Assessing impact on change in incidence on calibration performance across external validation

Ross D. Williams MSc (1), Jenna M. Reps PhD(2), Peter R. Runbeek PhD (1)

(1) Erasmus University Medical Centre, Rotterdam; (2) Janssen Research and Development, Raritan, NJ, USA

INTRODUCTION
Increasingly external validation is being seen as the gold standard in assessing prediction model performance[1]. One of the ways to measure the performance of a prediction model is to assess the calibration. This measures the agreement between predicted and observed risks. Calibration is essential to aiding decision making as using a poorly calibrated model would result in missing people who need intervention (if under-estimated risk) or giving an intervention unnecessarily (if over-estimating risk). A common issue that occurs when performing external validation is a worsening of calibration performance[2]. A possible reason for this could be due to the change in event rate between the development and validation environment. If this is the case it could be possible to correct some miscalibration by adjusting the model bias based upon the known differential event rates without retraining the model.

MATERIALS AND METHODS
We developed and externally validated models across a multitude of databases and problem settings. These databases included: CCAE, MDCH, MDOR, Optum claims and Optum EHR. The problems are specified elsewhere in two studies, one looking at hospitalization risk in covid-19 patients and another looking at predicting heart failure in type 2 diabetes patients [2,3]. All studies predict binary outcomes. We hypothesised that miscalibration is dependent on differential event rate. To test this, we compared the differential event rates (equation 1), to determine whether the model produced an over or underestimate of risk. We assessed this using the calibration-in-the-large and the intercept of the model. These are both metrics to assess calibration. Calibration-in-the-large checks the agreement between the mean predicted risk and the event rate, a value of 1 being optimal. The calibration intercept, obtained by fitting a linear model between the predicted and observed values, assesses whether the risks are over or underestimated. A value of zero being perfect and a negative value suggesting overestimation and positive value underestimation. A perfectly calibrated model would thus have a calibration-in-the-large of 1 and a calibration intercept of 0.

\[
\text{Incidence} = \frac{\text{event rate}_{\text{external database}} - \text{event rate}_{\text{development database}}}{\text{event rate}_{\text{development database}}} 
\]

(Eq. 1) 5 Incidence

Figure 1 shows a relationship between the differential event rates and the calibration statistics. Calibration in the large and Intersect were both correlated using Pearson’s test with a coefficient of 0.35 and 0.46 respectively.

CONCLUSION
The results show a relationship and suggest that using the differential event rate to create a correction factor for model recalibration is possible. The aim of future work is to extend it to include a correction factor based upon this relationship to provide a method of recalibration.

References