Overview of fracture prediction tools

John A Kanis¹,², Nicholas C Harvey³, Helena Johansson¹,³, Anders Odén¹, Eugene V. McCloskey¹, and William D. Leslie⁴
¹Centre for Metabolic Bone Diseases, University of Sheffield Medical School, UK
²Institute of Health and Ageing, Australian Catholic University, Melbourne, Australia
³MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK
⁴Centre for Bone and Arthritis Research (CBAR), Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
⁵University of Manitoba, Winnipeg, Canada

Abstract

The characterisation of risk factors for fracture that contribute significantly to fracture risk, over and above that provided by BMD, has stimulated the development of risk assessment tools. The more adequately evaluated tools, all available online include the FRAX® tool, the Garvan fracture risk calculator and, in the UK only, QFracture®. Differences in the input variables, output and model construct give rise to marked differences in the computed risks from each calculator. Reasons for the differences include the derivation of fracture probability (FRAX) rather than incidence (Garvan, QFracture), limited calibration (Garvan) and inappropriate source information (QFracture). These differences need to be taken into account in the evaluation of assessment guidelines.

Keywords

calibration; fracture risk; FRAX; Garvan fracture risk calculator; QFracture

Introduction

A significant advance over the past 15 years has been the development of medical interventions that have been shown in high quality randomised controlled trials to decrease the risk of fragility fractures [1, 2]. Unfortunately, a minority of men and women receive treatment even after sustaining a fragility fracture [1, 3]. The reason for a large treatment gap (the difference between the number of individuals at high risk and the proportion of the...
population that receives treatment) is complex and multifactorial. One of the reasons is, however, limitations in the assessment of fracture risk.

Although the diagnosis of the disease relies on the quantitative assessment of bone mineral density (BMD), a major determinant of bone strength, the clinical significance of osteoporosis lies in the fractures that arise. The causation of fractures is, however multifactorial. In this respect, there are some analogies with other multifactorial chronic diseases. For example, hypertension is diagnosed on the basis of blood pressure whereas an important clinical consequence of hypertension is stroke, the likelihood of which is dependent on multiple fractures including hypertension.

Despite many guidelines using BMD thresholds to determine whether treatments should be recommended, the multifactorial nature of fracture risk means that BMD does not capture non-skeletal determinants of fracture risk such as liability to fall. A number of risk factors for fracture has been identified that contribute significantly to fracture risk over and above that provided by BMD [4]. A good example is age where the same BMD has a different significance at different ages, such that fracture risk is much higher in the elderly than in the young [5, 6]. This is because age contributes to risk independently of BMD. The realisation that independent risk factors in combination with BMD predict fractures with greater accuracy than BMD alone [4] has led to the development of risk prediction tools to better categorise individuals at high or low risk and, in turn to optimise clinical decision making with regard to therapeutic intervention. This paper reviews the strengths and weaknesses of the most commonly espoused risk assessment tools.

## Fracture risk prediction tools

Several assessment tools have been derived most of which have been poorly validated [7, 8]. The more adequately evaluated tools, all available online include the FRAX® tool [7], the Garvan fracture risk calculator [9, 10] and, in the UK, QFracture® [11, 12].

### FRAX

The FRAX algorithm is based on a series of meta-analyses of data from 12 independent fracture studies from North America, Europe, Asia, and Australia [13–18], which comprised a total of 60,000 men and women with more than 250,000 person-years of follow-up, and included more than 1100 cases of hip fracture and 3300 osteoporotic fractures [7].

After the fracture risk algorithm had been constructed using primary data from these studies, a validation study was performed using individual-level data from 11 independent population-based cohorts that were not used in the development of the original model [19]. The latter comprised a total of 230,000 individuals with more than 1.2 million person-years of follow up. By reason of its large numbers, its international character, and the care taken in its construction and implementation, the FRAX algorithm is considered to have unique authority [20]. The algorithm provides information on the 10-year probabilities of hip fracture and any major osteoporotic fracture (defined as a hip, wrist, humerus, or clinical vertebral fracture). Further details are provided in an accompanying paper in this volume.
QFracture

The QFracture tool is based on a UK prospective open cohort study of routinely collected data from 357 general practices on over 2 million men and women aged 30-85 years (www.qfracture.org) [11]. Like the FRAX tool it takes into account history of smoking, alcohol, corticosteroid use, parental history (of hip fracture or osteoporosis) and several secondary causes of osteoporosis. Unlike FRAX it also includes a history of falls (yes/no only over an unspecified time frame), utilises a large number of clinical risk factors and no provision is made for BMD. It has been internally validated (i.e. from a stratum of the same population), and externally validated in a similar population (routinely collected data in general practitioner records). The performance characteristics and calibration in the UK have been compared with FRAX with comparable results for hip fracture. The tool is not calibrated to the epidemiology of other countries. A feature of QFracture is that it is more cumbersome (more questions), and does not accommodate the inclusion of BMD. BMD measurements are dismissed as “expensive and inconvenient tests” [11] and so the model ignores a wealth of data demonstrating the utility of BMD testing in fracture risk assessment.

Garvan

The Garvan tool (www.garvan.org.au) is based on many fewer men and women from a single study, the Australian Dubbo Osteoporosis Epidemiology Study (DOES) of approximately 2500 men and women age 60 years or more. It differs from FRAX by including a history of falls (categorised as 0, 1, 2, >2 in the previous year), and the number of previous fragility fractures (categorised as 0, 1, 2, >2), but does not include other FRAX variables such as parental history of hip fracture, secondary osteoporosis, rheumatoid arthritis, glucocorticoid use, smoking and intake of alcohol. The output of the tool differs from FRAX in that it reports the risk of a larger number of fracture sites (additionally includes fractures of the distal femur, proximal tibia/fibula, distal tibia/fibula, patella, pelvis, ribs sternum, hands and feet excluding digits). Further details are provided in an accompanying paper in this volume.

Comparative features

There are important differences in the input variables, output and model features that make comparison of the models problematic. A summary of input variables is given in Table 1.

With regard to input variables, both Garvan and QFracture include a history of falls whereas this is not an input variable in FRAX. Indeed the Garvan tool weights the number of falls in the past year. Whereas falls are a strong risk factor for fracture, the incorporation of falls into FRAX is problematic for several reasons. First, at the time of the release of FRAX, existing falls data were not of adequate quality, including the heterogeneous construct of questions on falls. Second, falls risk is inherently taken into account in the algorithm, though not as an input variable. Thus, the fracture probability given for any combination of risk factors assumes that the falls risk is that observed (but not documented) in the cohorts used to construct FRAX. Third, the interrelationship of falls risk with the other FRAX variables has
been inadequately explored on an international basis. Fourth, the relationship between the risk variable and mortality needs to be accounted for, but there are no data available.

These technical problems aside, risk assessment tools are intended to identify a risk that is amenable to a therapeutic intervention. However, falls as a risk variable does not consistently pass the test of reversibility of risk [21–23], a necessary feature of any risk variable used in tools to direct interventions [7, 24]. Recently, an analysis in elderly men, available as a meeting abstract, indicated that the predictive value of falls for fracture waned significantly with time [25]. If the phenomenon is replicated more generally, then this would further question the utility of falls history in the long term (e.g. 10-year) assessment of fracture risk. Thus, a useful role of falls history in fracture risk assessment remains sub judice. There are also few data that many of the QFracture risk factors (cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants usage, history of falls or liver disease) characterises a risk that is amenable to bone-targeted interventions.

Other important differences between models include the question construct for fracture history given as past fragility fracture (for FRAX), fractures since the age of 50 years (Garvan) or past wrist, spine, hip or shoulder fracture (QFracture). For BMD, the femoral neck is the reference site for FRAX and for Garvan but is not an input variable for QFracture.

Important differences in the output and model features are given in Table 2. The Garvan instrument includes many more fracture outcomes than QFracture or FRAX. Compared with FRAX, the inclusion of these additional fractures is expected to inflate fracture risks in women by 34-45% depending on age [26].

The outcome variable differs between models, not only in the fracture sites but also in the metric. In the case of FRAX, the algorithm computes a fracture probability (i.e. a metric that incorporates the death hazard) which is not synonymous with simple fracture incidence. This is illustrated in Figure 1 that shows the difference in outcome between hip fracture probability as estimated by FRAX and incidence calculated using the Garvan instrument. As a result, the comparisons are problematic [27].

**Comparison of output**

A comparison of the performance characteristics of the three models is beyond the scope of this review, but for hip fracture risk appear to be rather comparable [28] taking into account the methodological flaws in most of the comparative studies [27]. Given the considerations above, it is not surprising that each model identifies different segments of populations at high risk. Of equal importance is that very different estimates are derived in single populations. For example, the three models were compared in a referral population from Denmark. For hip fracture prediction, the mean 10-year risk (with interquartile range) was 11.0% (5.6-17.0%) for FRAX, 14.5% (7.3-23.0%) for QFracture and 26.4% (10.4-53.3%) for Garvan [29]. Irrespective of the model constructs, these differences are not surprising given that neither QFracture nor Garvan are calibrated for the country in which they were used.
When QFracture and FRAX are applied to the UK population, there is reasonable concordance for hip fracture risk since both are calibrated to the UK, though in different ways. The Garvan instrument is calibrated only to Dubbo and is the outlier (Figure 2). The concordance of the Garvan and FRAX tools reported in Canada is thus a fortuitous accident occasioned by the similar epidemiology between Canada and Dubbo. The claim of good calibration in Norway is not supported by the evidence.

Whereas QFracture and FRAX are comparably calibrated for hip fracture risk, a quite different pattern is evident for major osteoporotic fractures where the probabilities derived from FRAX are markedly higher than the incidences from QFracture. The Garvan instrument gives even higher values as shown in Figure 3 for the same clinical scenarios as in Figure 2. The Garvan tool provides the highest risks, in part because the output is the 10-year incidence of all fractures (minus those at the digits). QFracture gives the lowest estimates. This is confirmed in Figure 4 where the 10-year hip fracture rates/probabilities are shown in women at each decile of risk category.

The reason for the discrepancy is that QFracture is derived from General Practitioner records that are often incomplete for some important variables. For example, GP records are reasonably accurate for the documentation of hip fracture but notoriously unreliable for other major fractures, particularly vertebral fractures. Thus, the prevalence of a prior major fracture in the QFracture data base is 1.9%, whereas prior fracture is estimated at 21-45% in women from the UK, depending on age. Of these, approximately half will be major fractures. For a parental history of osteoporosis or hip fracture the prevalence is given at 0.3% in the QFracture data base whereas meta-analysis of prospective studies gives a prevalence of parental hip fracture at 13%. The impact of the inaccuracies is difficult to quantify but is likely to decrease the median of the distribution of 10-year risk in the population. Empirical observation supports this view in that at each tenth of risk category, QFracture risk is lower than FRAX-based probabilities (see Figure 4).

The poor and inaccurate capture of clinical risk factors is likely to bias the weights for both hip fracture risk and major fracture risk. In the case of FRAX and Garvan, the probability of fracture is approximately doubled with a prior history of fracture consistent with worldwide observation. In the case of major fracture incidence, QFracture determines an increase in risk ratio of approximately only 8%, rather than the expected doubling of risk. As expected from meta-analysis, the impact of a prior fracture is somewhat greater at younger ages and is accommodated in FRAX. In contrast, the weighting given for a prior fracture as a risk fracture is unrealistic for QFracture and does not vary with age (the latter, also the case for Garvan).

A further problem arises in considering the pattern of fractures with age. As expected, FRAX probabilities of a major fracture exceed that of hip fracture at all ages. In the case of QFracture the incidence of hip fracture and the incidence of major fracture are identical from the age of 85 years. This implies that no fractures of the spine, humerus or distal forearm arise in women from the age of 85 years. Again, this contrasts with empirical observation. Indeed, fragility fractures other than hip fracture account for 64-67% of fractures in women and men (respectively) aged 85-89 years.
These considerations indicate that little credence can be afforded for estimates of major fracture using the QFracture algorithm. They further indicate that the weights given to several of the clinical risk factors are inappropriate. Both factors result in a large underestimation of major fracture risk by QFracture. In contrast, the prevalence and weight of clinical risk factors, and the pattern of fractures with age of major fractures has been extensively validated with the use of FRAX [26, 38–41]. In summary, FRAX is well calibrated whereas QFracture under-predicts the risk of major fractures at all levels of risk.

**Conclusion**

There are important differences between the FRAX tool, the Garvan fracture risk calculator and QFracture® that make their comparison problematic. Differences in the input variables, output and model construct give rise to marked differences in the computed risks from each calculator. Reasons for the differences include the derivation of fracture probability (FRAX) rather than incidence (Garvan, QFracture), poor calibration (Garvan) and inappropriate source information (QFracture).

**Acknowledgements**


**References**


Figure 1.
The risk of hip fracture with age in a model that considers 10-year fracture risk alone (the Garvan tool) and FRAX which computes the probability of hip fracture from the fracture and death hazards (FRAX). The T-scores are set differently in the two models so that the risks are approximately equal at the age of 60 years. Data are computed from the respective web sites. [27] with kind permission from Springer Science+Business Media B.V]
Figure 2.
Comparison of the risk of a hip fracture using the Garvan calculator, FRAX and QFracture in women with a prior fracture from the UK by age. Height was set at 165cm (QFracture and FRAX) and weight at 65kg (all models).
Figure 3.
Comparison of the risk of any fragility fracture using the Garvan calculator, and the risk of a major osteoporotic fracture with FRAX and QFracture in women with a prior fracture from the UK by age. Height was set at 165cm (QFracture and FRAX) and weight at 65kg (all models).
Figure 4.
Comparison of the distribution of FRAX and QFracture (QF) model output by decile of risk in women for hip fracture (left panel [12] and major fracture (right panel) [33]. The diagonal line shows the line of identity. [34 with kind permission from Springer Science+Business Media B.V].
Table 1

Input risk variables for FRAX, QFracture and the Garvan assessment tools.

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>FRAX</th>
<th>Garvan</th>
<th>QFracture</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>+</td>
<td>+*</td>
<td>+</td>
<td>* only when BMD is not entered</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>+*</td>
<td>-</td>
<td>+</td>
<td>*US and Singapore</td>
</tr>
<tr>
<td>Previous fragility fracture</td>
<td>+</td>
<td>+*</td>
<td>+**</td>
<td>*from the age of 50y *also number of prior fractures **prior hip, spine or shoulder</td>
</tr>
<tr>
<td>BMD</td>
<td>+</td>
<td>+*</td>
<td>-</td>
<td>* site of BMD not specified on website but elsewhere stated to be femoral neck.</td>
</tr>
<tr>
<td>Parental history of hip fracture</td>
<td>+</td>
<td>-</td>
<td>+*</td>
<td>* Also, parental history of OP</td>
</tr>
<tr>
<td>Smoking</td>
<td>+*</td>
<td>-</td>
<td>+**</td>
<td>*current **categories of exposure</td>
</tr>
<tr>
<td>Alcohol intake</td>
<td>+*</td>
<td>-</td>
<td>+**</td>
<td>* ≥3 units per day **categories of exposure</td>
</tr>
<tr>
<td>Taking steroid tablets regularly</td>
<td>+*</td>
<td>-</td>
<td>+</td>
<td>*currently exposed to oral glucocorticoids or has been exposed to oral glucocorticoids for more than 3 months at a dose of prednisolone of 5mg daily or more</td>
</tr>
<tr>
<td>Osteogenesis imperfecta</td>
<td>+*</td>
<td>-</td>
<td>-</td>
<td>* grouped as secondary OP</td>
</tr>
<tr>
<td>Diabetes</td>
<td>+*</td>
<td>-</td>
<td>+**</td>
<td>*Type 1 grouped as secondary OP ** Type 1, type 2</td>
</tr>
<tr>
<td>In nursing home</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Falls history</td>
<td>-</td>
<td>+*</td>
<td>+</td>
<td>*categories of exposure</td>
</tr>
<tr>
<td>Dementia</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Asthma or COPD</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Heart attack, angina, stroke or TIA</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>+*</td>
<td>-</td>
<td>+</td>
<td>*grouped as secondary OP</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>-</td>
<td>-</td>
<td>+*</td>
<td>*stage 4 or 5</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis or SLE</td>
<td>+*</td>
<td>-</td>
<td>+</td>
<td>*RA alone</td>
</tr>
<tr>
<td>Malabsorption</td>
<td>+*</td>
<td>-</td>
<td>+**</td>
<td>*grouped as secondary OP **Crohn’s disease, ulcerative colitis, coeliac disease, steatorrhea or blind loop syndrome</td>
</tr>
<tr>
<td>Endocrine disorders</td>
<td>+*</td>
<td>-</td>
<td>+**</td>
<td>*untreated long-standing hyperthyroidism, hypogonadism **thyrotoxicosis, hyperparathyroidism, Cushing’s syndrome</td>
</tr>
<tr>
<td>Premature menopause</td>
<td>+*</td>
<td>-</td>
<td>-</td>
<td>* (&lt;45 years) grouped as secondary OP</td>
</tr>
<tr>
<td>Epilepsy or taking anticonvulsants</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Taking antidepressants</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Chronic malnutrition</td>
<td>+*</td>
<td>-</td>
<td>-</td>
<td>*grouped as secondary OP</td>
</tr>
<tr>
<td>HRT</td>
<td>-</td>
<td>-</td>
<td>+*</td>
<td>* oestrogen only</td>
</tr>
</tbody>
</table>

*J Clin Densitom. Author manuscript; available in PMC 2018 January 14.*
HRT hormone replacement therapy; OP osteoporosis; SLE systemic lupus erythematosus
Table 2

Comparative features of the Garvan tool, QFracture and FRAX.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Garvan</th>
<th>QFracture</th>
<th>FRAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Externally validated</td>
<td>Yes (Canada)</td>
<td>Yes (UK only)</td>
<td>Yes, internationally</td>
</tr>
<tr>
<td>Calibrated</td>
<td>No</td>
<td>Yes (hip only)</td>
<td>Yes</td>
</tr>
<tr>
<td>Applicability</td>
<td>Uncertain</td>
<td>UK</td>
<td>58 countries</td>
</tr>
<tr>
<td>Falls as an input variable</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>BMD as an input variable</td>
<td>Yes *</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Prior fracture as an input variable</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Family history as an input variable</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Outcome</td>
<td>All fractures excluding digits</td>
<td>Hip, forearm, spine, shoulder</td>
<td>Hip, forearm, spine, humerus</td>
</tr>
<tr>
<td>Outcome metric</td>
<td>Incidence</td>
<td>Incidence</td>
<td>Probability</td>
</tr>
</tbody>
</table>

\* site of BMD not specified on website