

Pitfalls in the external validation of FRAX: response to Bolland et al.

J. A. Kanis · A. Oden · H. Johansson · E. McCloskey

Received: 13 March 2012 / Accepted: 13 March 2012 / Published online: 15 May 2012
© International Osteoporosis Foundation and National Osteoporosis Foundation 2012

Dear Editor,

As Charles Buxton once said, “Silence is sometimes the severest criticism” but we felt compelled to respond to the comments raised by Bolland and colleagues as, in many respects, they have missed the point of our review [1].

1. We have nothing against simpler models. We only state the obvious—that the optimal predictive model in any data set would be that derived from that data set. This should not be misinterpreted to mean that validation should not be performed in independent data sets, but only that the results be more critically interpreted.
2. Their letter criticises the unavailability of the risk coefficients used in FRAX. The model is not as opaque as it may seem as great detail is provided in the WHO study report [2] and related publications. The coefficients are published as well as the interactions used and much of the mathematical approach.
3. Bolland raises the point that we did not participate in the study of Collins et al. [3] that examined the performance characteristics of QFracture in the THIN data from general practices in the UK. Had Bolland et al. (and Collins et al. for that matter) read the literature that they cite in their paper, they would have appreciated that we had validated FRAX in this cohort several years earlier [4].

4. We do not deprecate the use of ROC analysis—only its misinterpretation—a criticism from which Bolland et al. are not immune. In the discussion of their paper [5], they make comparative statements on the areas under curve (AUCs) from their study and from our own work [4] without accounting for duration of follow-up and age. The error is compounded by the fact that the AUCs that we reported were not even AUCs for fracture probability.
5. Bolland et al. wish to deny a charge that they compared fracture incidence with fracture probability. Probability differs from cumulative incidence in that the former accounts for death as well as the fracture hazard. There is no need for the jury to retire, since the 248 deaths in their study were not considered.
6. Bolland et al. contest a view that the cohort studied was too small. The cohort was small but we never claimed it was too small. The real question is whether the cohort is representative of the New Zealand population which, by their own admission in the letter, they acknowledge to be biased by the use of inclusion/exclusion criteria. While some of this bias can be accounted for, other factors not captured by FRAX that influence fracture risk are also likely to differ.
7. The authors seem surprised by our view to expect some differences in cross calibration when using regional cohorts. This seems surprising given the overwhelming evidence for heterogeneity of fracture risk within countries [6]. If a cohort finds discordance in the number of expected and observed hip fractures, then it is possible that the model fails because of an error in the national statistics (supplied to us by one of the authors) and hence the calibration of FRAX or, in the case of regional samples, it fails because of bias or regional variations in age- and sex-specific fracture incidence or mortality risk.

This reply refers to the comment available at doi:10.1007/s00198-012-1977-y.

J. A. Kanis (✉) · A. Oden · H. Johansson · E. McCloskey
WHO Collaborating Centre for Metabolic Bone Diseases,
University of Sheffield Medical School,
Beech Hill Road,
Sheffield S10 2RX, UK
e-mail: w.j.pontefract@sheffield.ac.uk

If one concluded that FRAX was not calibrated for a particular study population, a question arises about the desirability of a regional model. The most robust data for fracture, and certainly mortality, lie at a national level. If reliable un-biased data for both exist, a more localised model could be developed. But why stop there? Why not a FRAX model for Auckland, for example? Perhaps a FRAX model for Queen Street, Auckland? The absurdity is obvious, but no less so than the expectation that any fracture prediction model calibrated nationally will work perfectly within any selected population.

It is ironic that the vast majority of the points raised in our review were discussed in correspondence with Mark Bolland before submission of his paper. Indeed we declined co-authorship of the paper for this reason. This, and similar experience with several others, was the stimulus to write our review so that others do not ignore issues that we had hitherto only managed to express privately.

References

1. Kanis JA, Oden A, Johansson H, McCloskey E (2012) Pitfalls in the external validation of FRAX. *Osteoporos Int* 23:423–431. doi:10.1007/s00198-011-1846-0
2. Kanis JA on behalf of the World Health Organization Scientific Group (2008) Assessment of osteoporosis at the primary health-care level. Technical Report. WHO Collaborating Centre, University of Sheffield, UK. Available at <http://www.shef.ac.uk/FRAX/index.htm>
3. Collins GS, Mallett S, Altman DG (2011) Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFracture Scores. *BMJ* 342:d3651. doi:10.1136/bmj.d3651
4. Kanis JA, Oden A, Johnell O et al (2007) 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* 18:1033–1046
5. Bolland MJ, Siu AT, Mason BH et al (2011) Evaluation of the FRAX and Garvan fracture risk calculators in older women. *J Bone Miner Res* 26:420–427
6. Kanis JA, Hans D, Cooper C et al (2011) Interpretation and use of FRAX in clinical practice. *Osteoporos Int* 22:395–411