=z)\exp\{b\widetilde{X}(\boldsymbol{\delta}_0)\}I(U \geq u)\}}.\end{aligned} \tag{5.1}$$

Assuming the error terms ϵ_{ij} in W are symmetric and independent of all other random variables in the error and survival model, and assuming the biomarker subset is a random subset of the cohort, the first term on the r.h.s. is consistent for $[\mathcal{E}\{\exp(b\epsilon/2)\}]^{-2}$. Using the same assumptions, and that ν given a fixed Z is independent of the other random variables, the second term on the r.h.s. of (5.1) has the limit $\mathcal{E}\{I(Z=z)\exp(b\nu)\}[\mathcal{E}\{\exp(b\epsilon/2)\}]^2$.

Similarly, the third term on the r.h.s of (5.1) is consistent for

$$\int_0^t \frac{d\mathcal{E}\{I(Z=z)\Delta I(U \le u)\}}{\mathcal{E}\{I(Z=z)\exp(bX+b\nu)I(U \ge u)\}} = \frac{1}{\mathcal{E}\{I(Z=z)\exp(b\nu)\}} \int_0^t \frac{d\mathcal{E}\{I(Z=z)\Delta I(U \le u)\}}{\mathcal{E}\{I(Z=z)\exp(bX)I(U \ge u)\}}.$$

Thus $\widehat{\Lambda}_{z0}^{NP}(t;b;\boldsymbol{\delta}_0)$ is consistent for

$$\Lambda_{z0}(t;b) \quad = \quad \int_0^t \frac{d\mathcal{E}\{I(Z=z)\Delta I(U \le u)\}}{\mathcal{E}\{I(Z=z)\exp(bX)I(U \ge u)\}}.$$

It follows, that for consistent estimators of $\widehat{\beta}$ and $\widehat{\boldsymbol{\delta}}^0$, with some imposed regularity, one has $\widehat{\Lambda}_{z0}^{NP}(t;\widehat{\beta};\widehat{\boldsymbol{\delta}}^0)$ will be consistent for the stratum specific $\Lambda_{0z}(t;\beta)$. For the RC and RRC estimators, this also leads to the convenient overall estimator of Λ_0,

$$\widehat{\Lambda}_0(t) = \sum_{z\in\{Z\}} \int_0^t \exp(-\widehat{\beta}_2 z) n_z(u) n(u)^{-1} \widehat{\Lambda}_{z0}^{NP}(du;\widehat{\beta}_1,\widehat{\boldsymbol{\delta}}),$$

where $n_z(u)$ and $n(u)$ denote the risk set size in stratum $Z=z$ and the overall risk set size, respectively, at time u.

5.2 Asymptotic Normality of $\widehat{\Lambda}_{z0}^{NP}(t;\widehat{\beta};\widehat{\boldsymbol{\delta}}^0)$

In Appendix C of Huang and Wang (2000), the authors sketch out a proof for asymptotic normality for a similar estimate of $\Lambda_0(t;\beta)$. Their estimator, like ours, involves the Breslow estimator functional applied to an error prone random variable, multiplied by a correction factor. The crux of the proof, which can be applied here, relies on regularity conditions which ensure the uniform convergence of the empirical processes in the estimating equation for β and the estimator for $\Lambda_{0z}(t)$ (5.1). The estimating equation for β and the estimator for $\Lambda_{z0}(t)$ are continuous and differentiable functionals of empirical processes. With the imposed regularity, the functional delta method can then be applied to show the asymptotic normality of $\widehat{\Lambda}_{z0}^{NP}(t;\widehat{\beta};\widehat{\boldsymbol{\delta}}^0)$.

6 Distributions for the data in the WHI example

Figure 1 shows the distributions for log-energy consumption data from the food frequency questionnaire in the main WHI cohort and the biomarker measure in the WHI Nutritional Biomarker Subset, from the example (Section 5) of the manuscript, along with the corresponding normal density curve with the same mean and standard deviation.

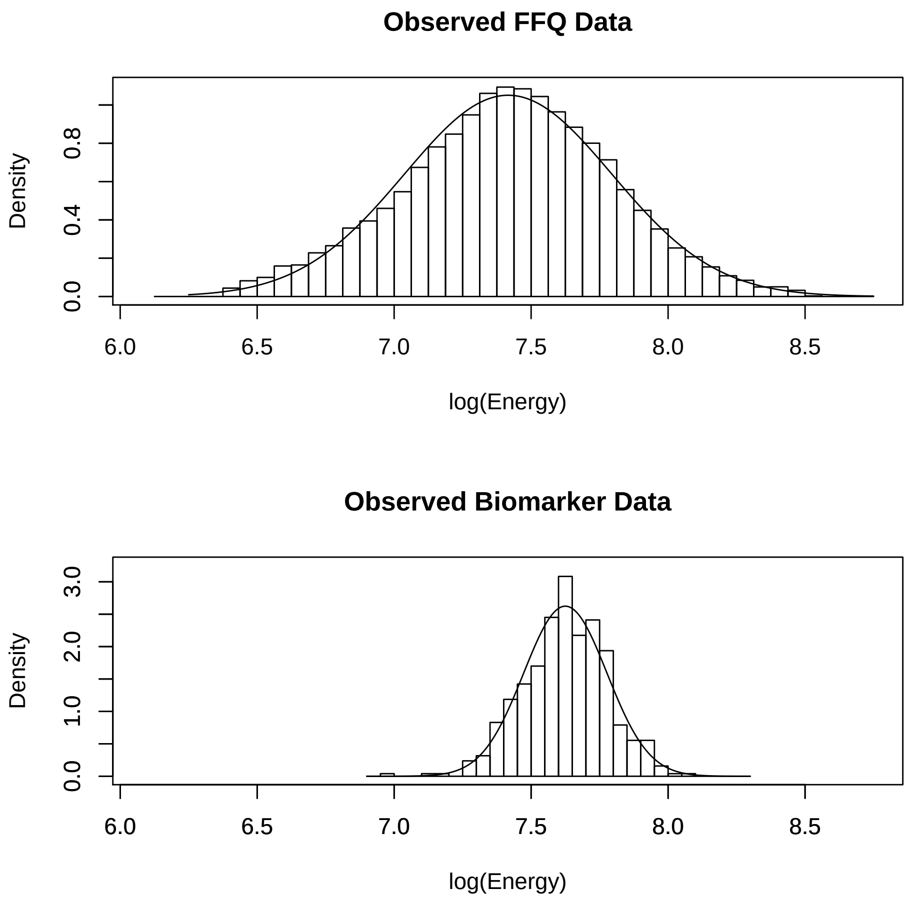

Figure 1: The logarithm of the self-reported (Q) and biomarker (W) measures for energy intake, along with the fitted normal density curves.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics* **10,** 1100–1120.

Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric correction approach. *Journal of the American Statistical Association* **95,** 1209–1219.

Neuhouser, M. L., Tinker, L., Shaw, P. A., Schoeller, D., Bingham, S. A., Van Horn, L., Beresford, S. A. A., Caan, B., Thompson, C., Satterfield, S., Kuller, L., Heiss, G., Smit, E., Sarto, G., Ockene, J., Stefanick, M. L., Assaf, A., Runswick, S., and Prentice, R. L. (2008). Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *American Journal of Epidemiology* **167,** 1247–1259.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics* **9,** 93–108.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88,** 447–458.

Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53,** 131–145.