

SUPPLEMENTARY MATERIAL

*Table S1: Summary of posterior distributions (based on 1000 samples) of the anticipated CPM performance, degradation and instability in a large evaluation dataset, for CPMs developed with particular modelling approaches using a sample size of 335 participants (~241 events) and 10 predictor parameters. Degradation is examined relative to the performance of the reference model shown in **Error! Reference source not found.**, which has a slope of 1, c-statistic of 0.76, and net benefit of 0.41 in the target population. Each shaded row gives results for an approximation (to improve computational speed) to the fully simulation-based approach of the previous row. ERVSI was very similar for the winning strategy.*

CPM development approach	Sample size approach	Error and uncertainty of predictions from CPM		Calibration, discrimination and clinical utility of CPM		
		MAPE: mean (95% range)	95% interval width: mean (95% range)	Calibration slope: mean (95% range) P(0.9<slope<1.1) P(0.85<slope<1.15)	C-statistic: mean (95% range) mean degradation (95% range)	Net benefit: mean (95% range) REVSI (95% RVSI range) P(RVSI>90%)
Unpenalised logistic regression (frequentist)	Fully simulation-based	0.06 (0.036 to 0.087)	0.29 (0.06 to 0.53)	0.85 (0.63 to 1.14) 0.26 0.43	0.74 (0.72 to 0.75) -0.02 (-0.04 to -0.01)	0.40 (0.38 to 0.40) 97.1% (94.1% to 99.3%) 1.0
Unpenalised logistic regression (frequentist)	Frequentist approximation via Fisher's information decomposition	0.059 (0.036 to 0.086)	0.28 (0.08 to 0.50)	0.88 (0.67 to 1.15) 0.31 0.51	0.74 (0.72 to 0.75) -0.020 (-0.04 to -0.01)	0.39 (0.38 to 0.40) 96.7% (93.4% to 98.7%) 1.0
Unpenalised logistic regression + heuristic shrinkage	Fully simulation-based	0.058 (0.033 to 0.085)	0.28 (0.09 to 0.48)	0.98 (0.69 to 1.43) 0.40 0.56	0.74 (0.72 to 0.75) -0.020 (-0.04 to -0.01)	0.39 (0.38 to 0.40) 96.7% (93.4% to 98.8%) 1.0
Ridge logistic regression (frequentist)	Fully simulation-based	0.056 (0.032 to 0.081)	0.24 (0.09 to 0.42)	1.12 (0.78 to 1.66) 0.36 0.52	0.74 (0.72 to 0.75) -0.02 (-0.04 to -0.01)	0.39 (0.38 to 0.40) 96.9% (93.9% to 98.9%) 1.0
Ridge logistic regression (Bayesian)	Fully simulation-based	0.058 (0.035 to 0.087)	0.27 (0.08 to 0.49)	0.99 (0.69 to 1.39) 0.44 0.60	0.74 (0.72 to 0.75) -0.021 (-0.04 to -0.01)	0.39 (0.38 to 0.40) 96.7% (93.1% to 98.7%) 1.0
Ridge logistic regression (Bayesian)	Bayesian approximation via Fisher's information decomposition	0.066 (0.036 to 0.096)	0.26 (0.12 to 0.47)	1.04 (0.78 to 1.47) 0.49 0.68	0.73 (0.70 to 0.75) -0.028 (-0.06 to -0.01)	0.39 (0.38 to 0.40) 96.0% (92.4% to 98.6%) 1.0
Lasso logistic regression (frequentist)	Fully simulation-based	0.059 (0.036 to 0.088)	0.27 (0.09 to 0.48)	1.00 (0.69 to 1.45) 0.44 0.63	0.74 (0.71 to 0.75) -0.02 (-0.04 to -0.01)	0.39 (0.38 to 0.40) 96.6% (93.4% to 98.6%) 1.0
Lasso logistic regression (Bayesian)	Fully simulation-based	0.061 (0.037 to 0.091)	0.30 (0.07 to 0.54)	0.85 (0.60 to 1.16) 0.30 0.46	0.74 (0.71 to 0.75) -0.022 (-0.04 to -0.01)	0.39 (0.38 to 0.40) 96.5% (92.9% to 98.7%) 1.0
Lasso logistic regression (Bayesian)	Bayesian approximation via Fisher's information decomposition	0.059 (0.035 to 0.089)	0.28 (0.09 to 0.50)	0.94 (0.73 to 1.25) 0.49 0.67	0.74 (0.71 to 0.75) -0.022 (-0.05 to -0.01)	0.39 (0.38 to 0.40) 96.5% (92.8% to 98.7%) 1.0
Random forest (100 trees, depth 3)	Fully simulation-based	0.092 (0.079 to 0.108)	0.23 (0.16 to 0.30)	1.53 (1.16 to 1.97) 0.01 0.02	0.72 (0.70 to 0.73) -0.04 (-0.06 to -0.03)	0.38 (0.36 to 0.39) 93.0% (89.7% to 95.7%) 0.95
Random forest (100 trees, depth 15)	Fully simulation-based	0.115 (0.100 to 0.133)	0.52 (0.21 to 0.71)	0.62 (0.53 to 0.71) 0 0	0.69 (0.66 to 0.71) -0.07 (-0.10 to -0.06)	0.37 (0.35 to 0.38) 90.4% (86.5% to 93.5%) 0.64

Table S2: Summary of posterior distributions (based on 1000 samples) of the anticipated CPM performance, degradation and instability in a large evaluation dataset, for CPMs developed with particular modelling approaches using a sample size of 75 participants (~51 events) and 20 candidate predictor parameters (including 10 noise predictor parameters). Degradation is examined relative to the performance of the reference model shown in **Error! Reference source not found.**, which has a slope of 1, c-statistic of 0.76, and net benefit of 0.41 in the target population. Each shaded row gives results for an approximation (to improve computational speed) to the fully simulation-based approach of the previous row. ERVSI was very similar for the winning strategy.

CPM development approach	Sample size approach	Error and uncertainty of predictions from CPM		Calibration, discrimination and clinical utility of CPM		
		MAPE: mean (95% range)	95% interval width: mean (95% range)	Calibration slope: mean (95% range) P(0.9<slope<1.1) P(0.85<slope<1.15)	C-statistic: mean (95% range) mean degradation (95% range)	Net benefit: mean (95% range) REVSI (95% RVSI range) P(RVSI>90%)
Unpenalised logistic regression (frequentist)	Fully simulation-based	0.24 (0.16 to 0.36)	0.94 (0.56 to 1.0)	0.17 (0 to 0.38) 0 0	0.63 (0.54 to 0.70) -0.12 (-0.22 to -0.06)	0.32 (0.25 to 0.36) 78.3% (61.9% to 89.5%) 0.02
Unpenalised logistic regression (frequentist)	Frequentist approximation via Fisher's information decomposition	0.16 (0.12 to 0.22)	0.73 (0.28 to 0.91)	0.39 (0.19 to 0.56) 0 0	0.66 (0.58 to 0.72) -0.10 (-0.18 to -0.04)	0.34 (0.28 to 0.37) 83.4% (67.6% to 92.0%) 0.10
Unpenalised logistic regression + heuristic shrinkage	Fully simulation-based	0.18 (0.11 to 0.33)	0.84 (0.46 to 0.99)	0.41 (-2.19 to 3.95) 0.04 0.06	0.62 (0.37 to 0.70) -0.13 (-0.38 to -0.06)	0.34 (0.28 to 0.38) 83.8% (68.0% to 92.2%) 0.11
Ridge logistic regression (frequentist)	Fully simulation-based	0.13 (0.09 to 0.17)	0.45 (0.33 to 0.64)	2.21 (0 to 7.22) 0.14 0.20	0.67 (0.51 to 0.72) -0.09 (-0.25 to -0.04)	0.37 (0.33 to 0.38) 90.1% (81.9% to 94.1%) 0.52
Ridge logistic regression (Bayesian)	Fully simulation-based	0.19 (0.11 to 0.34)	0.86 (0.37 to 1.00)	0.41 (0.01 to 1.21) 0.04 0.05	0.65 (0.56 to 0.70) -0.11 (-0.20 to -0.05)	0.34 (0.28 to 0.37) 84.0% (69.5% to 92.0%) 0.14
Ridge logistic regression (Bayesian)	Bayesian approximation via Fisher's information decomposition	0.17 (0.12 to 0.22)	0.46 (0.36 to 0.61)	0.51 (-0.74 to 1.52) 0.11 0.15	0.57 (0.43 to 0.67) -0.19 (-0.33 to -0.08)	0.33 (0.22 to 0.37) 82.2% (54.2% to 90.6%) 0.05
Lasso logistic regression (frequentist)	Fully simulation-based	0.14 (0.09 to 0.19)	0.49 (0.36 to 0.69)	0.93 (0 to 3.42) 0.11 0.16	0.62 (0.50 to 0.71) -0.14 (-0.26 to -0.05)	0.36 (0.32 to 0.38) 88.8% (78.8% to 93.1%) 0.24
Lasso logistic regression (Bayesian)	Fully simulation-based	0.21 (0.12 to 0.34)	0.88 (0.36 to 1.00)	0.26 (0.02 to 0.57) 0 0	0.65 (0.57 to 0.71) -0.11 (-0.19 to -0.05)	0.34 (0.28 to 0.37) 82.5% (69.0% to 92.1%) 0.093
Lasso logistic regression (Bayesian)	Bayesian approximation via Fisher's information decomposition	0.16 (0.11 to 0.21)	0.69 (0.48 to 0.86)	0.43 (0.11 to 0.72) 0 0	0.64 (0.53 to 0.71) -0.12 (-0.23 to -0.05)	0.33 (0.26 to 0.38) 82.1% (64.0% to 92.3%) 0.105
Random forest (100 trees, depth 3)	Fully simulation-based	0.14 (0.11 to 0.17)	0.38 (0.30 to 0.43)	1.04 (0.54 to 1.51) 0.31 0.44	0.64 (0.57 to 0.69) -0.12 (-0.19 to -0.07)	0.36 (0.34 to 0.38) 89.4% (83.0% to 92.3%) 0.37
Random forest (100 trees, depth 15)	Fully simulation-based	0.14 (0.11 to 0.17)	0.47 (0.35 to 0.55)	0.72 (0.36 to 1.01) 0.1 0.21	0.64 (0.56 to 0.68) -0.13 (-0.20 to -0.07)	0.36 (0.32 to 0.37) 87.7% (78.4% to 92.0%) 0.25