Letter to the Editor

Flawed validation of FRAX


Evaluating the performance of a prediction model in separate data from which the model was developed is important to judge its usefulness. The more external validation studies that are carried out showing good performance the more likely the model will be useful in untested settings. The recent study by Kim and colleagues sought to validate a South Korean version of the FRAX (without BMD) model in a moderately large South Korean cohort [1]. While this may seem entirely appropriate and indeed useful, there are a number of concerns, dare we say flaws, in the analysis and reporting that deserve highlighting so that future investigators do not replicate.

FRAX (version 3.9) is an opaque system of 62 multivariable models available for 57 countries, to predict the 10-year probability of major osteoporotic and hip fractures [2,3]. To evaluate the performance of FRAX requires a longitudinal design whereby individuals have their baseline information recorded (see www.shef.ac.uk/FRAX/) and then are followed up over time, ideally for 10-years (as this is the time horizon for which the FRAX model is predicting), or at least a sufficient number is followed up for 10 years. However, the study by Kim and colleagues, used a cross-sectional design (i.e. data collected from individuals at one specific point in time) and therefore evaluating the predictive performance of FRAX (often termed validation) in the true sense is not possible [4]. Comparing FRAX with and without BMD, as what the authors attempted to do might be of some use, but again without knowing which individuals went on to have a major osteoporotic or hip fracture, it is unclear on the value of this so-called validation. For example, if both models under- or over predict the 10-year probability of fracture, then correlating one against the other will not identify this.

It is widely accepted that external validation studies should assess both discrimination and calibration [5,6]. Discrimination is the ability of the model to differentiate between those with and without the fracture, quantified by the c-index, while calibration compares predictions from the model with observed outcomes, usually assessed graphically. FRAX was developed to account for competing risks (e.g. death) and therefore appropriate adjustment to traditional measures of discrimination and calibration needs to be made [7,8]. In the absence of an assessment of discrimination and calibration, then the study should not be perceived as an external validation study as it provides no meaningful information on how useful the models are.

We recommend to the authors and other investigators developing or validating a prediction model to consult to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) Statement [5]. In addition to providing guidance on key information to report when describing a prediction model study, the accompanying Explanation & Elaboration article also highlights numerous methodological aspects to consider when developing or validating a clinical prediction model [4].

Conflict of interest

Dr Collins is a member of the TRIPOD steering group which developed the TRIPOD Statement. No other disclosures were reported.

References


Gary S. Collins
Centre for Statistics in Medicine, Botnar Research Centre, University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom
Corresponding author.
E-mail address: gary.collins@csm.ox.ac.uk.

Yannick Le Manach
Department of Anesthesia & Clinical Epidemiology, Michael DeGroote School of Medicine, Faculty of Health Sciences, McMaster University, Perioperative Research Group, Population Health Research Institute, Hamilton, Canada
Department of Biostatistics, Michael DeGroote School of Medicine, Faculty of Health Sciences, McMaster University, Perioperative Research Group, Population Health Research Institute, Hamilton, Canada
E-mail address: yannick.lemanach@phri.ca.

3 April 2015