EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

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

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

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

Introduction

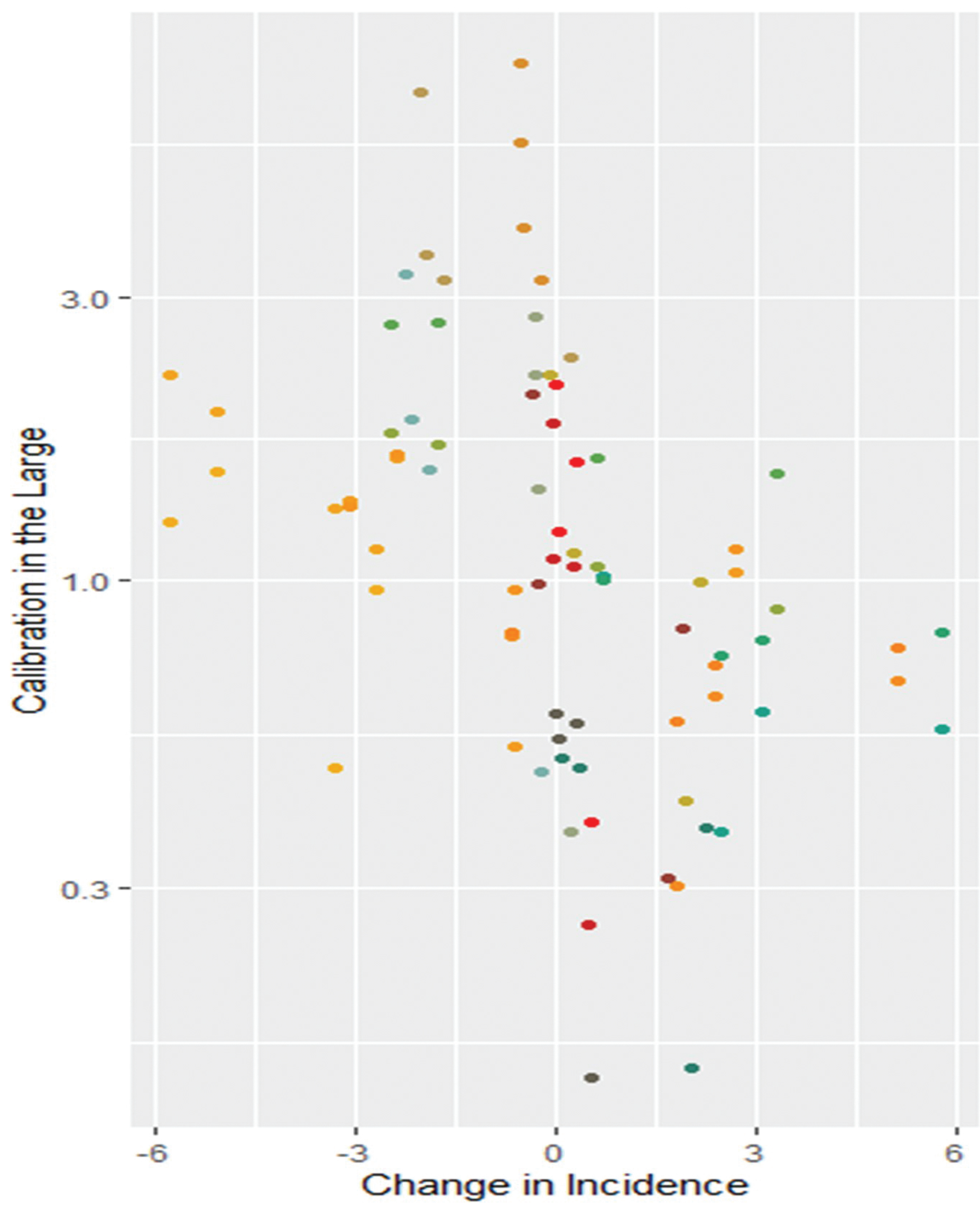
Increasingly external validation is being seen as the gold standard in assessing prediction model performance[1]. One of the ways to measure the performance of a prediction model is to assess the calibration. This measures the agreement between predicted and observed risks. Calibration is essential to 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, MDCD, MDCR, 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.

$$\text{(Eq. 1) } \delta\, \boldsymbol{Incidence} = \boldsymbol{event\ rate}_{external\ database} - \boldsymbol{event\ rate}_{development\ database}$$

Results



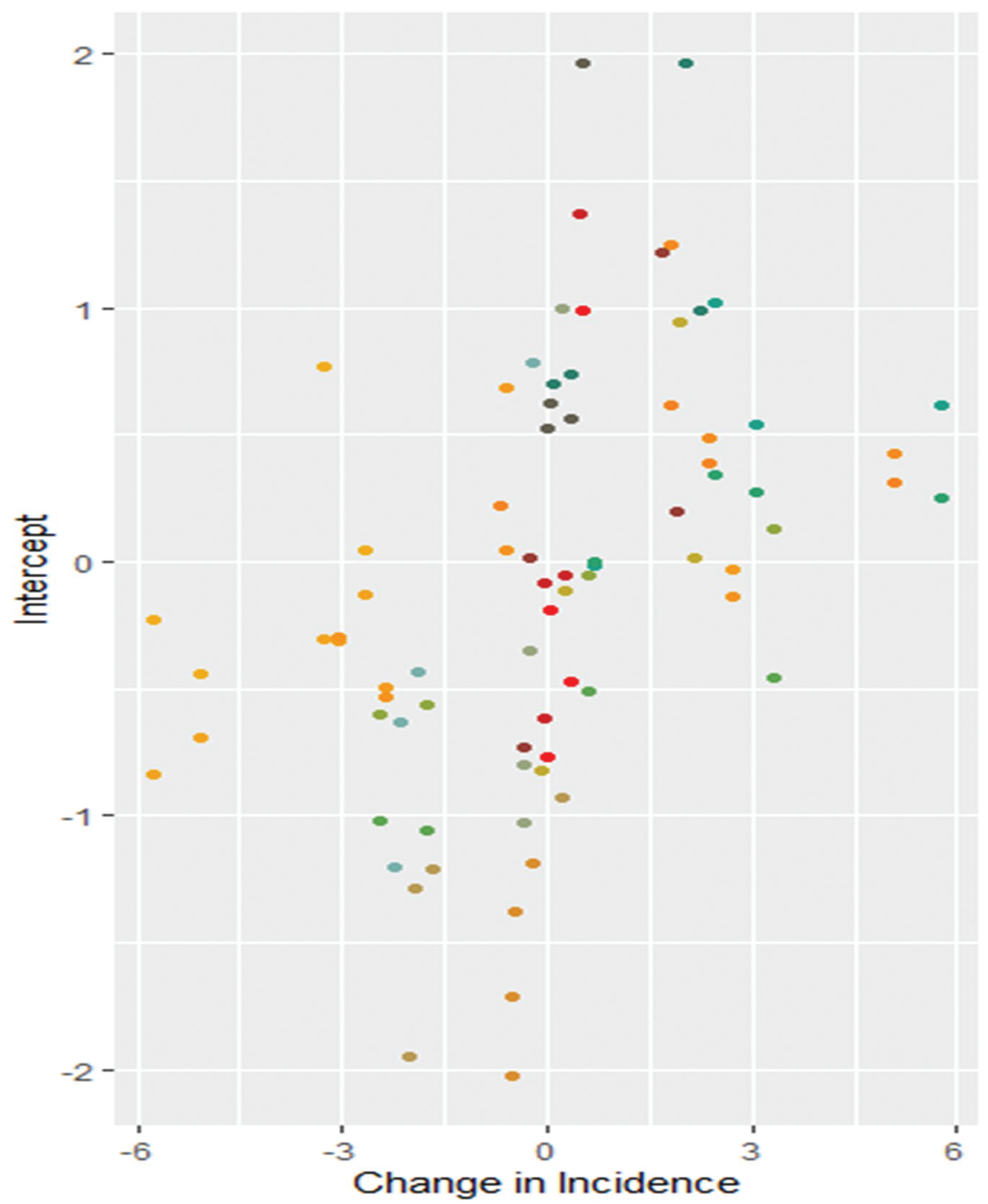


Figure 1 shows a relationship between the differential event rates and the calibration statistics. Calibration in the large and Intercept were both correlated using Pearsons test with a coefficient of 0.35 and 0.46 respectively.

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

1. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;69:245-247.
2. Williams RD, Markus AF, Yang C, et al. Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. *medRxiv.* 2020:2020.2005.2026.20112649.
3. Williams RD, Reps JM, Kors JA, et al. Using the OHDSI network to develop and externally validate a patient-level prediction model for Heart Failure in Type II Diabetes Mellitus. *medRxiv.* 2021:2021.2004.2006.21254966.

Erasmus MC

Email: r.williams@erasmusmc.nl

www.ehden.eu @IMI_EHDEN IMI_EHDEN github.com/EHDEN

imi innovative medicines initiative efpia

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.