

Simulation-based sample size calculations for studies externally validating a prediction model

Kym IE Snell¹, Lucinda Archer¹, Joie Ensor¹, Laura J Bonnett², Bob Phillips³, Gary S Collins⁴, Richard D Riley¹

¹ *Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Keele, Staffordshire, UK*

² *Department of Biostatistics, University of Liverpool, Liverpool, UK*

³ *Centre for Reviews and Dissemination, University of York, York, UK*

⁴ *Centre for Statistics in Medicine, University of Oxford, Oxford, UK*

Introduction

Sample size requirements for external validation of a prediction model are often based on 'rules-of-thumb' such as requiring at least 100 (or even 200) events or non-events. Although often overlooked, it is not simply the point estimates of performance measures that are of interest but also the precision in these estimates. Researchers should therefore ensure that validation studies are large enough to estimate performance measures with reasonable precision.

Objectives

To investigate factors affecting precision of performance measures, and demonstrate a simulation-based approach for determining appropriate sample sizes for external validation studies.

Methods

We conducted a simulation study to investigate the relationship between various factors (outcome prevalence, linear predictor distribution (LPSD), total sample size) and precision of performance measures for a logistic regression model, and developed a simulation-based approach to determine the minimum sample size required to achieve sufficiently narrow confidence intervals for all predictive performance measures of interest.

Results

The simulation study demonstrates that factors other than number of events affect precision of performance measures (including LPSD and total sample size) and that even with 100 or 200 events and non-events, 95% CIs remain wide in some settings. By specifying the desired precision of performance measures and distribution of the linear predictor (e.g. based on development data), our simulation-based approach can be used to tailor sample size calculations. The approach will be illustrated for designing a validation study for a diagnostic model for deep vein thrombosis, based on published data.

Conclusions

Sample size for validation of logistic models cannot be solved easily using closed form solutions and rules-of-thumb are often too simplistic and fail in individual settings. In situations where the distribution of the linear predictor can be ascertained, a simulation-based approach allows the sample size to be tailored to the setting.

Keywords

Sample size, validation, prediction, simulation