SUPPLEMENTARY MATERIAL:

Supplementary Material A - Sample Size Calculation

Following (Riley 2019, Riley 2020), the sample size is considered from four perspectives, and the largest sample size calculated is selected as the overall sample size needed.

1. Approximate 95% confidence interval for overall outcome proportion in study population

$$n = \left(\frac{1.96}{\delta}\right)^2 \hat{\theta}(1-\hat{\theta})$$

$\hat{\theta} = .10 \ or \ .15$ - overall outcome proportion in study population
Then for:
$\hat{\theta} = .10$, n=139
$\hat{\theta} = .15$, n=196

2. Mean absolute prediction error (MAPE) - average error in the model's outcome

$$n = \exp\left(\frac{-0.508 + 0.259\ln(\theta) + 0.504\ln(P) - \ln(MAPE)}{0.544}\right)$$

MAPE=0.050 - suggested MAPE is no larger than 0.050 (lower values in settings may be appropriate where precise predictions are needed if consequences of wrong decisions are large)

P=18 - number of predictors

For $\hat{\theta} = .10$, then n=274
For $\hat{\theta} = .15$, then n=332

3. Achieve expected uniform shrinkage factor S

$$n = \frac{P}{(S-1)ln\left(1-\frac{R_{cs}^2}{S}\right)}$$

R_{cs}^2 =0.10 or 0.15 - proportion of overall variation

explained P=19 - number of predictors

S=0.9 or 0.85 - suggested target for shrinkage of ≤10% (i.e. S≥0.9)

For R^2_{cs} = 0.10, S = 0.9 , then n=1529

For R^2_{cs} = 0.15, S = 0.9 , then n=988

For R^2_{cs} = 0.10, S = 0.85 , then n=959

For R^2_{cs} = 0.15, S = 0.85 , then n=1529

4. Ensure a small expected optimism in apparent R^2

$$n = \frac{P}{(S-1)ln\left(1-\frac{R^2_{cs}}{S}\right)}$$

Where

$$S = \frac{R^2_{cs}}{R^2_{cs} + \delta \max\left(R^2_{cs}\right)}$$

$$\max(R^2_{cs}) = 1 - exp\left(\frac{2lnL_{null}}{n}\right)$$

$$lnL_{null} = Eln\left(\frac{E}{n}\right) + (n-E)ln\left(1-\frac{E}{n}\right)$$

and consider $\frac{E}{n} = \theta$

For $\hat{\theta}$ = .10, R^2_{cs} = 0.10 then max(R^2_{cs})=0.48, S=0.81 and n=719

For $\hat{\theta}$ = .10, R^2_{cs} = 0.15 then max(R^2_{cs})=0.48, S=0.81 and n=463

For $\hat{\theta}$ = .15, R^2_{cs} = 0.10 then max(R^2_{cs})=0.57, S=0.84 and n=888

For $\hat{\theta}$ = .15, R^2_{cs} = 0.15 then max(R^2_{cs})=0.57, S=0.84 and n=572

References

Riley RD et al, Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441 doi: 10.1136/bmj.m441.

Riley RD, Snell KI, Ensor J, etal . Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38:1276-96. 10.1002/sim.7992 30357870