

## **SUPPLEMENTAL DATA**

# **Identifying patients at risk of prolonged hospital length of stay after total knee arthroplasty: a real-world study on the creation and validation of a cloud estimator**

**Table of contents.**

<b>Item</b>	<b>Description</b>	<b>Page</b>
Table S1	Details of complete and imputation datasets.	3-4
Table S2	Baseline characteristics of training and validation sets.	5-6
Table S3	Baseline characteristics of patients with TKA in the training cohort.	7-8
Table S4	Multivariable logistic regression results based on variables screened by BSR, LASSO, and BWS methods.	9
Table S5	Results of the final quasi-Poisson model using the backward stepwise analysis.	10
Figure S1	Association between continuous predictors and hospital prolonged LOS using RCS analysis.	11
Figure S2	Variable selection with the BSR and LASSO methods.	12
Figure S3	Bias plots for variables in the prediction model.	13
Figure S4	Snapshots of a dynamic web-based calculator for estimating prolonged LOS.	14

**Table S1. Details of complete and imputation datasets.**

Characteristics	Missing N (%)	Complete (N = 2246)	Imputation (N = 2600)	P Value*
Age, years, median (IQR)	NA	66 (61, 72)	66 (61, 72)	0.771
Race, n (%)	4 (0.2)			0.980
Chinese		1887 (84.0)	2192 (84.3)	
Malay		166 (7.4)	187 (7.2)	
Indian		132 (5.9)	148 (5.7)	
Others		61 (2.7)	73 (2.8)	
Gender, n (%)	NA			
Male		544 (24.2)	623 (24.0)	0.833
Female		1702 (75.8)	1977 (76.0)	
BMI, kg/m <sup>2</sup> , median (IQR)	66 (2.5)	27.3 (24.4, 30.5)	27.2 (24.5, 30.4)	0.886
Smoking, n (%)	NA			0.656
No		2041 (90.9)	2353 (90.5)	
Yes		205 (9.1)	247 (9.5)	
DM, n (%)	NA			0.527
No		1818 (80.9)	2123 (81.7)	
Yes		428 (19.1)	477 (18.3)	
IHD, n (%)	NA			0.275
No		2116 (94.2)	2468 (94.9)	
Yes		130 (5.8)	132 (5.1)	
CHF, n (%)	NA			0.731
No		2224 (99.0)	2577 (99.1)	
Yes		22 (1.0)	23 (0.9)	
CVA, n (%)	NA			0.618
No		2201 (98.0)	2553 (98.2)	
Yes		45 (2.0)	47 (1.8)	
Hb, g/dL, median (IQR)	NA	13.1 (12.3, 14.0)	13.1 (12.3, 14.0)	0.910
Creatinine, mg/dL, n (%)	292			0.766
≤ 2	(11.2)	2227 (99.2)	2580 (99.2)	
> 2		19 (0.9)	20 (0.8)	
ASA status, n (%)	NA			0.792
1 or 2		2111 (94.0)	2439 (93.8)	
3		135 (6.0)	161 (6.2)	
Type of anaesthesia, n (%)	NA			0.799
General		826 (36.8)	947 (36.4)	
Regional		1420 (63.2)	1653 (63.6)	
Procedure site, n (%)	NA			0.525
Unilateral		2079 (92.6)	2394 (92.1)	
Bilateral		167 (7.4)	206 (7.9)	

OD, mins, median (IQR)      NA      80 (65, 100)      80 (65, 100)      0.834

---

Table S1. Continued.

Characteristics	Missing N (%)	Complete (N = 2246)	Imputation (N = 2600)	<i>P</i> Value*
LOS, days, n (%)	NA			0.292
≤ 6		1801 (80.2)	2053 (79.0)	
> 6		445 (19.8)	547 (21.0)	

---

\**P* values between groups were assessed by the Chi-square and Mann-Whitney test.

IQR: Interquartile range; NA, not applicable; BMI: Body mass index; DM: Diabetes mellitus; IHD: Ischaemic heart disease; CHF: Congestive heart failure; CVA: Cerebrovascular accident; Hb: Haemoglobin; ASA: American Society of Anesthesiologist; OD: Operation duration.

**Table S2. Baseline characteristics of training and validation sets.**

Characteristics	Training Set (N = 2081)	Validation Set (N = 519)	P Value*
Age, years, median (IQR)	66 (61, 72)	66 (60, 72)	0.328
Race, n (%)			0.947
Chinese	1755 (84.3)	437 (84.2)	
Malay	147 (7.1)	40 (7.7)	
Indian	120 (5.8)	28 (5.4)	
Others	59 (2.8)	14 (2.7)	
Gender, n (%)			0.234
Male	509 (24.5)	114 (22.0)	
Female	1572 (75.5)	405 (78.0)	
BMI, kg/m <sup>2</sup> , median (IQR)	27.2 (24.4, 30.4)	27.3 (24.6, 30.3)	0.697
Smoking, n (%)			0.652
No	1886 (90.6)	467 (90.0)	
Yes	195 (9.4)	52 (10.0)	
DM, n (%)			0.061
No	1714 (82.4)	409 (78.8)	
Yes	367 (17.6)	110 (21.2)	
IHD, n (%)			0.232
No	1970 (94.7)	498 (96.0)	
Yes	111 (5.3)	21 (4.1)	
CHF, n (%)			0.291
No	2060 (99.0)	517 (99.6)	
Yes	21 (1.0)	2 (0.4)	
CVA, n (%)			0.888
No	2043 (98.2)	510 (98.3)	
Yes	38 (1.8)	9 (1.7)	
Hb, g/dL, median (IQR)	13.1 (12.3, 14.0)	13.1 (12.3, 13.9)	0.760
Creatinine, mg/dL, n (%)			0.154
≤ 2	2062 (99.1)	518 (99.8)	
> 2	19 (0.9)	1 (0.2)	
ASA status, n (%)			0.322
1 or 2	1957(94.0)	482 (93.0)	
3	124 (6.0)	37 (7.1)	
Type of anaesthesia, n (%)			0.186
General	745 (35.8)	202 (38.9)	
Regional	1336 (64.2)	317 (61.1)	
Procedure site, n (%)			0.601
Unilateral	1919 (92.2)	475 (91.5)	
Bilateral	162 (7.8)	44 (8.5)	

---

OD, mins, median (IQR)	80 (65, 100)	80 (65, 95)	0.765
------------------------	--------------	-------------	-------

---

\**P* values between groups were assessed by the Chi-square, Fisher's exact and Mann-Whitney tests.

IQR: Interquartile range; BMI: Body mass index; DM: Diabetes mellitus; IHD: Ischaemic heart disease; CHF: Congestive heart failure; CVA: Cerebrovascular accident; Hb: Haemoglobin; ASA: American Society of Anesthesiologist; OD: Operation duration.

**Table S3. Baseline characteristics of patients with TKA in the training cohort.**

Characteristics	Normal LOS (N = 1643)	Prolonged LOS (N = 438)	P Value*
Age, years, median (IQR)	66 (60, 72)	68 (63, 74)	< <b>0.001</b>
Race, n (%)			<b>0.041</b>
Chinese	1382 (84.1)	373 (85.2)	
Malay	127 (7.7)	20 (4.6)	
Indian	93 (5.7)	27 (6.2)	
Others	41 (2.5)	18 (4.1)	
Gender, n (%)			0.372
Male	409 (24.9)	100 (22.8)	
Female	1234 (75.1)	338 (77.2)	
BMI, kg/m <sup>2</sup> , median (IQR)	27.3 (24.5, 30.4)	27.1 (24.2, 30.4)	0.330
Smoking, n (%)			0.994
No	1489 (90.6)	397 (90.6)	
Yes	154 (9.4)	41 (9.4)	
DM, n (%)			<b>0.026</b>
No	1369 (83.3)	345 (78.8)	
Yes	274 (16.7)	93 (21.2)	
IHD, n (%)			<b>0.011</b>
No	1566 (95.3)	404 (92.2)	
Yes	77 (4.7)	34 (7.8)	
CHF, n (%)			0.062
No	1630 (99.2)	430 (98.2)	
Yes	13 (0.8)	8 (1.8)	
CVA, n (%)			< <b>0.001</b>
No	1622 (98.7)	421 (96.1)	
Yes	21 (1.3)	17 (3.9)	
Hb, g/dL, median (IQR)	13.2 (12.4, 14.0)	12.8 (11.8, 13.8)	< <b>0.001</b>
Hb, g/dL, n (%) <sup>†</sup>			< <b>0.001</b>
< 13	683 (41.6)	234 (53.4)	
≥ 13	960 (58.4)	204 (46.6)	
Creatinine, mg/dL, n (%)			<b>0.002</b>
≤ 2	1634 (99.5)	428 (97.7)	
> 2	9 (0.6)	10 (2.3)	
ASA status, n (%)			< <b>0.001</b>
1 or 2	1567 (95.4)	390 (89.0)	
3	76 (4.6)	48 (11.0)	
Type of anaesthesia, n (%)			<b>0.002</b>
General	560 (34.1)	185 (42.2)	
Regional	1083 (65.9)	253 (57.8)	

Table S3. Continued.

Characteristics	Normal LOS (N = 1643)	Prolonged LOS (N = 438)	P Value*
Procedure site, n (%)			<b>&lt; 0.001</b>
Unilateral	1551 (94.4)	368 (84.0)	
Bilateral	92 (5.6)	70(16.0)	
OD, mins, median (IQR)	80 (65, 95)	85 (70, 105)	<b>&lt; 0.001</b>

\*P values between groups were assessed by the Chi-square, Fisher's exact and Mann-Whitney tests. Bold values indicate statistical significance.

† The cut-off value was determined by the restricted cubic spline model and adjusted for variables including age, race, DM, IHD, CVA, creatinine, ASA status, type of anaesthesia, procedure site, and OD.

TKA: Total knee arthroplasty; IQR: Interquartile range; LOS: Length of stay; BMI: Body mass index; DM: Diabetes mellitus; IHD: Ischaemic heart disease; CHF: Congestive heart failure; CVA: Cerebrovascular accident; Hb: Haemoglobin; ASA: American Society of Anesthesiologist; OD: Operation duration.

**Table S4. Multivariable logistic regression results based on variables screened by BSR, LASSO, and BWS methods.**

	BWS		BSR		LASSO	
	Coefficient	OR	Coefficient	OR	Coefficient	OR
Intercept	-4.38	0.01	-4.40	0.01	-4.45	0.01
Age, years	0.04	1.04	0.04	1.04	0.04	1.04
Race						
Chinese	Ref.	Ref.				
Malay	-0.44	0.60	NA	NA	NA	NA
Indian	0.16	1.28				
Others	0.53	1.59				
DM						
No	Ref.	Ref.	NA	NA	NA	NA
Yes	0.26	1.50				
IHD						
No	NA	NA	NA	NA	NA	NA
Yes						
CVA						
No	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Yes	1.08	2.83	1.09	2.99	1.08	2.95
Hb, g/dL						
< 13	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
≥ 13	-0.39	0.64	-0.41	0.66	-0.40	0.67
Creatinine, mg/dL						
≤ 2	Ref.	Ref.	NA	NA	NA	NA
> 2	1.15	2.85				
ASA status						
1 or 2	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
3	0.72	1.62	0.78	2.19	0.80	2.22
Type of anesthesia						
General	Ref.	Ref.	Ref.	Ref.	NA	NA
Regional	-0.37	0.64	-0.37	0.69		
Procedure site						
Unilateral	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Bilateral	1.08	2.34	1.04	2.82	1.08	2.95
OD, mins	0.01	1.01	0.01	1.01	0.01	1.01

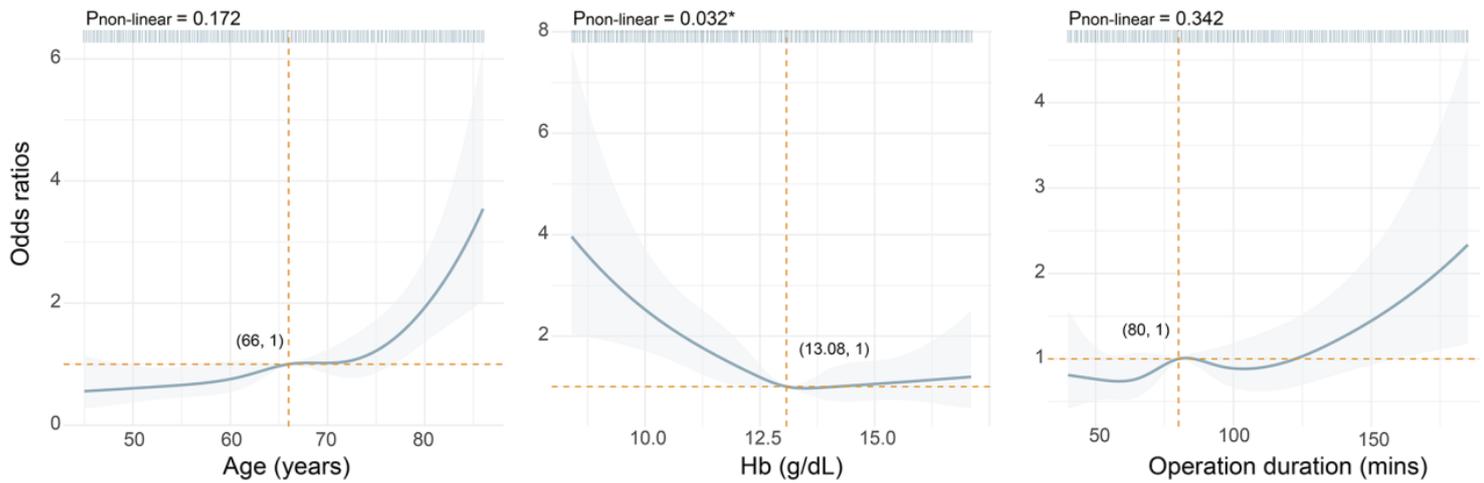
**Note:** Fields highlighted in green correspond to variables included in the model.

BSR: Best subset regression; LASSO: Least absolute shrinkage and selection operator; BWS: Backward stepwise; DM: Diabetes mellitus; IHD: Ischaemic heart disease; CVA: Cerebrovascular accident; Hb: Haemoglobin; ASA: American Society of Anesthesiologist; OD: Operation duration; OR: Odds ratio; NA, not applicable.

**Table S5. Results of the final quasi-Poisson model using the backward stepwise analysis.**

	Coefficient	RR (95% CI)	P Value
Intercept	0.62	1.87 (1.36, 2.57)	<0.001
Age, years	0.01	1.01 (1.01, 1.02)	<0.001
Race			
Chinese	Ref.	Ref.	
Malay	-0.08	0.92 (0.80, 1.06)	0.257
Indian	-0.03	0.97 (0.83, 1.12)	0.653
Others	0.09	1.09 (0.89, 1.32)	0.371
DM			
No	Ref.	Ref.	
Yes	0.82	1.09 (0.99, 1.18)	0.062
IHD			
No	Ref.	Ref.	
Yes	0.09	1.09 (0.95, 1.25)	0.201
Hb, g/dL			
< 13	Ref.	Ref.	
≥ 13	-0.10	0.91 (0.85, 0.97)	<b>0.006</b>
Creatinine, mg/dL			
≤ 2	Ref.	Ref.	
> 2	0.51	1.67 (1.28, 2.13)	<0.001
ASA status			
1 or 2	Ref.	Ref.	
3	0.40	1.49 (1.32, 1.67)	<0.001
Type of anaesthesia			
General	Ref.	Ref.	
Regional	-0.16	0.85 (0.79, 0.92)	<0.001
Procedure site			
Unilateral	Ref.	Ref.	
Bilateral	0.18	1.20 (1.05, 1.37)	<b>0.006</b>
OD, mins	0.003	1.00 (1.00, 1.00)	<0.001

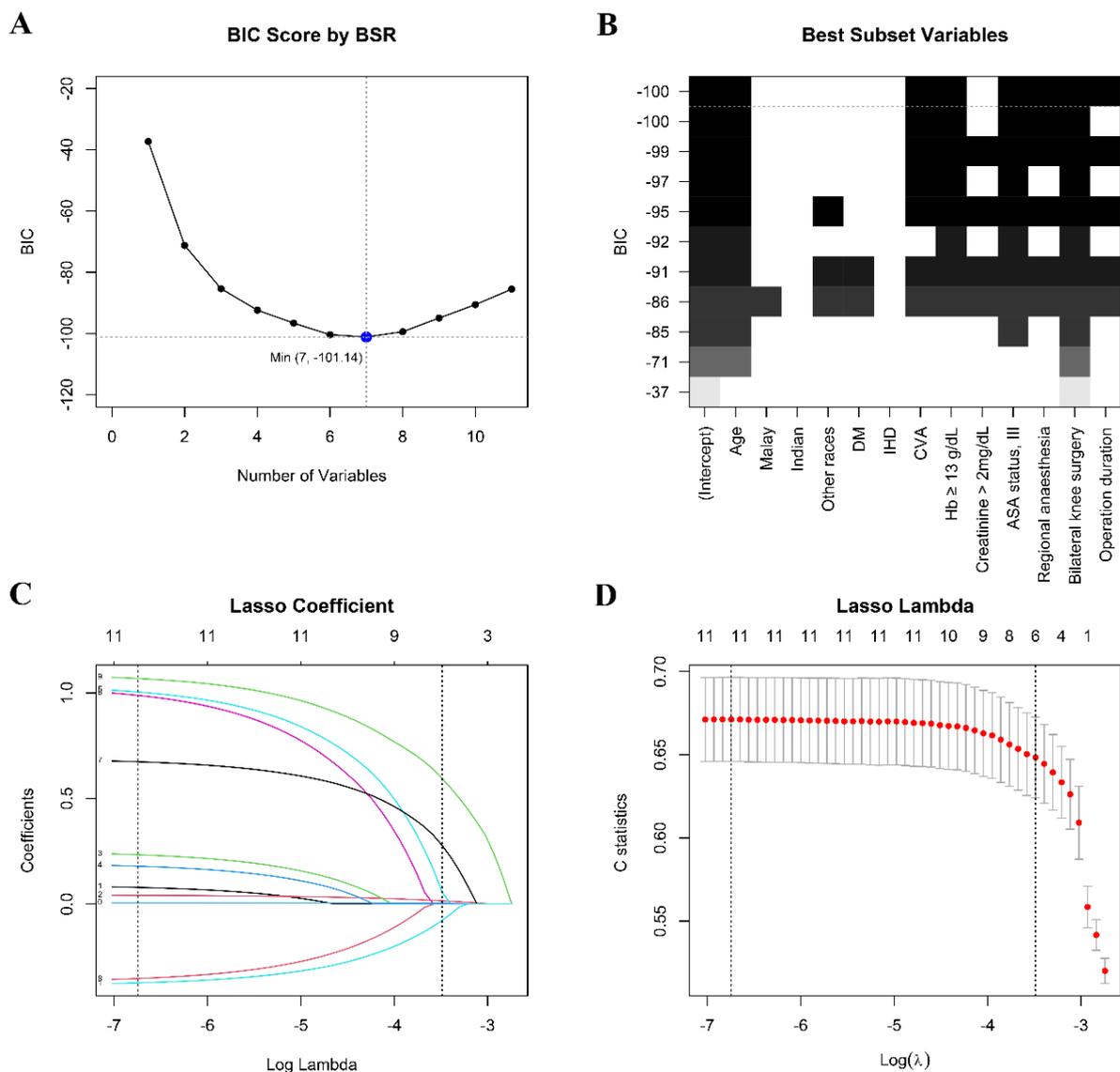
RR: Risk ratio; CI: Confidence interval; DM: Diabetes mellitus; IHD: Ischaemic heart disease; Hb: Haemoglobin; ASA: American Society of Anesthesiologist; OD: Operation duration.



**Figure S1.**

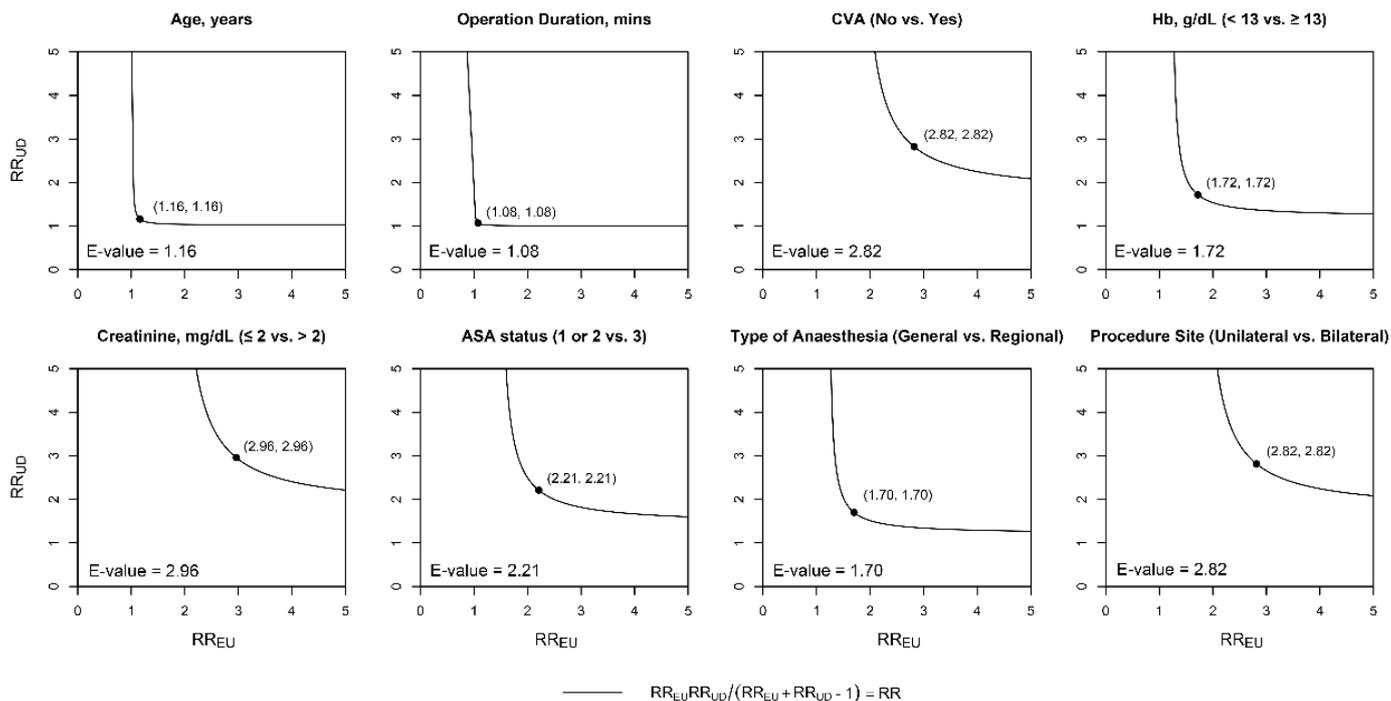
Association between continuous predictors and hospital prolonged LOS using RCS analysis. For each curve, five knots at the 5th, 35th, 50th, 65th, and 95th centiles were selected. Odds ratios are indicated by solid lines and 95% CIs by shaded areas. The reference point is the first value closest to or equal to the odds ratio at one (i.e., the intersection of the two yellow dashed lines).

LOS: Length of stay; RCS: Restricted cubic spline; Hb: Hemoglobin; CI: Confidence interval.



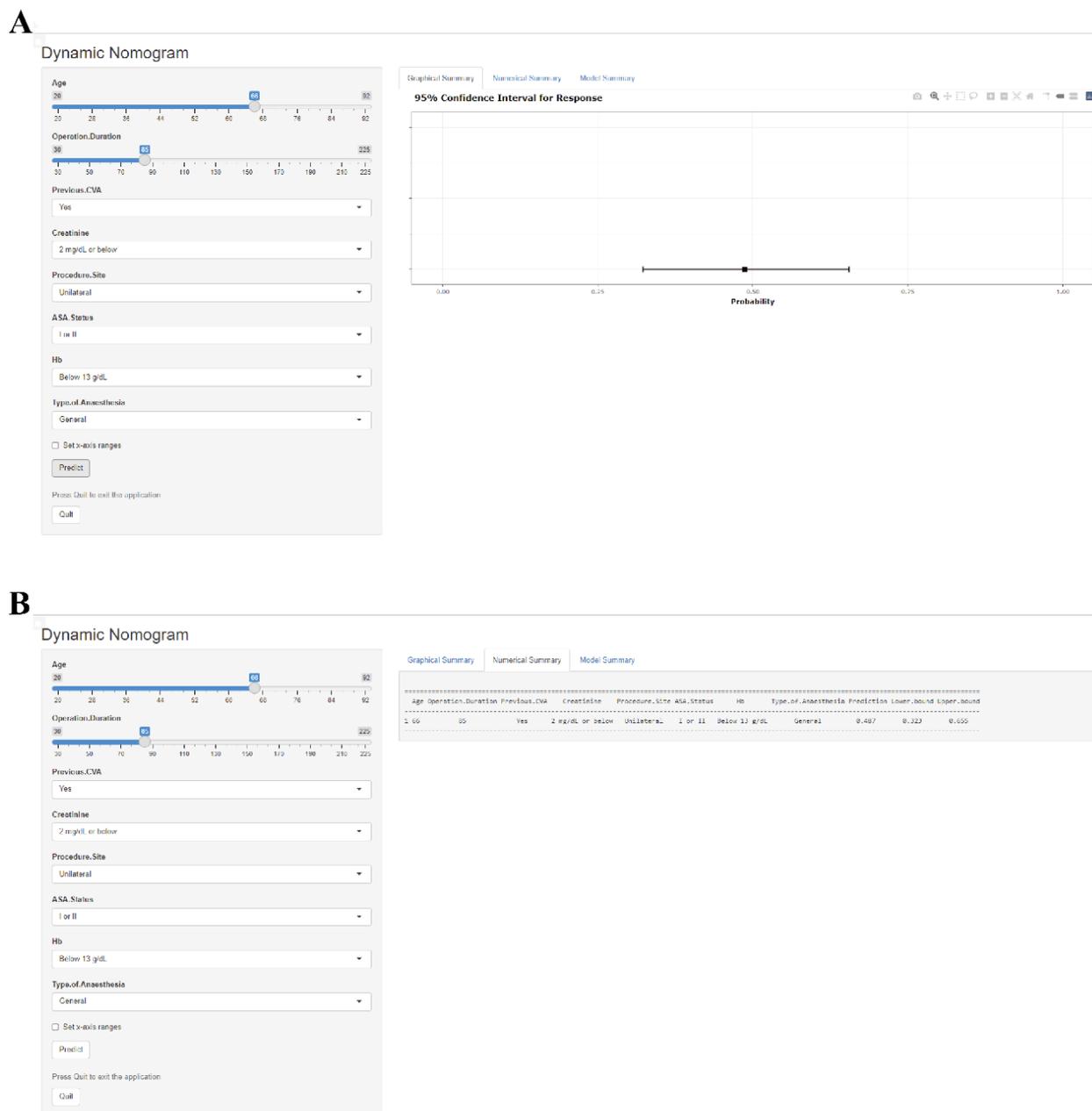
**Figure S2.** Variable selection with the BSR and LASSO methods. (A) The blue dot represents the number of variables optimally selected by the BSR, which is 7 as determined by the smallest BIC value of -101.14. (B) Small black squares above the dotted line indicate variables that were finally picked using BSR. (C) The LASSO coefficient profile of 11 significant variables was screened by univariate logistic regression analysis in the training cohort. Colored lines represent different variables. The coefficient of each variable tends to be zero as the corresponding penalty parameter lambda ( $\lambda$ ) increases, thus reducing the number of nonzero coefficient variables. (D) Ten-fold cross-validation (CV) for tuning parameter ( $\lambda$ ) selection. Two dashed vertical lines refer to two  $\lambda$  values that differ by one standard error, with the smaller one corresponding to the maximum C-statistic.

BIC: Bayesian information criterion; BSR: Best subsets regression; LASSO: Least absolute shrinkage and selection operator; DM: Diabetes mellitus; IHD: Ischaemic heart disease; CVA: Cerebrovascular accident; Hb: Haemoglobin; ASA: American Society of Anesthesiologist.



**Figure S3.** Bias plots for variables in the prediction model. RREU is the risk ratio for the relationship between measured variables (i.e., exposure variables) and unmeasured confounders, and RRUD for the relationship of unmeasured confounders to the outcome. The E-value is a continuous measure of how robust the observed association is to potential uncontrolled confounders. The lowest possible E-value is 1 (i.e., no unmeasured confounding is needed to explain away the observed association). The higher the E-value, the stronger the confounder-outcome association would have to be to explain away the effect.

CVA: Cerebrovascular accident; Hb: Haemoglobin; ASA: American Society of Anesthesiologist.



**Figure S4.** Snapshots of a dynamic web-based calculator for estimating prolonged LOS (<https://songandwen.shinyapps.io/PredictLOS/>). (A) Homepage. Continuous variable values can be chosen by using the slider, and categories corresponding to categorical variables can be selected by pulling down the drop-down checkbox. Then click on the "Predict" button and predicted probabilities with confidence intervals will be presented in the graph. (B) We can check out the results by switching to the "Numeric Summary" column.

LOS: Length of stay; TKA: Total knee arthroplasty; CVA: Cerebrovascular accident; ASA: American Society of Anesthesiologist; Hb: Haemoglobin.