Short summary

Background: The assessment of calibration performance of risk prediction models based on regression or more flexible machine learning algorithms receives little attention.

Main text: Herein, we argue that this needs to change immediately because poorly calibrated algorithms can be misleading and potentially harmful for clinical decision-making. We summarize how to avoid poor calibration at algorithm development and how to assess calibration at algorithm validation, emphasizing balance between model complexity and the available sample size. At external validation, calibration curves require sufficiently large samples. Algorithm updating should be considered for appropriate support of clinical practice.

Conclusion: Efforts are required to avoid poor calibration when developing prediction models, to evaluate calibration when validating models, and to update models when indicated. The ultimate aim is to optimize the utility of predictive analytics for shared decision-making and patient counseling.