

Supplementary Material - Determining number of covariates (parameters) for prediction model in ARCHIE

Based on R.Riley paper (*Statistics in Medicine*. 2019;38:1276–1296)

LR from logit model in ARCHIE dataset with only age as predictor: 0.76

From $n = 265$ (61 events, 204 no events)

Using this and the estimation from Riley 2019 Stats in Med:

$$R^2_{CS_app} = 1 - \exp(-LR/n) = 1 - \exp(-0.76/265) = 0.00286$$

Based on a shrinkage factor (as recommended by Riley 2019) of 0.9 we get the following values and table:

$$S_{VH} = 0.9,$$

$$R^2_{CS_adj} = S_{VH} R^2_{CS_app} = 0.9 (0.00286) = 0.00258,$$

and

$$n = p / \{ (S_{VH} - 1) \ln \{ 1 - (R^2_{CS_adj} / S_{VH}) \} \} = p / 0.00029 .$$

Which gives n (total sample) for a given p (number of parameters included in model):

n	p
3486.8	1.0
6973.7	2.0
10460.5	3.0
13947.4	4.0
17434.2	5.0
20921.1	6.0
