PREDICT: Prognostic Models for Warfarin, Clopidogrel, or Statin Exposure

Yaping Shi, MS and Jonathan Schildcrout, PhD

March 8, 2011

Contents

List of Tables

1 Introduction

This document describes the models that will be used to identify those who are at highest risk for being exposed to Warfarin, Clopidogrel, or a Statin, and who are therefore candidates for preemptive genotyping. These models are based on Cox proportional hazards (PH) regression analysis and a broader goal here is to develop a modeling and implementation strategy that can be translated into real-time practice without requiring specialized statistical software (e.g., R). The output from these models is risk (and associated uncertainty) of being put on one of these three medications within one, two, three, four, and five years following the date the patient's "medical home" was established.

2 Methods

Two models were developed. Both models included the pre-specified, baseline risk factors age, gender, race, history of type II diabetes, history of coronary disease, history of atrial fibrillation, history of hypertension, history of atherosclerosis, history of CHF, previous clot, and history of dialysis. When height was available, BMI was included in the model (model 1) and when it was absent, weight was used (model 2). Since continuous variable effects on patient outcomes are often non-linear, age, BMI, and weight, were entered into the regression models using flexible restricted cubic spline functionals.

Because the Cox model is semi-parametric, translating model results into non-statistical languages required a multi-step strategy. This strategy involved estimates of 1) the baseline hazard function, 2) the linear predictor function (covariate effects) for the log hazard, 3) standard errors of the log survival function, and a couple of formula that combine 1), 2), and 3) into risk and confidence intervals for risk. Since the baseline hazard is unspecified in the cox model, we use the non-parametric Breslow estimator at one, two, three, four, and five years. The linear predictor for the log hazard function is the parametric component of the Cox model, and it's output can be used directly in non-statistical languages. The standard errors of log survival are relatively complex to calculate outside of a statistical package, and so for 3) we constructed a linear regression model to approximate these log survival standard errors. Results from this regression model are parametric and can be used directly in calculating estimates of estimated standard errors.

Let $S_0(t)$ denote the baseline survival function at time t, or the probability of being medication free at time t when all covariate are set to their reference level value (all covariates are 0) and $S(t)$ is the survival function at time t. Then we can write the patient risk of exposure by time t with,

$$R(t;X) = 1 - S(t) = 1 - S_0(t;X)^{exp(X\beta)} \tag{1}$$

where X is a vector of covariate values and β is the vector of log hazard ratios. The value $X\beta$ is referred to as the linear predictor. The 95% confidence interval for risk is then given by,

$$95\%CI = 1 - exp[log(S(t)) \pm Z_{0.975} \times SE(log(S(t)))] \tag{2}$$

where $Z_{0.975} \approx 1.96$ is the 97.5^{th} percentile of a standard normal distribution, and $SE(logS(t))$ is the standard error for the $log\{S(t)\}$ that is approximated by 3) above. In the next section, we describe results from models 1 and 2. Subsections detail the output, show how to combine results to obtain $R(t;X)$ and the 95% confidence interval, describe covariate effects, and display model diagnostics.

3 Results

3.1 Predicted risk and example calculation

3.1.1 Model 1: BMI model

The linear predictor, $lp.bmi = X\beta$, estimated from the Cox PH model is given by,

lp.bmi ← function (baseline.age = 51, BMI = 28.051752, gender = "F", race3 = "White", diabetes = 0, coronary.disease = 0, afib = 0, hypertension = 0, atherosclerosis = 0, chf = 0, previous.clot = 0, dialysis = 0) { -4.7305438 + 0.065459713 * baseline.age - 3.0194841e-05 * pmax(baseline.age - 24, 0)^3 + 0.0001037356 $*$ $pmax(baseline.age - 41, 0)^3 - 0.00015032806 * pmax(baseline.age - 51, 0)^3 + 0.0001060992 * pmax(baseline.age - 60, 0)^3 - 2.9311903e-05 * pmax(baseline.age - 76, 0)^3 + 0.065024817 * BMI - 0.00010665707 * pmax(BMI - 20.07749, 0)^3 + 0.00012222854 * pmax(BMI - 24.578615, 0)^3 - 0.00010824019 * pmax(BMI - 28.051752, 0)^3 + 0.00016694303 * pmax(BMI - 32.435083, 0)^3 - 7.4274303e-05 * pmax(BMI - 43.639397, 0)^3 + 0.23279775 * (gender == "M") + 0.0860601 * (race3 == "Other") + 0.049906138 * (race3 == "White") + 0.58213821 * diabetes + 0.21707034 * coronary.disease + 0.4003414 * afib - 0.0013109944 * hypertension + 0.036763009 * atherosclerosis + 0.19367406 * chf + 0.11672828 * previous.clot + 0.67819823 * dialysis$}

Table 1: BMI model: baseline survivals at time t (days)

t	S(t)
0	1.000
365	0.926
730	0.863
1095	0.797
1460	0.725
1825	0.680

Using equation (1) and results from the Breslow estimator of baseline survival shown in table 1, the three-year risk (where t=1095) of medication exposure for two patients with demographic and clinical data shown in table 2 is calculated as follows:

Table 2: Patient input

	Predictor	Pt 1 value	Pt 2 value
1	age	54	61
2	BMI	25.80	23.48
3	gender	F	F
4	race	White	White
5	diabetes	no	no
6	coronary disease	no	no
7	afib	no	no
8	hypertension	no	yes
9	atherosclerosis	no	no
10	chf	no	no
11	previous clot	no	no
12	dialysis	no	no

$$\begin{aligned} R_1(1095; X_1) &= 1 - S_0(1095)^{exp(lp.bmi(X_1))} \\ &= 1 - 0.797^{exp(-0.079)} \\ &= 1 - 0.811 \\ &= 0.189 \end{aligned}$$

$$\begin{aligned} R_2(1095; X_2) &= 1 - S_0(1095)^{exp(lp.bmi(X_2))} \\ &= 1 - 0.797^{exp(-0.016)} \\ &= 1 - 0.800 \\ &= 0.200 \end{aligned}$$

To approximate the standard errors of $log\{S(t)\}$ used in calculations of confidence intervals for risk, we constructed a linear regression model that regressed the (relatively complicated) estimated standard error of log survival (Tsiatis, 1990; Harrell 2001) from the Cox PH model on flexible functionals of all covariates included in the regression models. The idea behind this strategy is to develop an easily implementable estimate of the estimated standard error. In this case, the regression resulted in an (R^2=0.97). That is, using flexible functionals of covariates, we can predict 97% of variation in the standard errors. Letting $\hat{}$ denoting estimated values, the estimated 95% confidence interval of survival (probability of being medication-free at year 3) can be obtained by

$$exp(log(\hat{S(t)}) \pm 1.96 \times \hat{SE})$$

where $\hat{SE} = exp(linear.predictor)$, and

linear.predictor ← function (baseline.age = 50, BMI = 28.390178, gender = "F", race3 = "White", diabetes = 0, coronary.disease = 0, afib = 0, hypertension = 0, atherosclerosis = 0, chf = 0, previous.clot = 0, dialysis = 0) { -5.4538714 + 0.021251831 * baseline.age + 8.413256e-06 * pmax(baseline.age - 24, 0)^3 $- 4.0672572e-05 * pmax(baseline.age-41,0)^3$+3.8010225$e-05*pmax(baseline.age-50,0)^3$+2.2098643$e-07*pmax(baseline.age-59,0)^3 - 5.971895e-06 * pmax(baseline.age-75,0)^3 - 0.019089831 * BMI + 0.00096806017 * pmax(BMI - 20.293553,0)^3 - 0.0028627413 * pmax(BMI - 24.678555,0)^3 + 0.0026124876 * pmax(BMI - 28.390178,0)^3 - 0.0007432145 * pmax(BMI - 32.677073,0)^3 + 2.5408079e-05 * pmax(BMI - 44.081553,0)^3 + 0.24241605 * (gender == "M")+0.52928324*(race3 == "Other")-0.10294989*(race3 == "White")+0.69835658*diabetes+ 0.15456796 * coronary.disease + 2.4086646 * afib + 0.067561462 * hypertension + 0.38618988 * atherosclerosis + 0.53195157 * chf + 0.88238038 * previous.clot + 1.5397411 * dialysis\}$

For patient 1, the 95% CI for medication-free survival, estimated from the Cox PH model, was (0.796, 0.8259). This value is approximated, using the *linear.predictor* function, with (0.7951, 0.8269). From the approximation, the outputted 95% CI of risk $\{1 - S(t)\}$ at three years is (0.1731, 0.2049).

Similarly, for patient 2, the approximated 95% CI of risk at three years is (0.1843, 0.2157), and the full calculation yields (0.1809, 0.219).

3.1.2 Model 2: Weight model

The following are linear predictor from the Cox PH target model, the Breslow estimator of the baseline survival, and the linear predictor of the standard error of log survival for the model that included weight as opposed to BMI.

Linear predictor function for the Cox PH model:

lp.wei ← function (baseline.age = 51, weight = 81.215, gender = "F", race3 = "White", diabetes = 0, coronary.disease = 0, afib = 0, hypertension = 0, atherosclerosis = 0, chf = 0, previous.clot = 0, dialysis = 0) { -4.093934 + 0.062887454 * baseline.age - 2.7663616e-05 * pmax(baseline.age - 25, 0)^3 + $8.704417e-05 * pmax(baseline.age-41,0)^3 - 0.00012894443 * pmax(baseline.age-51,0)^3 + 9.7498419e-05 * pmax(baseline.age-60,0)^3 - 2.7934541e-05 * pmax(baseline.age-77,0)^3 + 0.014879719 * weight + 4.7831395e-06 * pmax(weight-53.52,0)^3 - 2.5301908e-05 * pmax(weight-69,0)^3 + 3.0660501e-05 * pmax(weight-81.215,0)^3 - 9.0792408e-06 * pmax(weight-95.28,0)^3 - 1.062491e-06 * pmax(weight-127.2335,0)^3 + 0.10159946 * (gender == "M") + 0.12720848 * (race3 == "Other") + 0.047369679 * (race3 == "White") + 0.59062019 * diabetes + 0.26573081 * coronary.disease + 0.78302641 * afib - 0.03548041 * hypertension + 0.050988742 * atherosclerosis + 0.24422313 * chf + 0.14582975 * previous.clot + 0.70935756 * dialysis$}

Table 3: Weight model: baseline survivals at time t (days)

t	S(t)
0	1.000
365	0.927
730	0.865
1095	0.795
1460	0.714
1825	0.662

Linear predictor function of the estimated standard error of $log(S(t))$:

linear.predictor ← function (baseline.age = 51, weight = 80.625, gender = "F", race3 = "White", diabetes = 0, coronary.disease = 0, afib = 0, hypertension = 0, atherosclerosis = 0, chf = 0, previous.clot = 0, dialysis = 0) { -5.4855211 + 0.019960095 * baseline.age + 1.1353636e-05 * pmax(baseline.age - 24, 0)^3 $- 6.3361016e-05 * pmax(baseline.age-42,0)^3 + 7.2839462e-05 * pmax(baseline.age-51,0)^3 - 1.6598067e-05 * pmax(baseline.age-60,0)^3 - 4.234014e-06 * pmax(baseline.age-78,0)^3 - 0.0065294144 * weight + 3.0281935e-05 * pmax(weight-53.98,0)^3 - 0.00010398669 * pmax(weight-68.95,0)^3 + 0.00010944161 * pmax(weight-80.625,0)^3 - 3.8917852e-05 * pmax(weight - 94.8055,0)^3 + 3.180992e-06 * pmax(weight - 126.1115,0)^3 + 0.1120716 * (gender == "M") + 0.57705027 * (race3 == "Other") - 0.11366799 * (race3 == "White") + 0.70217287 * diabetes + 0.16584989 * coronary.disease + 2.4671425 * afib + 0.035538372 * hypertension + 0.42986377 * atherosclerosis + 0.52882789 * chf + 0.96592082 * previous.clot + 1.5594269 * dialysis$}

3.2 Covariate Effects

3.2.1 Model 1: with BMI

13

Cox Proportional Hazards Model

```
cph(formula = s ~ rcs(baseline.age, 5) + rcs(BMI, 5) + gender +
    race3 + diabetes + coronary.disease + afib + hypertension +
    atherosclerosis + chf + previous.clot + dialysis, data = Baseline.bmi,
    x = TRUE, y = TRUE, surv = TRUE, time.inc = 365)
```

	Model Tests		Discrimination Indexes	
Obs 16020	LR χ^2	1228.56	R^2	0.075
Events 3752	d.f.	19	g	0.728
Center 4.731	Pr($>\chi^2$)	0.0000	g_r	2.071
	Score chi2	1232.03		
	Pr($>\chi^2$)	0.0000		

	Coef	S.E.	Wald Z	Pr($>\|Z\|$)
baseline.age	0.065	0.010	6.73	$<$ 0.0001
baseline.age'	-0.082	0.036	-2.30	0.0217
baseline.age"	0.281	0.166	1.69	0.0919
baseline.age"'	-0.406	0.254	-1.60	0.1090
BMI	0.065	0.021	3.04	0.0023
BMI'	-0.059	0.209	-0.28	0.7772
BMI"	0.068	0.688	0.10	0.9214
BMI"'	-0.060	0.697	-0.09	0.9313
gender=M	0.233	0.034	6.91	$<$ 0.0001
race3=Other	0.086	0.107	0.80	0.4214
race3=White	0.050	0.046	1.08	0.2814
diabetes	0.582	0.040	14.71	$<$ 0.0001
coronary.disease	0.217	0.091	2.39	0.0166
afib	0.400	0.448	0.89	0.3710
hypertension	-0.001	0.037	-0.04	0.9714
atherosclerosis	0.037	0.074	0.50	0.6203
chf	0.194	0.076	2.55	0.0108
previous.clot	0.117	0.144	0.81	0.4166
dialysis	0.678	0.154	4.40	$<$ 0.0001

$$\text{Prob}\{T \geq t\} = S_0(t)^{e^{X\beta}}, \quad \text{where}$$

$$
\begin{aligned}
X\hat{\beta} = \\
&-4.730544 \\
&+0.06545971\text{baseline.age} - 3.019484\times10^{-5}(\text{baseline.age}-24)_+^3 \\
&+0.0001037356(\text{baseline.age}-41)_+^3 - 0.0001503281(\text{baseline.age}-51)_+^3 \\
&+0.0001060992(\text{baseline.age}-60)_+^3 - 2.931190\times10^{-5}(\text{baseline.age}-76)_+^3 \\
&+0.06502482\text{BMI} - 0.0001066571(\text{BMI}-20.07749)_+^3 \\
&+0.0001222285(\text{BMI}-24.57861)_+^3 - 0.0001082402(\text{BMI}-28.05175)_+^3 \\
&+0.0001669430(\text{BMI}-32.43508)_+^3 - 7.42743\times10^{-5}(\text{BMI}-43.6394)_+^3 \\
&+0.2327978\{\text{M}\} \\
&+0.08606010\{\text{Other}\} + 0.04990614\{\text{White}\} + 0.5821382\ \text{diabetes} \\
&+0.2170703\ \text{coronary.disease} + 0.4003414\ \text{afib} - 0.001310994\ \text{hypertension} \\
&+0.03676301\ \text{atherosclerosis} + 0.1936741\ \text{chf} + 0.1167283\ \text{previous.clot} \\
&+0.6781982\ \text{dialysis}
\end{aligned}
$$

and $\{c\} = 1$ if subject is in group c, 0 otherwise; $(x)_+ = x$ if $x > 0$, 0 otherwise.

t	$S_0(t)$
0	1.000
365	0.926
730	0.863
1095	0.797
1460	0.725
1825	0.680

Table 4: Wald Statistics for `s`

	χ^2	$d.f.$	P
baseline.age	477.94	4	< 0.0001
Nonlinear	73.68	3	< 0.0001
BMI	190.79	4	< 0.0001
Nonlinear	39.96	3	< 0.0001
gender	47.75	1	< 0.0001
race3	1.35	2	0.5082
diabetes	216.30	1	< 0.0001
coronary.disease	5.73	1	0.0166
afib	0.80	1	0.3710
hypertension	0.00	1	0.9714
atherosclerosis	0.25	1	0.6203
chf	6.50	1	0.0108
previous.clot	0.66	1	0.4166
dialysis	19.36	1	< 0.0001
TOTAL NONLINEAR	124.79	6	< 0.0001
TOTAL	1151.59	19	< 0.0001

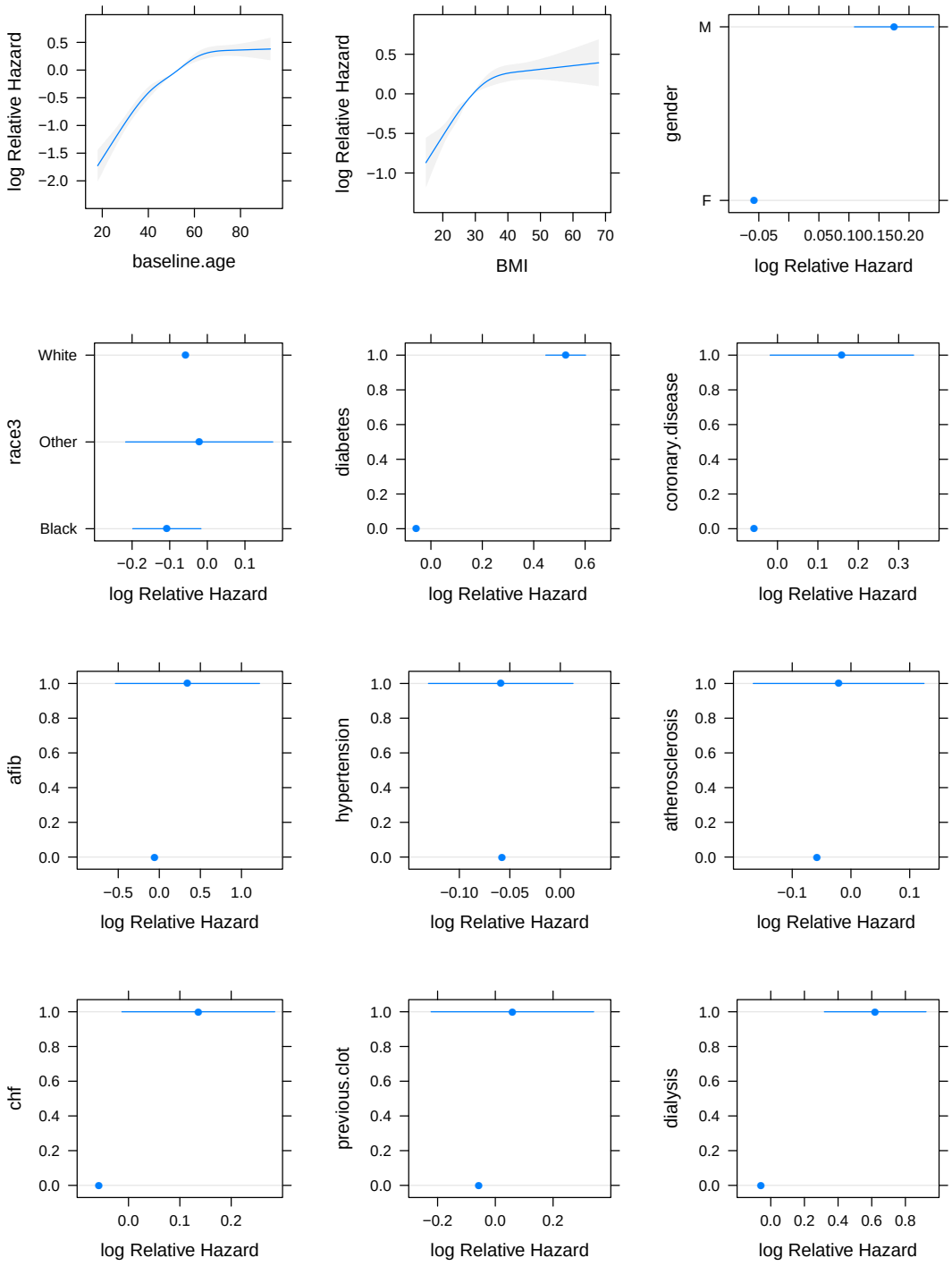


Figure 1: BMI model: Predictor effects on the log hazard ratio scale with pointwise .95 confidence bands

3.3 Model 2: with weight

13

Cox Proportional Hazards Model

```
cph(formula = s ~ rcs(baseline.age, 5) + rcs(weight, 5) + gender +
    race3 + diabetes + coronary.disease + afib + hypertension +
    atherosclerosis + chf + previous.clot + dialysis, data = Baseline.wei,
    x = TRUE, y = TRUE, surv = TRUE, time.inc = 365)
```

	Model Tests		Discrimination Indexes	
Obs 18340	LR χ^2	1410.55	R^2	0.075
Events 4477	d.f.	19	g	0.714
Center 4.094	Pr($>\chi^2$)	0.0000	g_r	2.041
	Score chi2	1430.00		
	Pr($>\chi^2$)	0.0000		

	Coef	S.E.	Wald Z	Pr($>\|Z\|$)
baseline.age	0.063	0.009	7.18	$<$ 0.0001
baseline.age'	-0.075	0.036	-2.08	0.0378
baseline.age"	0.235	0.158	1.49	0.1371
baseline.age"'	-0.349	0.233	-1.50	0.1340
weight	0.015	0.006	2.68	0.0074
weight'	0.026	0.046	0.56	0.5744
weight"	-0.137	0.154	-0.89	0.3731
weight"'	0.167	0.167	1.00	0.3184
gender=M	0.102	0.033	3.12	0.0018
race3=Other	0.127	0.101	1.26	0.2063
race3=White	0.047	0.042	1.12	0.2636
diabetes	0.591	0.037	15.86	$<$ 0.0001
coronary.disease	0.266	0.083	3.19	0.0014
afib	0.783	0.324	2.42	0.0156
hypertension	-0.035	0.034	-1.05	0.2928
atherosclerosis	0.051	0.070	0.73	0.4645
chf	0.244	0.066	3.70	0.0002
previous.clot	0.146	0.135	1.08	0.2783
dialysis	0.709	0.149	4.78	$<$ 0.0001

$$\text{Prob}\{T \geq t\} = S_0(t)^{e^{X\beta}}, \quad \text{where}$$

$$
\begin{aligned}
X\hat{\beta} = & \\
& -4.093934 \\
& +0.06288745\text{baseline.age} - 2.766362\times10^{-5}(\text{baseline.age}-25)_+^3 \\
& +8.704417\times10^{-5}(\text{baseline.age}-41)_+^3 - 0.0001289444(\text{baseline.age}-51)_+^3 \\
& +9.749842\times10^{-5}(\text{baseline.age}-60)_+^3 - 2.793454\times10^{-5}(\text{baseline.age}-77)_+^3 \\
& +0.01487972\text{weight} + 4.783139\times10^{-6}(\text{weight}-53.52)_+^3 \\
& -2.530191\times10^{-5}(\text{weight}-69)_+^3 + 3.06605\times10^{-5}(\text{weight}-81.215)_+^3 \\
& -9.07924\times10^{-6}(\text{weight}-95.28)_+^3 - 1.062491\times10^{-6}(\text{weight}-127.2335)_+^3 \\
& +0.1015995\{\text{M}\} \\
& +0.12720848\{\text{Other}\} + 0.04736968\{\text{White}\} + 0.5906202\ \text{diabetes} \\
& +0.2657308\ \text{coronary.disease} + 0.7830264\ \text{afib} - 0.03548041\ \text{hypertension} \\
& +0.05098874\ \text{atherosclerosis} + 0.2442231\ \text{chf} + 0.1458297\ \text{previous.clot} \\
& +0.7093576\ \text{dialysis}
\end{aligned}
$$

and $\{c\} = 1$ if subject is in group c, 0 otherwise; $(x)_+ = x$ if $x > 0$, 0 otherwise.

t	$S_0(t)$
0	1.000
365	0.927
730	0.865
1095	0.795
1460	0.714
1825	0.662

Table 5: Wald Statistics for `s`

	χ^2	$d.f.$	P
baseline.age	596.78	4	< 0.0001
Nonlinear	100.38	3	< 0.0001
weight	161.57	4	< 0.0001
Nonlinear	35.30	3	< 0.0001
gender	9.74	1	0.0018
race3	2.10	2	0.3496
diabetes	251.58	1	< 0.0001
coronary.disease	10.19	1	0.0014
afib	5.84	1	0.0156
hypertension	1.11	1	0.2928
atherosclerosis	0.53	1	0.4645
chf	13.67	1	0.0002
previous.clot	1.18	1	0.2783
dialysis	22.81	1	< 0.0001
TOTAL NONLINEAR	150.96	6	< 0.0001
TOTAL	1336.70	19	< 0.0001

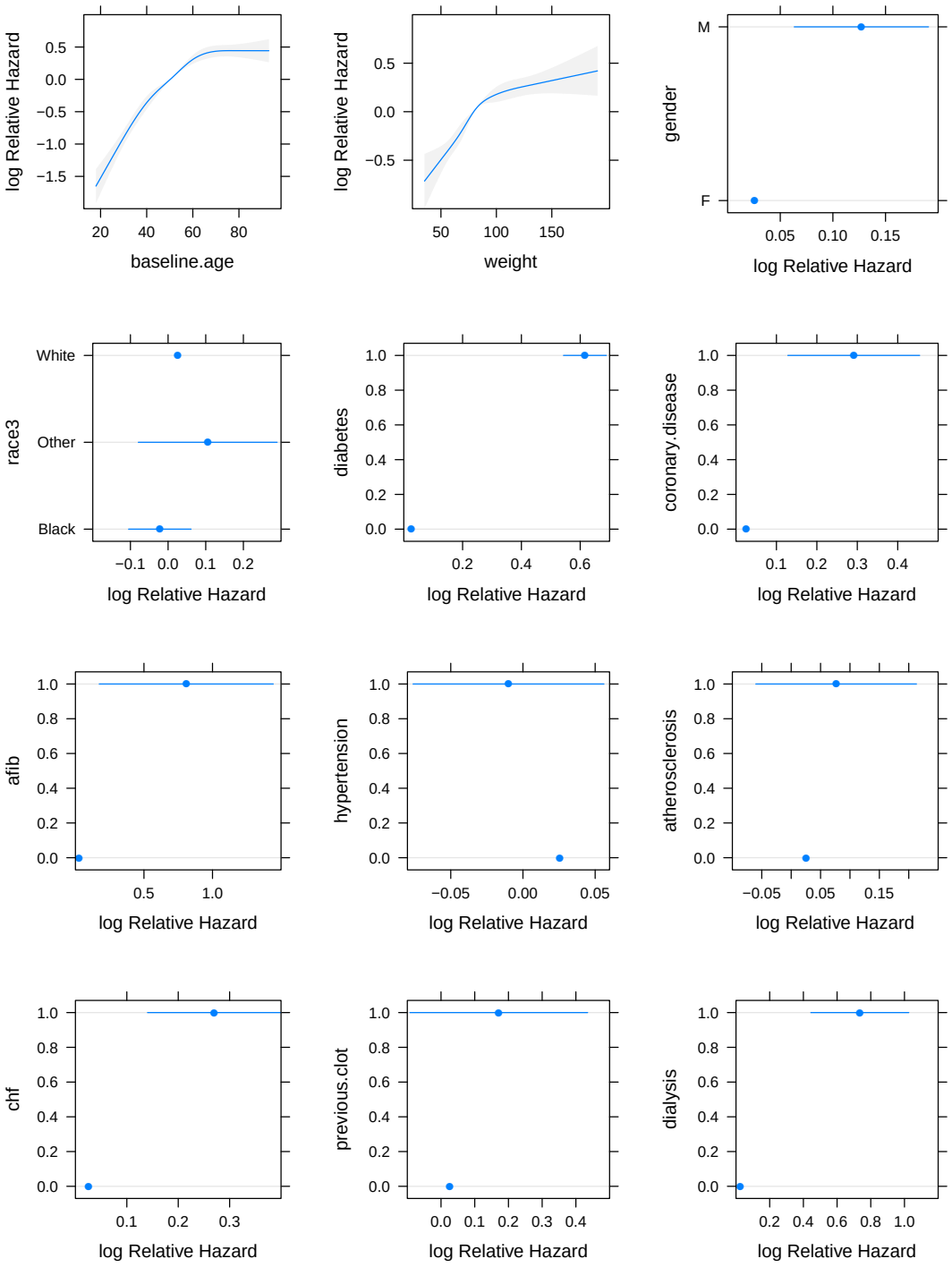


Figure 2: Weight model: Predictor effects on the log hazard ratio scale with pointwise .95 confidence bands

3.4 Primary Models: Evaluation and Diagnostics

3.4.1 Prediction and discrimination

Table 6: The distribution (percentiles) of medication free survival at three years

	Percentile	BMI model	Weight model
1	0%	0.159	0.107
2	1%	0.407	0.413
3	5%	0.542	0.545
4	10%	0.616	0.623
5	25%	0.705	0.702
6	50%	0.783	0.782
7	75%	0.853	0.851
8	90%	0.912	0.909
9	95%	0.937	0.934
10	99%	0.964	0.961
11	100%	0.982	0.976

3.4.2 Discrimination: C index

Table 7: C index

	BMI model	Weight model
1	0.666	0.663

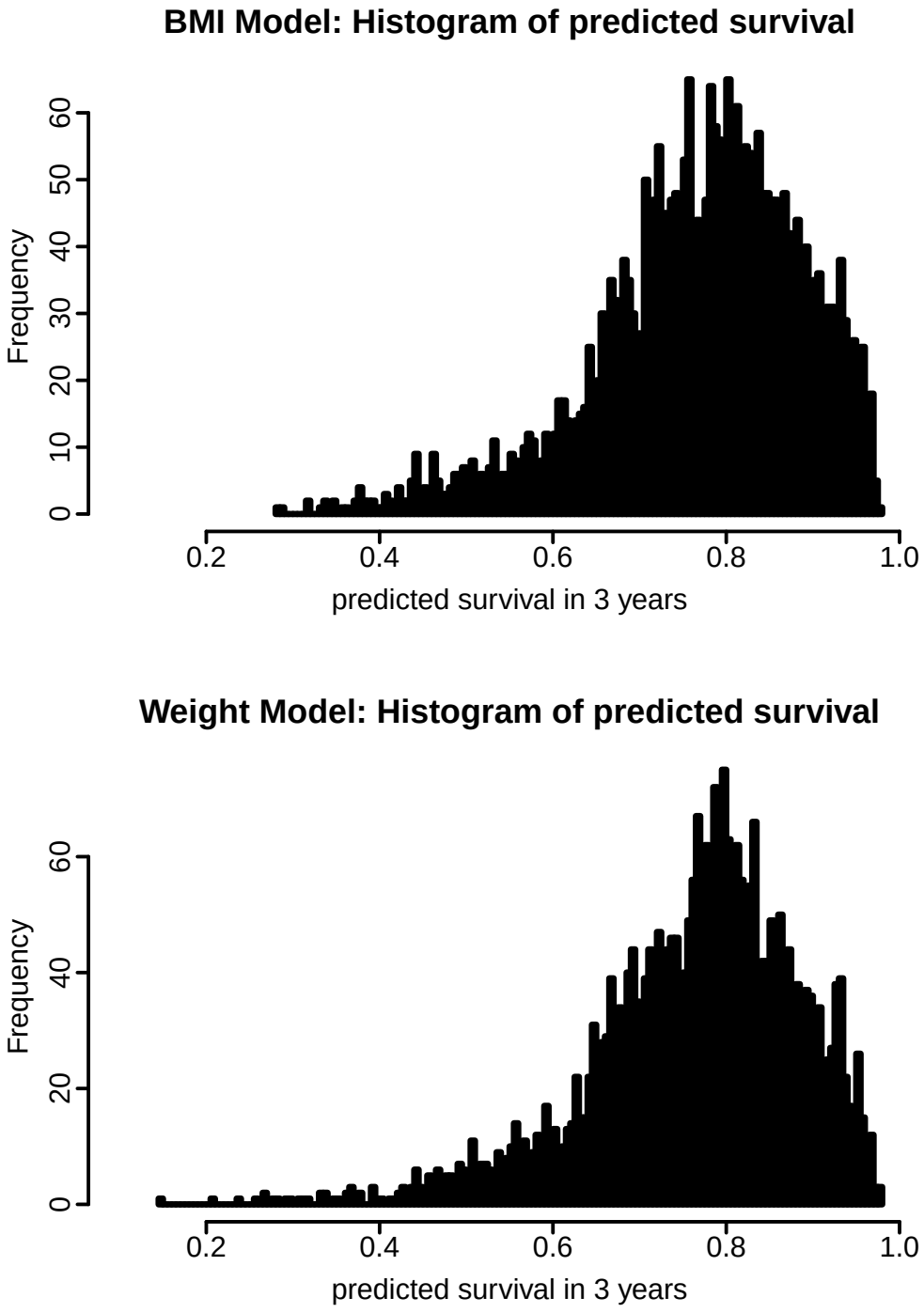


Figure 3: Histograms of predicted probability of medication free survival at 3 years

4 Validation and Calibration

2in

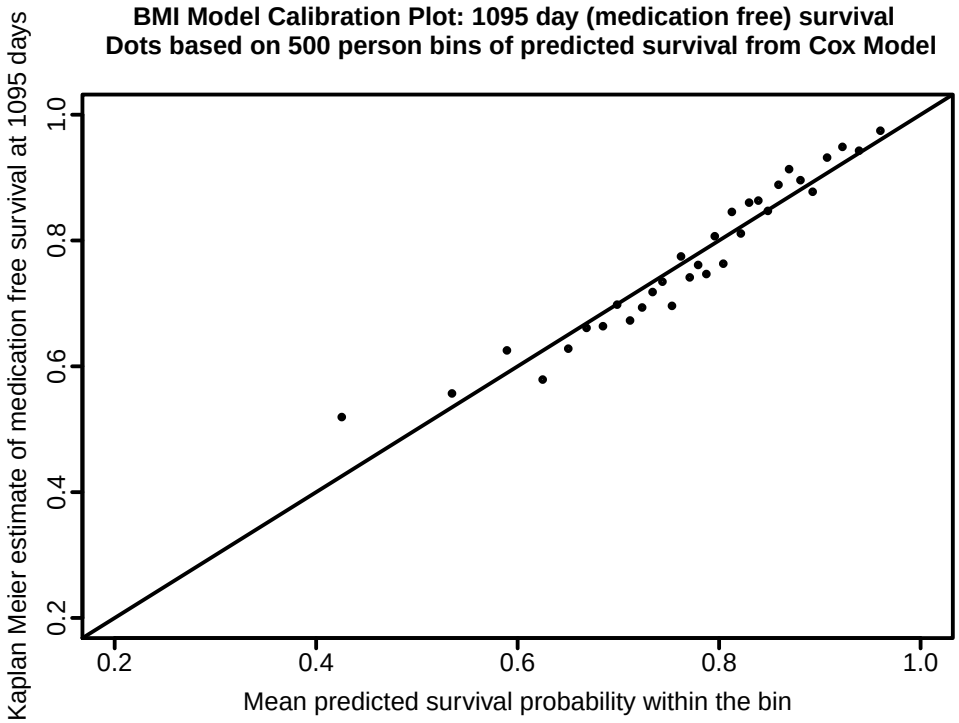
BMI Model Validation

C index	Index	Original Sample	Training Sample	Test Sample	Optimism	Corrected Index	n
D_{xy}		−0.332	−0.335	−0.330	−0.005	−0.327	100
R^2		0.075	0.076	0.074	0.002	0.072	100
Slope		1.000	1.000	0.980	0.020	0.980	100
D		0.018	0.018	0.018	0.001	0.017	100
U		0.000	0.000	0.000	0.000	0.000	100
Q		0.018	0.018	0.018	0.001	0.017	100
g		0.728	0.736	0.720	0.016	0.712	100

2in

Weight Model Validation

Index	Original Sample	Training Sample	Test Sample	Optimism	Corrected Index	n
D_{xy}	−0.325	−0.326	−0.323	−0.003	−0.322	100
R^2	0.075	0.076	0.074	0.002	0.073	100
Slope	1.000	1.000	0.981	0.019	0.981	100
D	0.017	0.017	0.017	0.001	0.016	100
U	0.000	0.000	0.000	0.000	0.000	100
Q	0.017	0.017	0.017	0.001	0.016	100
g	0.714	0.721	0.706	0.015	0.699	100



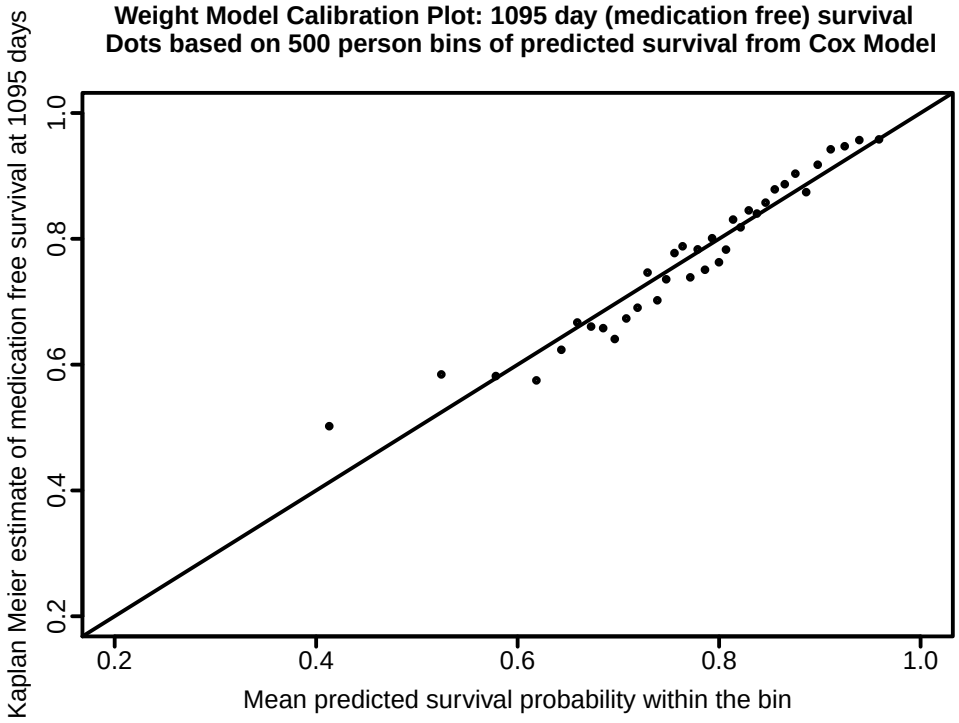


Figure 4: Calibration plot

4.1 Linear regression model to estimate the standard error of $log(S(t))$ and therefore the confidence intervals for $R(t;X)$

A randomly selected sample of n = 3000 patient records was used to develop this predictive ordinary least squares (OLS) model to estimate the standard error of log survival at year 3 (i.e., the probability of being medications-free at year 3).

4.1.1 Model 1: BMI model

$$\text{E(log(sqrt(varS.bmi))) } = X\beta, \quad \text{where}$$

$$\begin{aligned} X\hat{\beta} = & \\ & -5.453871 \\ & +0.02125183\text{baseline.age} + 8.413256\times10^{-6}(\text{baseline.age} - 24)_+^3 \\ & -4.067257\times10^{-5}(\text{baseline.age} - 41)_+^3 + 3.801022\times10^{-5}(\text{baseline.age} - 50)_+^3 \\ & +2.209864\times10^{-7}(\text{baseline.age} - 59)_+^3 - 5.971895\times10^{-6}(\text{baseline.age} - 75)_+^3 \\ & -0.01908983\text{BMI} + 0.0009680602(\text{BMI} - 20.29355)_+^3 \\ & -0.002862741(\text{BMI} - 24.67856)_+^3 + 0.002612488(\text{BMI} - 28.39018)_+^3 \\ & -0.0007432145(\text{BMI} - 32.67707)_+^3 + 2.540808\times10^{-5}(\text{BMI} - 44.08155)_+^3 \\ & +0.2424161\{\text{M}\} \\ & +0.5292832\{\text{Other}\} - 0.1029499\{\text{White}\} + 0.6983566\ \text{diabetes} \\ & +0.1545680\ \text{coronary.disease} + 2.408665\ \text{afib} + 0.06756146\ \text{hypertension} \\ & +0.3861899\ \text{atherosclerosis} + 0.5319516\ \text{chf} + 0.8823804\ \text{previous.clot} \\ & +1.539741\ \text{dialysis} \end{aligned}$$

and $\{c\} = 1$ if subject is in group c, 0 otherwise; $(x)_+ = x$ if $x > 0$, 0 otherwise.

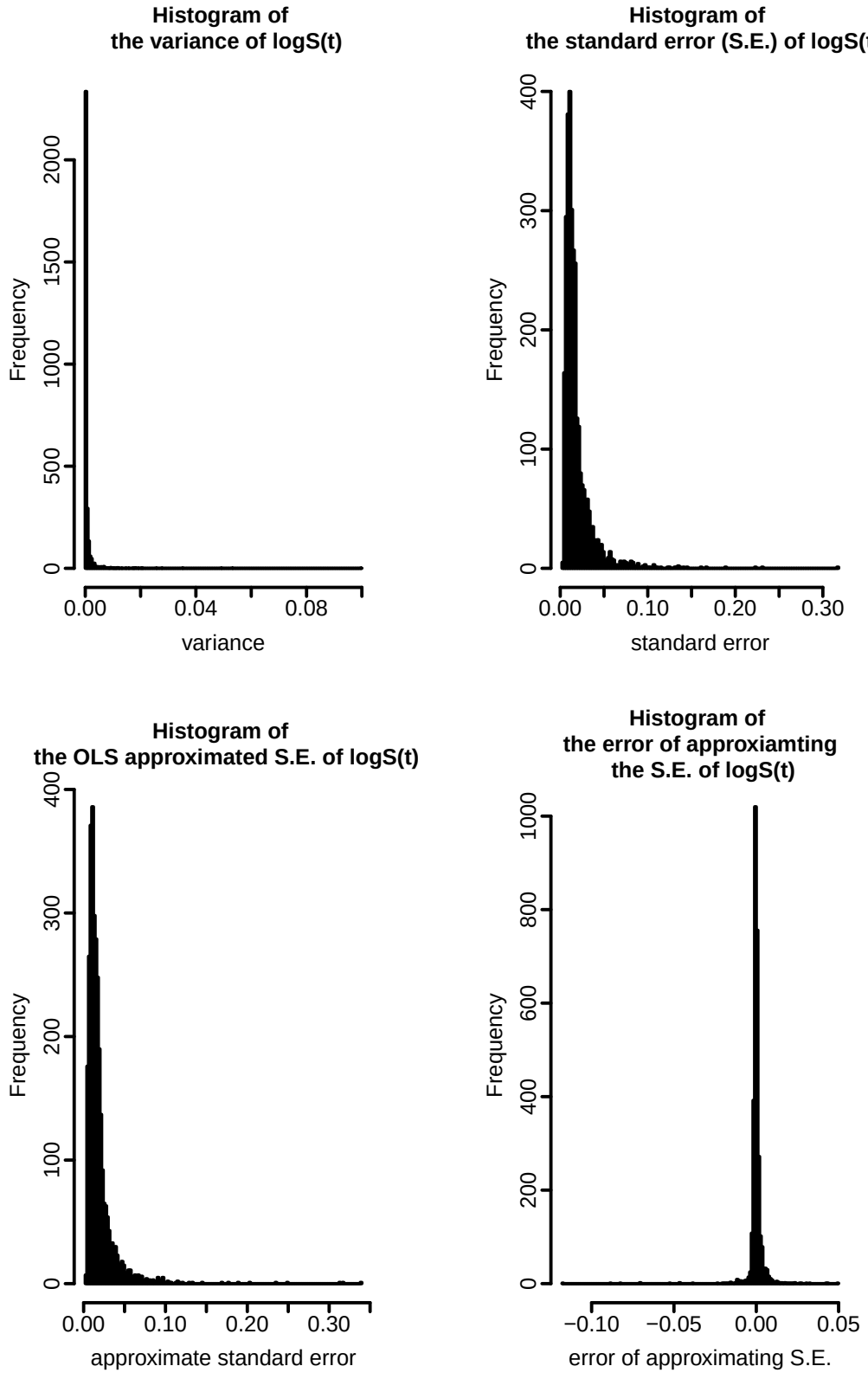
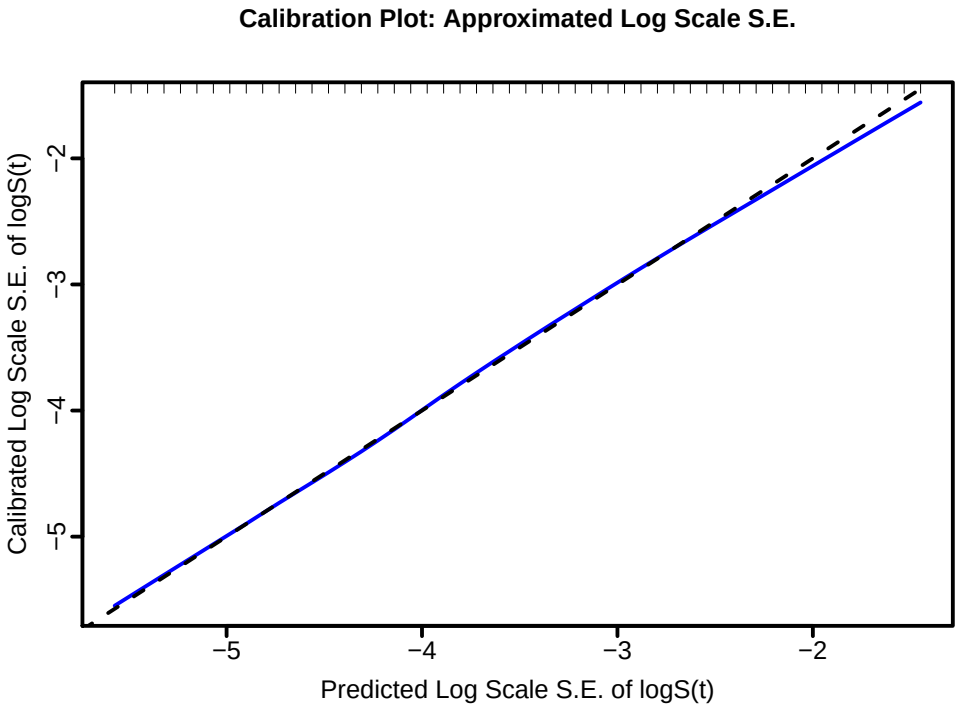


Figure 5: BMI OLS histogram: exact and approximated standard errors

```
The distribution(percentiles) of the error of approximating the S.E. of logS(t):

            0%             1%             2%             3%             4%
-1.176767e-01 -1.096718e-02 -5.810560e-03 -3.618146e-03 -2.769619e-03
            5%             6%             7%             8%             9%
-2.410362e-03 -2.216936e-03 -2.020481e-03 -1.904462e-03 -1.788277e-03
           10%            11%            12%            13%            14%
-1.700923e-03 -1.565624e-03 -1.468889e-03 -1.421017e-03 -1.358445e-03
           15%            16%            17%            18%            19%
-1.304102e-03 -1.242342e-03 -1.181983e-03 -1.119286e-03 -1.071203e-03
           20%            21%            22%            23%            24%
-1.012435e-03 -9.635687e-04 -9.308945e-04 -8.936793e-04 -8.579272e-04
           25%            26%            27%            28%            29%
-8.204427e-04 -7.881170e-04 -7.497822e-04 -7.199267e-04 -6.942550e-04
           30%            31%            32%            33%            34%
-6.601604e-04 -6.345486e-04 -6.110876e-04 -5.885188e-04 -5.624600e-04
           35%            36%            37%            38%            39%
-5.249614e-04 -4.946561e-04 -4.726863e-04 -4.481089e-04 -4.312221e-04
           40%            41%            42%            43%            44%
-3.947659e-04 -3.569881e-04 -3.264038e-04 -2.983732e-04 -2.860814e-04
           45%            46%            47%            48%            49%
-2.632062e-04 -2.306273e-04 -2.004323e-04 -1.694991e-04 -1.443536e-04
           50%            51%            52%            53%            54%
-1.127867e-04 -8.368236e-05 -6.483425e-05 -3.244242e-05 -8.139664e-06
           55%            56%            57%            58%            59%
 2.389796e-05  5.434737e-05  7.957481e-05  1.062302e-04  1.450137e-04
           60%            61%            62%            63%            64%
 1.769499e-04  2.032281e-04  2.421421e-04  2.838736e-04  3.204993e-04
           65%            66%            67%            68%            69%
 3.478454e-04  3.767256e-04  4.164681e-04  4.459058e-04  4.822977e-04
           70%            71%            72%            73%            74%
 5.165405e-04  5.513201e-04  5.964796e-04  6.436732e-04  6.889327e-04
           75%            76%            77%            78%            79%
 7.443022e-04  8.026600e-04  8.518220e-04  9.079256e-04  9.794699e-04
           80%            81%            82%            83%            84%
 1.056948e-03  1.128638e-03  1.202669e-03  1.300093e-03  1.392860e-03
           85%            86%            87%            88%            89%
 1.473258e-03  1.606273e-03  1.709243e-03  1.888099e-03  2.107408e-03
           90%            91%            92%            93%            94%
 2.412608e-03  2.767337e-03  3.087406e-03  3.313184e-03  3.763320e-03
           95%            96%            97%            98%            99%
 4.312942e-03  5.237838e-03  6.127191e-03  7.375056e-03  1.073018e-02
          100%
 4.926210e-02
```



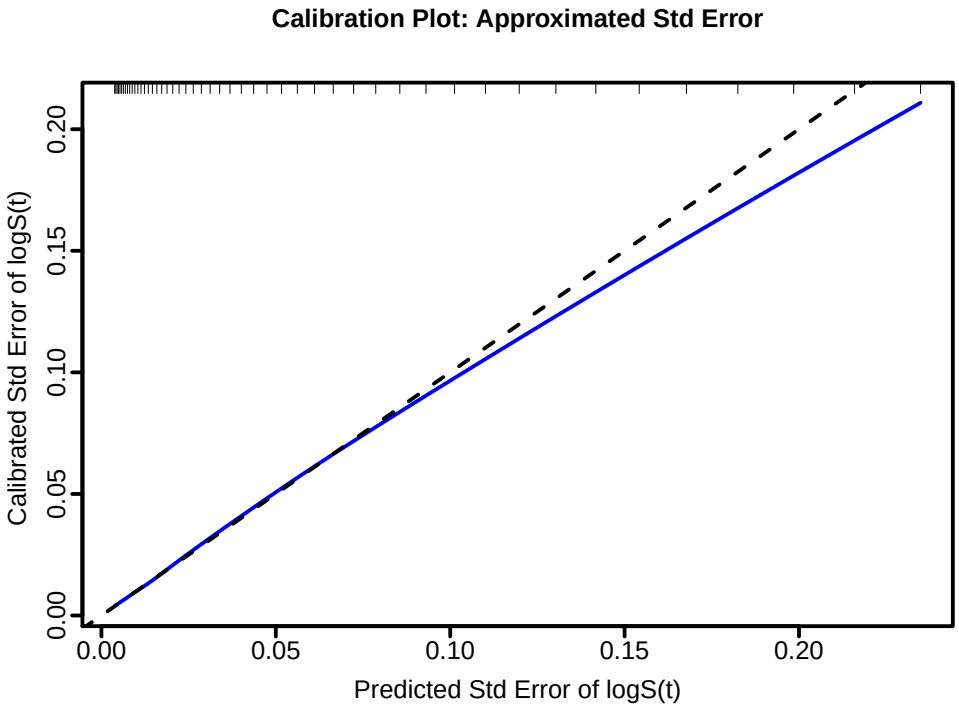


Figure 6: Calibration plot: predictive model on for standard errors of log(S(t)) (BMI model)

4.1.2 Model 2: Weight model

$$\text{E}(\log(\text{sqrt}(\text{varS.wei}))) = X\beta, \quad \text{where}$$

$$\begin{aligned}
X\hat{\beta} = \\
&-5.485521 \\
&+0.01996009\text{baseline.age} + 1.135364\times10^{-5}(\text{baseline.age} - 24)_+^3 \\
&-6.336102\times10^{-5}(\text{baseline.age} - 42)_+^3 + 7.283946\times10^{-5}(\text{baseline.age} - 51)_+^3 \\
&-1.659807\times10^{-5}(\text{baseline.age} - 60)_+^3 - 4.234014\times10^{-6}(\text{baseline.age} - 78)_+^3 \\
&-0.006529414\text{weight} + 3.028193\times10^{-5}(\text{weight} - 53.98)_+^3 \\
&-0.0001039867(\text{weight} - 68.95)_+^3 + 0.0001094416(\text{weight} - 80.625)_+^3 \\
&-3.891785\times10^{-5}(\text{weight} - 94.8055)_+^3 + 3.180992\times10^{-6}(\text{weight} - 126.1115)_+^3 \\
&+0.1120716\{\text{M}\} \\
&+0.5770503\{\text{Other}\} - 0.1136680\{\text{White}\} + 0.7021729\ \text{diabetes} \\
&+0.1658499\ \text{coronary.disease} + 2.467142\ \text{afib} + 0.03553837\ \text{hypertension} \\
&+0.4298638\ \text{atherosclerosis} + 0.5288279\ \text{chf} + 0.9659208\ \text{previous.clot} \\
&+1.559427\ \text{dialysis}
\end{aligned}$$

and $\{c\} = 1$ if subject is in group c, 0 otherwise; $(x)_+ = x$ if $x > 0$, 0 otherwise.

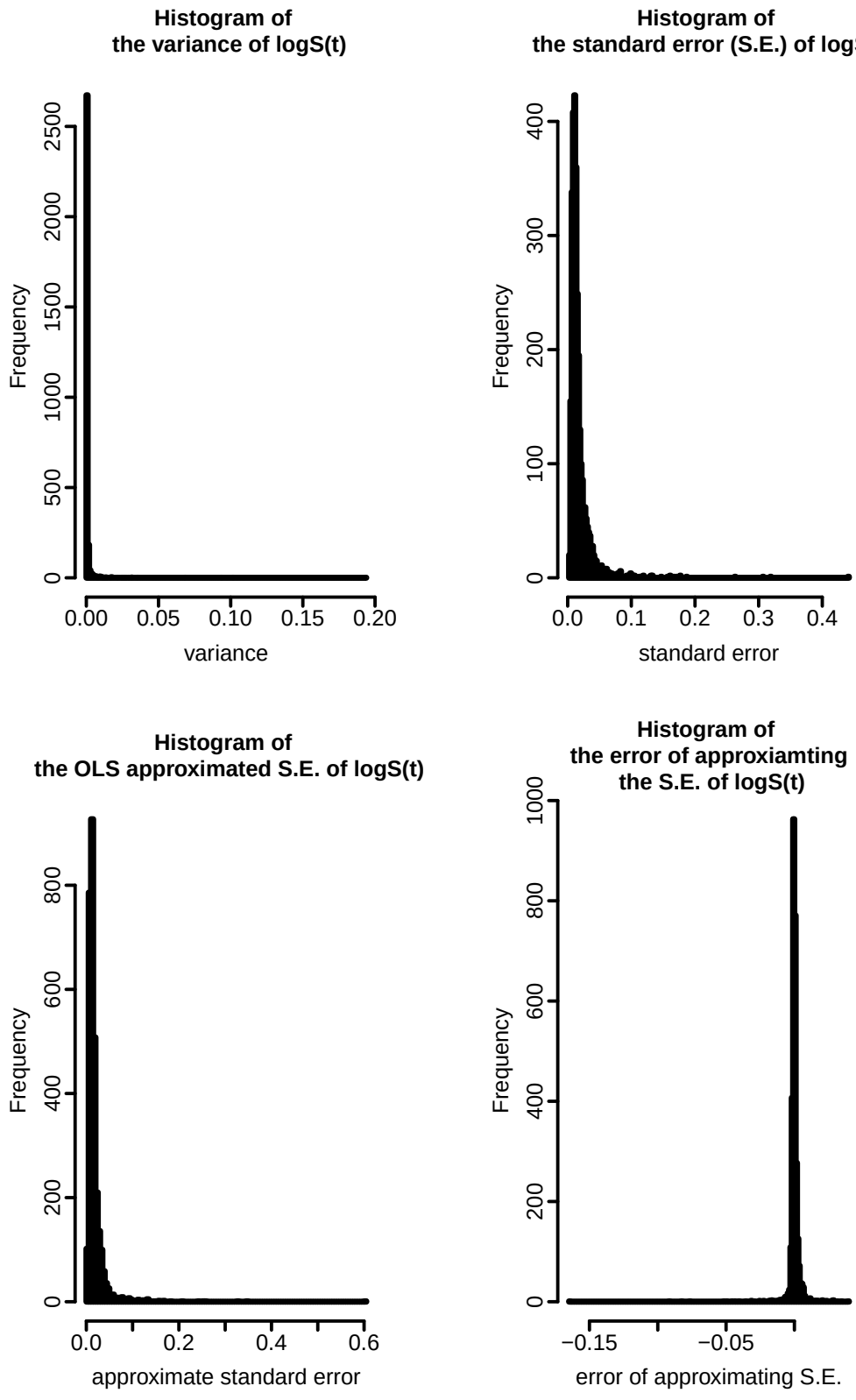


Figure 7: Weight OLS histogram: exact and approximated standard error

The distribution(percentiles) of the error of approximating the standard error of log(S(t)):

0%	1%	2%	3%	4%
-1.176767e-01	-1.096718e-02	-5.810560e-03	-3.618146e-03	-2.769619e-03
5%	6%	7%	8%	9%
-2.410362e-03	-2.216936e-03	-2.020481e-03	-1.904462e-03	-1.788277e-03
10%	11%	12%	13%	14%
-1.700923e-03	-1.565624e-03	-1.468889e-03	-1.421017e-03	-1.358445e-03
15%	16%	17%	18%	19%
-1.304102e-03	-1.242342e-03	-1.181983e-03	-1.119286e-03	-1.071203e-03
20%	21%	22%	23%	24%
-1.012435e-03	-9.635687e-04	-9.308945e-04	-8.936793e-04	-8.579272e-04
25%	26%	27%	28%	29%
-8.204427e-04	-7.881170e-04	-7.497822e-04	-7.199267e-04	-6.942550e-04
30%	31%	32%	33%	34%
-6.601604e-04	-6.345486e-04	-6.110876e-04	-5.885188e-04	-5.624600e-04
35%	36%	37%	38%	39%
-5.249614e-04	-4.946561e-04	-4.726863e-04	-4.481089e-04	-4.312221e-04
40%	41%	42%	43%	44%
-3.947659e-04	-3.569881e-04	-3.264038e-04	-2.983732e-04	-2.860814e-04
45%	46%	47%	48%	49%
-2.632062e-04	-2.306273e-04	-2.004323e-04	-1.694991e-04	-1.443536e-04
50%	51%	52%	53%	54%
-1.127867e-04	-8.368236e-05	-6.483425e-05	-3.244242e-05	-8.139664e-06
55%	56%	57%	58%	59%
2.389796e-05	5.434737e-05	7.957481e-05	1.062302e-04	1.450137e-04
60%	61%	62%	63%	64%
1.769499e-04	2.032281e-04	2.421421e-04	2.838736e-04	3.204993e-04
65%	66%	67%	68%	69%
3.478454e-04	3.767256e-04	4.164681e-04	4.459058e-04	4.822977e-04
70%	71%	72%	73%	74%
5.165405e-04	5.513201e-04	5.964796e-04	6.436732e-04	6.889327e-04
75%	76%	77%	78%	79%
7.443022e-04	8.026600e-04	8.518220e-04	9.079256e-04	9.794699e-04
80%	81%	82%	83%	84%
1.056948e-03	1.128638e-03	1.202669e-03	1.300093e-03	1.392860e-03
85%	86%	87%	88%	89%
1.473258e-03	1.606273e-03	1.709243e-03	1.888099e-03	2.107408e-03
90%	91%	92%	93%	94%
2.412608e-03	2.767337e-03	3.087406e-03	3.313184e-03	3.763320e-03
95%	96%	97%	98%	99%
4.312942e-03	5.237838e-03	6.127191e-03	7.375056e-03	1.073018e-02
100%				
4.926210e-02				

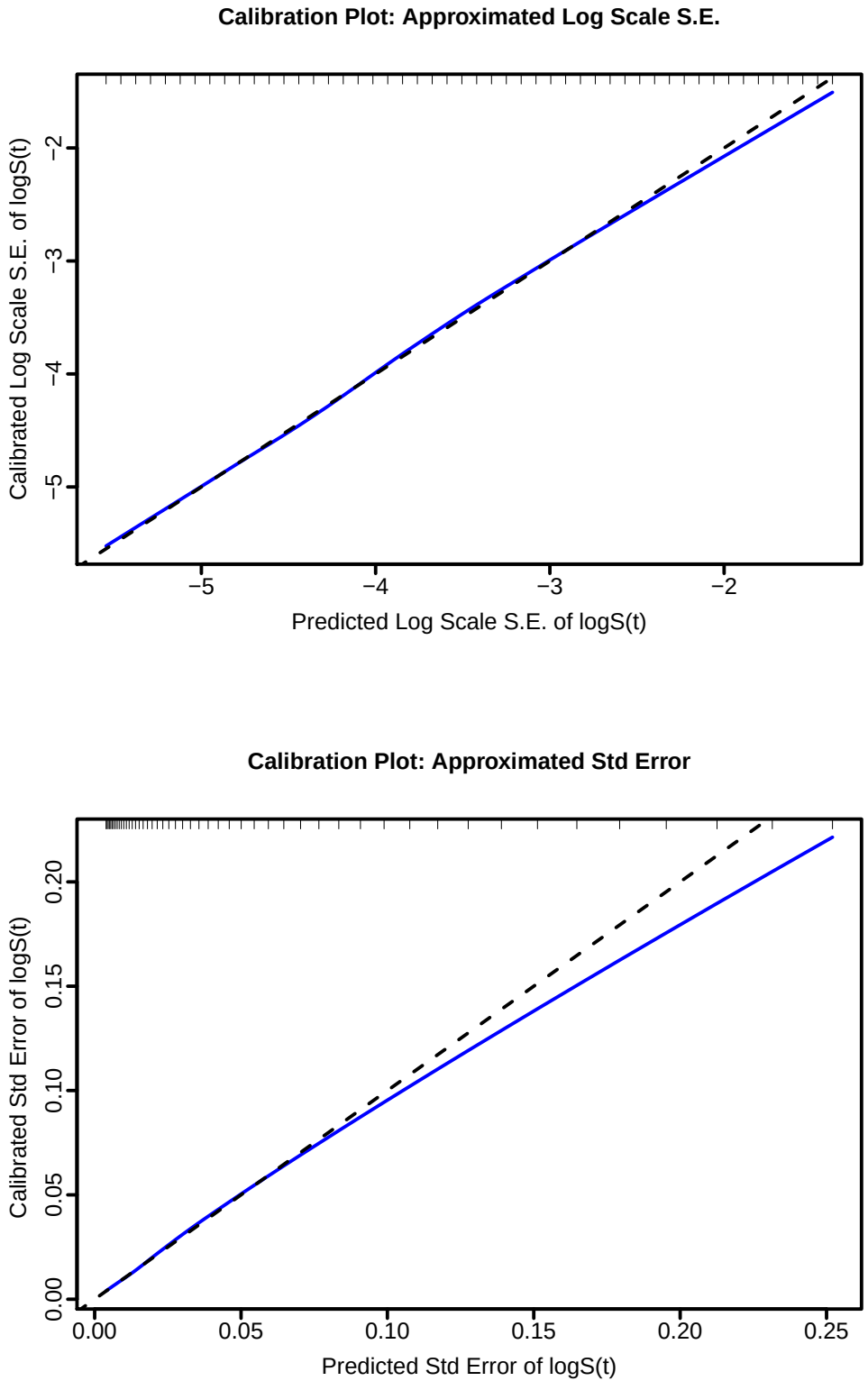


Figure 8: Calibration plot: prediction model for the standard errors (Weight model)

5 Limitations

1. As far as our records show, all patients were Warfarin, Statin, and Clopidogrel naive at the time medication home was established. This has two practical implications: 1) we do not make use of data for those with a history of these drugs, and 2) patients in our database might have been prescribed one of these medications at another institution (and so are not naive to the medications).

2. Medical home is established at one point in time in the database. However, individuals are being observed over time and so true risk is dynamic

3. The synthetic derivative was used to create the database. Are these patients different from the broader Vanderbilt population of patients (e.g. less healthy) and is there a difference in how we had coded data and how we intend to code data (e.g., diagnoses).

4. Secular trends in

   - Medication prescribing practices. We used data as far back as 2005. These trends could be caused by new warnings, and other reasons, including implementation of PREDICT.
   - The Vanderbilt patient population (size and makeup).
   - Secular trends in how we records diagnoses (e.g., how complete is it?)

5. People will drop out: People move or use other facilities. We are predicting an X year probability of being put on a medication, but it will not necessarily happen at Vanderbilt.

6. People will drop in: It may be that people get prescribed one of these medications prior to having medication home established. (insufficient opportunity to genotype).

7. We use one model for 3 drugs which may not be as good as three separate models and then putting results together (this would be difficult)

8. The accuracy of the phenotype elements (inaccurate recordings of height and weight, diagnoses, and the requirement of >=2 codes for most means that we will lose some sensitivity and also do not get them the first time they are diagnosed)

(Feel free to add...)