Additional file 1

The value of pre-season screening for injury prediction: The development and internal validation of a multivariable prognostic model to predict indirect muscle injury risk in elite football (soccer) players. Sports Medicine - Open.

Hughes, T., Riley, R.D. Sergeant, J.C., Callaghan, M.J. (2020)

Corresponding author: Tom Hughes

Email: tom.hughes.physio@manutd.co.uk

Correspondence address: Manchester United Football Club, AON Training Complex, Birch Road, Off Isherwood Road, Carrington, Manchester. UK. M31 4BH.
Tel: 0161 868 8754
Sample size calculation

Prognostic models suffer from optimism, where prediction performance is worse when applied in datasets that are not considered during development [1]. This is primarily due to overfitting, where the number of included candidate PFs is too large relative to the number of outcomes, or if irrelevant candidate PFs are included [2]. Because our sample size was fixed, to determine the maximum number of candidate PFs for inclusion, we used a minimum of 10 events per variable (EPV) rule, recommended to reduce statistical overfitting of logistic regression models [3]. Note that ‘variable’ means any parameter included (or considered for inclusion) that corresponded to a PF.

During the 5-seasons, 138 I-IMIs were recorded. We restricted the number of included parameters to 12, which corresponded to >10 EPV. This also met the criteria to minimise overfitting proposed by Riley et al [4]; our calculation assumed the model would have a modest Nagelkerke $R^2$ of 25%, so with an outcome proportion of 0.435, our 12 variables corresponded to 15% overfitting [4]. This was a suitable compromise between increasing the number of PF variables and minimising overfitting.

References


