

TITLE: Flawed validation of FRAX

Gary S. Collins, *associate professor*

Centre for Statistics in Medicine, Botnar Research Centre,
University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom
Email: gary.collins@csm.ox.ac.uk

Yannick Le Manach, *assistant professor*

Departments of Anesthesia & Clinical Epidemiology and Biostatistics, Michael
DeGroote School of Medicine, Faculty of Health Sciences, McMaster University and
the Perioperative Research Group, Population Health Research Institute, Hamilton,
Canada
Email: yannick.lemanach@phri.ca

The authors report no conflict of interest.

Letter Re: Kim JW, Koh JM, Park JH, Chang SK. Validation of FRAX without BMD: An age-related analysis of the Fifth Korean National Health and Nutrition Examination Survey (KNHANES V-1, 2010). *Bone* 2015; 75: 27-31.

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]. Whilst 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 are 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 of 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 the 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, whilst calibration compares predictions from the model with observed outcomes, usually assessed graphically. FRAX was developed accounting for competing risks (e.g. death) and therefore appropriate adjustment to traditional measures of discrimination and calibration need 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 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 validation a clinical prediction model [4].

REFERENCES

- [1] Kim JW, Koh JM, Park JH, Chang JS. Validation of FRAX without BMD: An age-related analysis of the Fifth Korean National Health and Nutrition Examination Survey (KNHANES V-1, 2010). *Bone* 2015;75: 27-31.
- [2] Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, Burckhardt P, Cooper C, Christiansen C, Cummings S, Eisman JA, Fujiwara S, Gluer C, Goltzman D, Hans D, Krieg MA, La Croix A, McCloskey E, Mellstrom D, Melton LJ, 3rd, Pols H, Reeve J, Sanders K, Schott AM, Silman A, Torgerson D, van Staa T, Watts NB, Yoshimura N. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int* 2007;18: 1033-1046.
- [3] Collins GS, Michaëlsson K. Fracture risk assessment: state of the art, methodologically unsound, or poorly reported? *Curr Osteoporos Rep* 2013;10: 199-207.
- [4] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162: W1-W73.

1 [5] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of
2 a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD):
3 The TRIPOD Statement. Ann Intern Med 2015;162: 55-63.

4 [6] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models:
5 seven steps for development and an ABCD for validation. Eur Heart J 2014;35: 1925-
6 1931.

7
8
9 [7] Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction
10 models in the presence of competing risks. Statistics in Medicine 2014: n/a-n/a.

11
12 [8] Wolbers M, Blanche P, Koller MT, Witteman JC, Gerds TA. Concordance for
13 prognostic models with competing risks. Biostatistics 2014;15: 526-539.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65