Independent external validation of nomograms for predicting risk of low-trauma fracture and hip fracture

Lisa Langsetmo PhD, Tuan V. Nguyen PhD, Nguyen D. Nguyen PhD MD, Christopher S. Kovacs MD, Jerilynn C. Prior MD, Jacqueline R. Center PhD, Suzanne Morin MSc MD, Robert G. Josse MD, Jonathan D. Adachi MD, David A. Hanley MD, John A. Eisman PhD; and the Canadian Multicentre Osteoporosis Study Research Group

See related commentary by Bolland

Abstract

Background: A set of nomograms based on the Dubbo Osteoporosis Epidemiology Study predicts the five- and ten-year absolute risk of fracture using age, bone mineral density and history of falls and low-trauma fracture. We assessed the discrimination and calibration of these nomograms among participants in the Canadian Multicentre Osteoporosis Study.

Methods: We included participants aged 55–95 years for whom bone mineral density measurement data and at least one year of follow-up data were available. Self-reported incident fractures were identified by yearly postal questionnaire or interview (years 3, 5 and 10). We included low-trauma fractures before year 10, except those of the skull, face, hands, ankles and feet. We used a Cox proportional hazards model.

Results: Among 4152 women, there were 583 fractures, with a mean follow-up time of 8.6 years. Among 1606 men, there were 116 fractures, with a mean follow-up time of 8.3 years. Increasing age, lower bone mineral density, prior fracture and prior falls were associated with increased risk of fracture. For low-trauma fractures, the concordance between predicted risk and fracture events (Harrell C) was 0.69 among women and 0.70 among men. For hip fractures, the concordance was 0.80 among women and 0.85 among men. The observed fracture risk was similar to the predicted risk in all quintiles of risk except the highest quintile of women, where it was lower. The net reclassification index (19.2%, 95% confidence interval [CI] 6.3% to 32.2%), favours the Dubbo nomogram over the current Canadian guidelines for men.

Interpretation: The published nomograms provide good fracture-risk discrimination in a representative sample of the Canadian population.

Competing interests: See end of article for competing interests.

This article has been peer reviewed.

Correspondence to:
Dr. David Hanley,
dahanley@ucalgary.ca

***CMAJ* 2011. DOI:10.1503/cmaj.100458**

Current recommendations for the treatment of osteoporosis are in transition. The T-score-based definition of osteoporosis and osteopenia by the expert committee of the World Health Organization on bone mineral density has been used in many guidelines to set intervention thresholds for treatment. However, studies have consistently reported that the highest number of fractures in a given population occurs in those with osteopenic or normal bone mineral density.[1,2] In fact, the National Osteoporosis Foundation has singled out people with osteopenic bone mineral density as a population in which assessment for fracture risk is merited.[3]

Nevertheless, appropriate prevention and treatment strategies for such people are uncertain.[4] Recent developments include the assessment of absolute fracture risk based on bone mineral density and other risk factors. Current Canadian methodology determines categorical risk based on age, sex, T-score, fracture history and glucocorticoid use.[5] These criteria were derived from Swedish data, but have been assessed and validated in a cohort of Manitoba women.[6] Newer nomograms based on the Australian cohort of the Dubbo Osteoporosis Epidemiology Study[7] are now available for the calculation of low-trauma hip fracture[8] and any fracture.[9] These nomograms provide continuous estimates for five- and 10-year absolute fracture risk in both men and women (available at http://fracturerisk calculator.com). The use of factors in addition to bone mineral density may provide a

better assessment of fracture risk for people who are near the T-score thresholds and facilitate decisions regarding therapeutic intervention.

A key step in the development of any prediction model is the assessment of its validity.[10] The aim of our study was to assess the performance of the Australian-derived nomogram among community-dwelling Canadians aged 55–95 years old. The first part of this assessment was a comparison of the nomogram model using the same variables, but using data from a Canadian population — participants in the Canadian Multicentre Osteoporosis Study (www.camos.org). The second part involved computing the calibration and discrimination of the nomogram in a Canadian cohort. The final part was comparison of the new assessments with the existing Canadian risk classification system.

Methods

Participants

The study sample included all participants in the Canadian Multicentre Osteoporosis Study cohort who were 55–95 years old at baseline, who had undergone measurement of bone mineral density and had at least one year of follow-up data. Of 6539 women and 2884 men in the entire baseline cohort, 4940 women and 1883 men met the age criteria; 788 women and 277 men were excluded based on missing data, leaving a study sample of 4152 women and 1606 men. Details of how the original cohort was randomly sampled from nine Canadian communities have been previously published[11] and are available at www.camos.org. Written informed consent was obtained from all participants and ethics approval was granted through McGill University and the ethics review boards for each participating centre.

Data collection

At baseline, all participants were given an interviewer-administered questionnaire, which determined their demographic characteristics, medical history and risk factors for fractures. Follow-up visits were scheduled in year 3 for participants aged 40 to 60 years only and in years 5 and 10 for all participants. In all other years, a self-administered, fracture-related questionnaire was mailed. We included incident fragility fractures that occurred between baseline and the tenth annual (2005–2006) follow-up.

Nomogram risk factors

The Dubbo nomograms were derived using model selection, and the final model included age, bone mineral density T-score of the femoral neck, number of prior fractures (after age 50) and number of falls in the previous year.[8,9] For this study, bone mineral density T-scores were based on published reference standards for Canadians.[12] All lunar measurements were converted to equivalent hologic values (g/cm^2) using standard reference formulas.[13] Detailed densitometer quality control is described elsewhere.[14] We used number of falls in the preceding one-month (as opposed to one-year) period because this was the period examined in the CaMos questionnaire administered at baseline. The Dubbo nomogram for estimating five- and ten-year risk of hip fracture and any fracture in women and men is shown in Appendix 1 (available at www.cmaj.ca/cgi/content/full/cmaj.100458/DC1).

Fracture assessment

Self-reported incident fractures were identified by annual follow-up and confirmed by structured interview (telephone or in-person). Information that was gathered included fracture site, circumstance, treatment, and x-ray or medical report (if available). As in the Dubbo Osteoporosis Epidemiology Study, we included only low-trauma fractures (i.e., without trauma or caused by a fall from standing height or less) and excluded fractures of the skull, face, hands, ankles and feet.

Risk category comparison

Risk categories based on the predicted 10-year osteoporotic fracture risk in the Dubbo low-trauma fracture nomogram were as follows: low risk = 0%–10%, moderate = 10%–20% and high ≥ 20%. Categories based on the T-score thresholds of the World Health Organization were low risk = T-score > –1 with no prior fracture, moderate = –2.5 < T-score ≤ –1 with no prior fracture, and high = T-score ≤ –2.5 or prior osteoporotic fracture. Ten-year fracture risk categories in the Canadian guidelines were as follows: low risk = 0%–10%, moderate = 10%–20% and high > 20%,[5] and were derived from age, minimum T-score (lumbar spine, total hip, femoral neck, trochanter),[15] glucocorticoid use and history of fracture after age 40. We used current systemic glucocorticoid use together with at least three months of total use as a proxy for three months in the past year.

Statistical methods

We used Cox's proportional hazards to develop a comparison model based on study participants with either hip fracture or low-trauma fracture as the outcome. Model entry was time of study enrolment and model exit was the first of the fol-

lowing: 10-year follow-up, fracture, loss to follow-up, or death. Model diagnostic tests included assessment of linearity of continuous variables, proportional hazards and overall model fit.

Validation of all models included both discriminative ability and calibration. Discriminative ability was assessed using the Harrell C, a statistic for survival analysis that is analogous to C-statistic or the area under the curve as used for diagnostic tests.[16] Calibration was assessed by dividing the cohort into quintiles according to risk predicted by the nomogram and comparing the observed risk and predicted risk. We used Kaplan–Meier methods to compute observed risk. Cross-tabulation of categories, the Aickin kappa[17] and the net reclassification index[18] were used for all comparisons. The net reclassification index is a measure indicating the likelihood that people were correctly reclassified (i.e., that events were reclassified higher risk and non-events reclassified lower risk) versus incorrectly reclassified (i.e., that events were reclassified lower risk and non-events reclassified higher risk). In simulated data, adding a biomarker with an odds ratio of 2 to an existing classification resulted in a net reclassification index of 5%.[19]

Results

The baseline characteristics of participants are shown in Table 1. Among 4152 women, there were 583 fractures with a mean follow-up time of 8.6 years. Among 1606 men, there were 116 fractures with a mean follow-up time of 8.3 years. Distribution of first fractures by skeletal site was as follows: forearm or wrist 174, ribs 133, upper arm or shoulder 100, hip 97, spine 89, leg 52, pelvis 27, multiple sites 27. We obtained documented confirmation of 78% of these fractures. The overall 10-year observed risk for low-trauma fracture was 16.2% (95% CI 15.0 to 17.5) among women and 8.7% (95% CI 7.26 to 10.4) among men. These values were slightly lower than the Dubbo nomogram mean predicted risk, which was 18.3% among women and 11.8% among men. The 10-year observed risk for hip fracture among women was 2.8% (95% CI 2.3 to 3.4), which was lower than the Dubbo nomogram mean predicted risk of 5.6%. The 10-year observed risk for hip fracture among men was 2.4% (95% CI 1.7 to 3.5), which was similar to the mean predicted risk of 2.6%.

A comparison of the model coefficients based on the Canadian Multicentre Osteoporosis Study cohort and the Dubbo Osteoporosis Epidemiology Study cohort is shown in Table 2. The association between prior fracture and future hip fracture or low-trauma fracture was notably stronger in the Dubbo cohort. Overall individual risk profiles derived from the Canadian and Australian models were similar despite the modest difference in some model parameters. Pearson correlations between log hazards were $r = 0.96$ (women) and $r = 0.82$ (men) for low-trauma fracture and $r = 0.93$ (women) and $r = 0.99$ (men) for hip fracture.

Table 1: Baseline characteristics of participants

Characteristic	Men *n = 1606*	Women *n = 4152*
Age, yr, mean (SD)	67.60 (7.60)	67.70 (7.60)
Height, cm, mean (SD)	172.80 (7.00)	159.10 (6.30)
Weight, kg, mean (SD)	81.30 (13.20)	68.70 (13.10)
Bone mineral density, g/cm^2, mean (SD)		
Femoral neck	0.78 (0.13)	0.68 (0.12)
Total hip	0.99 (0.15)	0.83 (0.14)
Lumbar spine	1.05 (0.18)	0.91 (0.17)
Femoral neck T-score, mean (SD)	–1.00 (1.00)	–1.43 (0.93)
Minimum T-score (lumbar spine, trochanter, femoral neck, total hip), mean (SD)	–1.19 (0.91)	–1.78 (1.02)
Dubbo nomogram 10-year predicted fracture risk,* %, mean (SD)	11.75 (12.74)	18.33 (14.04)
Dubbo nomogram 10-year predicted hip-fracture risk,* %, mean (SD)	2.66 (6.16)	5.63 (10.31)
Fracture history (after age 50), no. (%)		
None	1518 (94.5)	3628 (87.4)
One fracture	25 (1.6)	115 (2.8)
Two fractures	54 (3.4)	54 (7.7)
Three or more fractures	9 (0.6)	90 (2.2)
Previous falls (1 mo before baseline), no. (%)		
None	1516 (94.4)	3887 (93.6)
One fall	73 (4.6)	235 (5.7)
Two falls	13 (0.8)	21 (0.5)
Three or more falls	4 (0.3)	9 (0.2)
Glucocorticoid use (at baseline and history of ≥ 3 mo), no. (%)	17 (1.1)	59 (1.4)
Antiresorptive use, no. (%)	5 (0.3)	1235 (29.7)
Tobacco use, no. (%)	232 (14.4)	502 (12.1)
Alcohol use (any), no. (%)	1150 (71.6)	2237 (52.9)
Vertebral deformity, no. (%)		
Normal	894 (55.7)	2337 (56.3)
Grade 1	206 (12.8)	565 (13.6)
Grade 2	117 (7.3)	387 (9.3)
Missing data	389 (24.2)	863 (20.8)

*Nomogram risk factors were age, femoral neck T-score, fracture history after age 50 and falls in the last year (proxy: in the last month).

For low-trauma fractures, the concordance (expressed as Harrell's C) between predicted risk as assessed by the Dubbo nomogram and fracture outcomes in the study sample was C = 0.69 among women and C = 0.70 among men. For hip fractures, the concordance was C = 0.80 among women and C = 0.85 among men. Similar concordance was found when comparing risk derived from the study sample model and fracture outcomes. Calibration plots are shown in Figure 1. The observed low-trauma fracture risk was lower than the predicted risk in the highest quintile for both men and women. The observed hip fracture risk was also lower than the predicted risk in the highest quintile for women, but not for men. The observed number of hip fractures among women and the observed number of low-trauma fractures among men were both slightly lower than the predicted values across quintiles as well.

A comparison of the risk classification based on the Dubbo low-trauma fracture nomogram with that based on established T-score thresholds of the World Health Organization is shown in Appendix 2 (available at www.cmaj.ca/cgi/content/full/cmaj.100458/DC1). Overall, 58.8% ($\kappa = 0.38$) of men and 77.4% ($\kappa = 0.66$) of women were in the same fracture risk category, with stronger consistency among women. The nomogram reclassified more men as lower risk than as higher risk (26.7% v. 14.4%) but more women as higher risk than as lower risk (15.4% v. 7.2%). The net reclassification index favouring the Dubbo nomogram was equal to 6.7% (95% CI –6.0% to 19.4%) among men and 1.5% (95% CI –2.6% to 5.6%) among women, but neither was significant.

We also compared the risk classification based on the Dubbo low-trauma fracture nomogram with that appearing in Canadian guidelines.[5] The underlying thresholds based on age and T-score are similar among women (Table 3), but radically different among men. Under both schemes, 68.3% of men ($\kappa = 0.47$) were in the same frac-

Table 2: Comparison of hazard ratios derived from the Canadian Multicentre Osteoporosis Study cohort and those derived from the Australian Dubbo cohort

Outcome	Unit	Canadian Multicentre Osteoporosis study, HR (95% CI)	Australian Dubbo nomogram, HR (95% CI)
Low-trauma fracture			
Women			
Age	1 year	1.04 (1.03–1.05)	1.03 (1.02–1.05)
Femoral neck T-score	1 SD*	0.61 (0.55–0.68)	0.67 (0.60–0.75)
Prior fracture	1 fracture	1.37 (1.25–1.49)	1.77 (1.59–1.96)
Falls	1 fall	1.60 (1.20–2.12)	1.23 (1.08–1.29)
Men			
Age	1 year	1.05 (1.02–1.07)	1.09 (1.06–1.12)
Femoral neck T-score	1 SD*	0.49 (0.39–0.61)	0.74 (0.63–0.87)
Prior fracture	1 fracture	1.29 (0.96–1.71)	2.33 (1.87–2.90)
Falls	1 fall	1.43 (0.75–2.74)	1.10 (0.87–1.40)
Low-trauma hip fracture			
Women			
Age	1 year	1.10 (1.07–1.13)	1.05 (1.02–1.09)
Femoral neck T-score	1 SD*	0.44 (0.34–0.58)	0.43 (0.34–0.55)
Prior fracture	1 fracture	1.48 (1.22–1.79)	2.25 (1.41–3.59)
Falls	1 fall	0.64 (0.23–1.75)	1.44 (1.12–1.84)
Men			
Age	1 year	1.09 (1.04–1.15)	1.11 (1.04–1.09)
Femoral neck T-score	1 SD*	0.38 (0.24–0.60)	0.37 (0.24–0.56)
Prior fracture	1 fracture	1.55 (0.98–2.45)	1.82 (1.11–2.97)
Falls	1 fall	0.90 (0.21–3.82)	1.24 (0.69–2.20)

Note: CI = confidence interval, HR = hazard ratio, SD = standard deviation.
*1 SD is the standard T-score definition and refers to the standard deviation of bone mineral density scores in reference data based on healthy adults at time of peak bone mass.

ture risk category, compared to 79.1% ($\kappa = 0.69$) of women (Appendix 3, available at www.cmaj.ca/cgi/content/full/cmaj.100458/DC1). More men were reclassified as higher risk than as lower risk (30.2% v. 1.5%), which resulted in a drastic shift in risk prevalence and significant improvement in risk classification with a net reclassification index equal to 19.2% (95% CI 6.3% to 32.2%). The reclassification for women resulted in a slight decline in risk classification with a net reclassification index equal to –5.5% (95% CI –9.5% to –1.5%).

The prevalence of high fracture risk based on the three risk classification schemes is shown in Figure 2. The strongest age-gradient in high-risk prevalence was seen by using the Dubbo nomogram and resulted in the identification of roughly 90% of men and women older than 80 years old as high risk. For women, the overall prevalence of high risk was similar using the Dubbo nomogram or Canadian guidelines, but for men it was highest using the nomogram. We found that 40% of men with osteoporotic fractures were identified as high risk using the nomogram, which was notably higher than the percentage identified using other risk assessments.

Interpretation

Our study provides independent external validation of published nomograms based on the Dubbo Osteoporosis Epidemiology Study cohort of Australia for predicting absolute fracture risk in a Canadian population. The ability of the nomograms to discriminate between those who will and those who will not have hip fracture was excellent, whereas the discrimination for low-trauma fracture was more modest. The calibration of the Dubbo nomogram was very good. The main discrepancy between observed and predicted risks was in the highest quintiles. Possible explanations include treatment effects, competing mortality, the differing periods over which falls were assessed, or model shrinkage (i.e., the tendency of statistical models to overestimate the difference between low and high risk when using independent data).[10] The discrepancy between predicted and observed fracture risk would not affect clinical decision-making because people in the highest quintile would be considered for treatment even with the lower observed risk.

The classification performance of categories based on the low-trauma fracture nomogram among women was similar to current Canadian guidelines,[5] whereas for men there was substantially better performance. This improvement in classification was a result of better identification of men at high risk, given that the current model identifies very few of these people. Inclusion of fracture sites that better reflect osteoporosis among men may explain the substantial shift in thresholds. For women, the existing guidelines have previously been validated in a cohort of women in the province of Manitoba.[6] Our analysis also shows that the continuous (rather than categorical) nomogram-

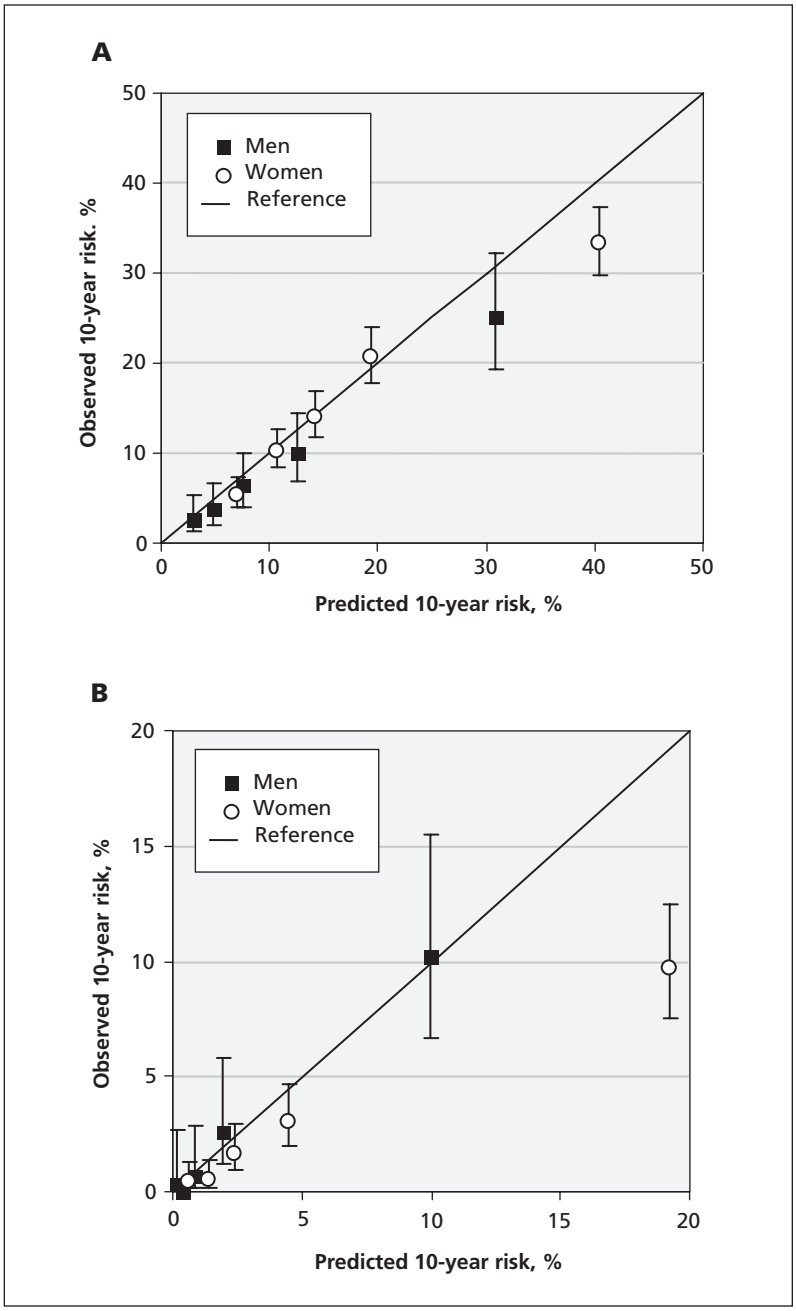

Figure 1: Comparison of the predicted versus observed 10-year risk of (A) low-trauma fracture and (B) hip fracture in the Canadian Multicentre Osteoporosis Study cohort, according to the Dubbo nomogram. Quintile cutoffs for the predicted 10-year risk of low-trauma fracture were 3.8%, 6.1%, 9.5% and 16.4% for men and 9.2%, 12.3%, 16.4% and 23.7% for women. Quintile cutoffs for the predicted 10-year risk of hip fracture were 0.3%, 0.6%, 1.2% and 3.1% for men and 1.0%, 1.8%, 3.2% and 6.4% for women.

based risk prediction is applicable to Canadian women and hence provides a refinement of the existing criteria.

The FRAX model is another tool for fracture-risk prediction that was constructed using a different methodology and is derived from a meta-analysis of nine studies, including the Dubbo Osteoporosis Epidemiology Study and the Canadian Multicentre Osteoporosis Study.[20] The discriminative ability of the Dubbo nomogram is slightly higher than that found in the external validation of the FRAX model in most of the test cohorts;[20] furthermore, only one of the test cohorts included men. Discrimination of the Dubbo nomogram is nearly the same as that of the Canadian FRAX model in a Manitoba cohort.[21] The validation of the model for men is reassuring, given that an independent assessment of the FRAX model in a small sample showed poor discrimination in men.[22] For women, results from the Study of Osteoporotic Fractures showed little benefit of the FRAX model compared with a model based on age, bone mineral density and previous fracture.[23] The Dubbo nomogram appears robust to change in study population, as indicated by similarities between internal and external validation.[8,9] The discriminative ability of the Dubbo nomogram is similar to the widely used Framingham risk score for cardiovascular events.[24]

There may be underlying differences in risk between cohorts beyond that of the measured risk factors, either directly attributable to fracture propensity, or indirectly as the result of competing risks. We note that the distribution of bone mineral density is roughly similar in Canada[12] and Australia.[25,26] However, we also note that the geographic variation in fracture does not always reflect underlying variation in bone mineral density.[27] A recent comparison of surveillance rates of hip fracture among Canada, the United States and Germany has shown differences in hip fracture incidence that are both age- and sex-dependent.[28]

Table 3: Comparison by age of bone mineral density T-score thresholds for low-trauma fracture risk categories determined by Dubbo nomogram* and Canadian guidelines for 2005†

	Dubbo nomogram		Canadian guidelines	
Age, yr	Low risk 0–10%	High risk > 20%	Low risk 0–10%	High risk > 20%
Men				
50	> –2.7	< –3.8	> –3.4	None
55	> –2.4	< –3.5	> –3.1	None
60	> –2.1	< –3.2	> –3.0	None
65	> –1.8	< –2.9	> –2.7	None
70	> –1.5	< –2.5	> –2.1	< –3.9
75	> –1.2	< –2.2	> –1.5	< –3.2
80	> –0.9	< –1.9	> –1.2	< –3.0
85	> –0.6	< –1.6	> –1.3	< –3.3
Women				
50	> –2.2	< –3.7	> –2.3	< –3.9
55	> –1.8	< –3.4	> –1.9	< –3.4
60	> –1.5	< –3.0	> –1.4	< –3.0
65	> –1.1	< –2.6	> –1.0	< –2.6
70	> –0.7	< –2.2	> –0.8	< –2.2
75	> –0.3	< –1.9	> –0.7	< –2.1
80	> 0.0	< –1.5	> –0.6	< –2.0
85	> 0.4	< –1.1	> –0.7	< –2.2

*Based on T-score for bone mineral density of femoral neck and assuming no additional risk factors (e.g., fracture after age 50 or falls in last year).
†Based on minimum T-score for bone mineral density (femoral neck, lumbar spine, total hip, trochanter) and assuming no additional risk factors (e.g., fracture after age 40, glucocorticoids).

Limitations

Risk assessment is used to identify those who are above (or below) a certain level of risk for fracture. The label "high risk" should be a stimulus to physician-patient discussion of management options, but does not necessarily translate into pharmacologic interventions. We included participants who were receiving osteoporosis therapy. Gaps in calibration occurred in the direction predicted by therapy use among those at risk. The limited number of fractures among men results in uncertainty in the model parameters and in the tests of discrimination and calibration. We included all clinical fractures by self-report, which is reliable for hip and wrist fracture, but may result in some misclassification.[29] Model calibration is not static, and recalibration may be necessary due to secular changes.[30] Nonresponse bias may be present, but as previously shown, is minor except among those over 80 years of age.[31] Finally, the results may not be generalizable to those in institutional care.

Conclusion

The Dubbo fracture risk nomogram was validated in an independent, population-based Canadian cohort of community-dwelling men and women and was shown to provide good discrimination for risk of future fracture and hip fracture. These simple nomograms for absolute fracture risk have potential to inform clinical decisions, notably those related to the large numbers of men and women with osteopenia who are at moderate risk.

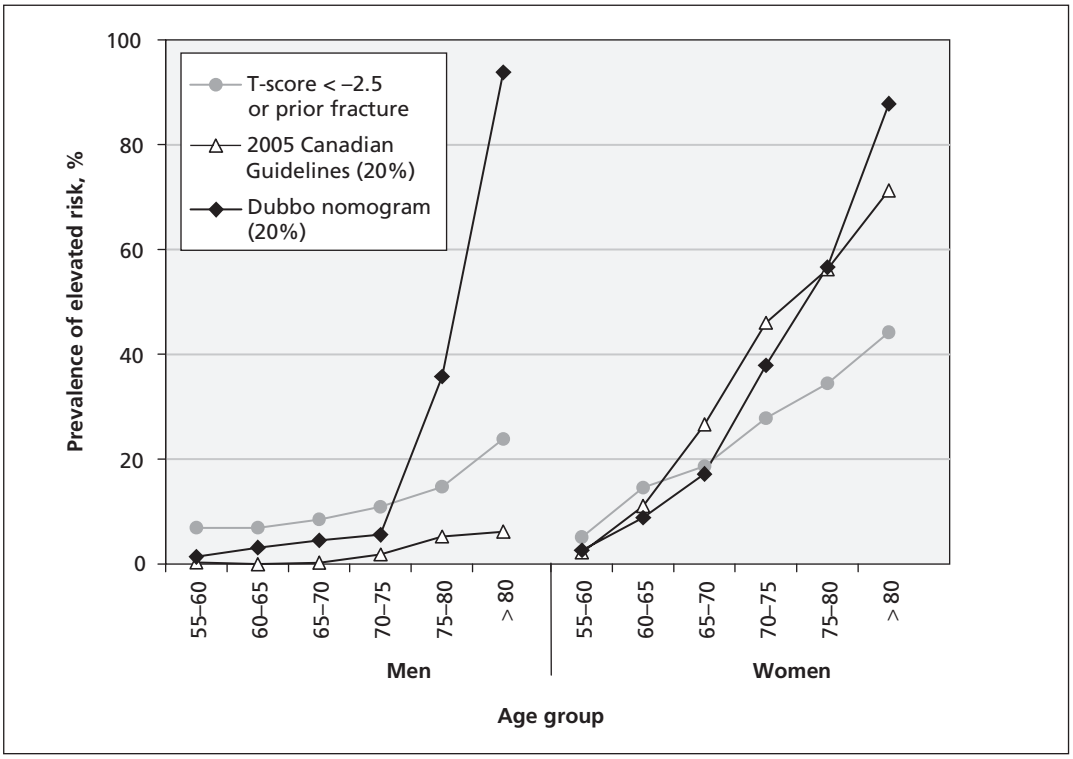

Figure 2: Comparison of the prevalence of elevated risk among participants in the Canadian Multicentre Osteoporosis Study cohort by age group, based on 1) femoral neck T-score ≤ –2.5 or prior fracture, 2) assessed 10-year risk ≥ 20% using Canadian guidelines for 2005 and 3) assessed 10-year risk ≥ 20% using the Dubbo nomogram. Summary estimates (*) were weighted according to Canada census population figures for 2006.

References

1. Siris ES, Chen YT, Abbott TA, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med* 2004;164:1108-12.
2. Pasco JA, Seeman E, Henry MJ, et al. The population burden of fractures originates in women with osteopenia, not osteoporosis. *Osteoporos Int* 2006;17:1404-9.
3. National Osteoporosis Foundation. *Clinician's guide to prevention and treatment of osteoporosis* [pamphlet]. Washington (DC): The Foundation; 2008.
4. Khosla S, Melton LJ III. Clinical practice. Osteopenia. *N Engl J Med* 2007;356:2293-300.
5. Siminoski K, Leslie WD, Frame H, et al. Recommendations for bone mineral density reporting in Canada. *Can Assoc Radiol J* 2005;56:178-88.
6. Leslie WD, Tsang JF, Lix LM. Simplified system for absolute fracture risk assessment: clinical validation in Canadian women. *J Bone Miner Res* 2009;24:353-60.
7. Nguyen T, Sambrook P, Kelly P, et al. Prediction of osteoporotic fractures by postural instability and bone density. *BMJ* 1993;307:1111-5.
8. Nguyen ND, Frost SA, Center JR, et al. Development of a nomogram for individualizing hip fracture risk in men and women. *Osteoporos Int* 2007;18:1109-17.
9. Nguyen ND, Frost SA, Center JR, et al. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. *Osteoporos Int* 2008;19:1431-44.
10. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605. doi: 10.1136/bmj.b605.
11. Kreiger N, Tenenhouse A, Joseph L, et al. Research notes: the Canadian Multicentre Osteoporosis Study (CaMos) — background, rationale, methods. *Can J Aging* 1999;18:376-87.
12. Tenenhouse A, Joseph L, Kreiger N, et al. Estimation of the prevalence of low bone density in Canadian women and men using a population-specific DXA reference standard: the Canadian Multicentre Osteoporosis Study (CaMos). *Osteoporos Int* 2000;11:897-904.
13. Genant HK. Universal standardization for dual X-ray absorptiometry: patient and phantom cross-calibration results. *J Bone Miner Res* 1995;10:997-8.
14. Berger C, Langsetmo L, Joseph L, et al. Change in bone mineral density as a function of age in women and men and association with the use of antiresorptive agents. *CMAJ* 2008;178:1660-8.
15. Looker AC, Wahner HW, Dunn WL, et al. Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int* 1998;8:468-89.
16. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
17. Aickin M. Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's kappa. *Biometrics* 1990;46:293-302.
18. Pencina MJ, D'Agostino RB Sr., D'Agostino RB Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-72.
19. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;54:17-23.
20. Kanis JA, Oden A, Johnell O, et al. 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;18:1033-46.
21. Leslie WD, Lix LM, Johansson H, et al. Independent clinical validation of a Canadian FRAX((R)) tool: Fracture prediction and model calibration. *J Bone Miner Res* 2010; Apr. 30 [Epub ahead of print].
22. Sandhu SK, Nguyen ND, Center JR, et al. Prognosis of fracture: evaluation of predictive accuracy of the FRAX algorithm and Garvan nomogram. *Osteoporos Int* 2010;21:863-71.
23. Ensrud KE, Lui LY, Taylor BC, et al. A comparison of prediction models for fractures in older women: is more better? *Arch Intern Med* 2009;169:2087-94.
24. Liao Y, McGee DL, Cooper RS, et al. How generalizable are coronary risk prediction models? Comparison of Framingham and two national cohorts. *Am Heart J* 1999;137:837-45.
25. Henry MJ, Pasco JA, Nicholson GC, et al. Prevalence of osteoporosis in Australian women: Geelong Osteoporosis Study. *J Clin Densitom* 2000;3:261-8.
26. Henry MJ, Pasco JA, Korn S, et al. Bone mineral density reference ranges for Australian men: Geelong Osteoporosis Study.

Osteoporos Int 2010;21:909-17.
27. Ismail AA, Pye SR, Cockerill WC, et al. Incidence of limb fracture across Europe: results from the European Prospective Osteoporosis Study (EPOS). *Osteoporos Int* 2002;13:565-71.
28. Leslie WD, O'Donnell S, Lagacé C, et al. Population-based Canadian hip fracture rates with international comparisons. *Osteoporos Int* 2010; 21:1317-22.
29. Chen Z, Kooperberg C, Pettinger MB, et al. Validity of self-report for fractures among a multiethnic cohort of postmenopausal women: results from the Women's Health Initiative observational study and clinical trials. *Menopause* 2004;11:264-74.
30. Leslie WD, O'Donnell S, Jean S, et al. Trends in hip fracture rates in Canada. *JAMA* 2009;302:883-9.
31. Kmetic A, Joseph L, Berger C, et al. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. *Epidemiology* 2002;13:437-44.

Competing interests: Nguyen Nguyen is supported by a grant from the Australian Medical Bioinformatics Resource (AMBeR). Christopher Kovacs has served as a consultant or received grants from Amgen, Eli Lilly, GlaxoSmithKline, Merck, Novartis, Proctor and Gamble, Sanofi-Aventis, Servier and Novo Nordisk. Jacqueline Center has been supported by or given educational talks for Eli Lilly, Merck Sharp and Dohme, and Sanofi-Aventis. Suzanne Morin has served as a consultant or received payment for lectures given to Warner-Chilcott, Amgen, Eli Lilly, Novartis and Merck. Robert Josse serves as a consultant for Amgen, Bayer, Eli Lilly, Glaxo-SmithKline, Merck, Novartis, Proctor and Gamble, Sanofi-Aventis, Servier and Wyeth-Ayerst. Jonathan Adachi served as a consultant or worked on clinical trials for Amgen, Astra Zeneca, Eli Lilly, GlaxoSmithKline, Merck Frosst, Novartis, Proctor and Gamble, Roche, Sanofi-Aventis, Servier, Pfizer and Wyeth-Ayerst. David Hanley serves as a consultant or has received grants from Amgen, Eli Lilly, Merck, Novartis, Proctor and Gamble, Sanofi-Aventis, Servier, Wyeth-Ayerst and Nycomed. John Eisman served as a consultant or worked on clinical trials for Amgen, Eli Lilly, Merck, Novartis, Sanofi-Aventis and Servier. No competing interests declared by Lisa Langsetmo, Tuan Nguyen or Jerilynn Prior.

Affiliations: From the CaMos National Coordinating Centre (Langsetmo), McGill University, Montréal, Que.; the Bone and Mineral Research Program (Nguyen T, Nguyen D, Center, Eisman), Garvan Institute of Medical Research, St Vincent's Hospital, Darlinghurst, Sydney, Australia; the Division of Endocrinology and Metabolism, Department of Medicine, Memorial University (Kovacs), St. John's NL; the Department of Medicine and Endocrinology, University of British Columbia (Prior), Vancouver, BC; the Department of Medicine, McGill University (Morin), Montréal, Que.; the Department of Medicine, University of Toronto (Josse), Toronto, Ont.; the Department of Medicine, McMaster University (Adachi), Hamilton, Ont.; and the Department of Medicine and Community Health, University of Calgary (Hanley), Calgary, Alta.

Contributors: Lisa Langsetmo, Tuan Nguyen, Nguyen Nguyen, Christopher Kovacs, Jerilynn Prior, David Hanley and John Eisman contributed to the design of the study. Lisa Langsetmo, Tuan Nguyen and John Eisman worked on the statistical analysis. All of the authors contributed to the interpretation of the data and the drafting and critical revision of the manuscript, and all of them approved the final version submitted for publication. Members of the Canadian Multicentre Osteoporosis Research Group were involved in the initial study design, recruitment of participants, data collection, quality control, review of projects, retention of cohort and other projects.

Funding: The Canadian Multicentre Osteoporosis Study was funded by the Canadian Institutes of Health Research (CIHR), Merck Frosst Canada, Eli Lilly Canada, Novartis, and the Alliance for Better Bone Health (Sanofi-Aventis, Procter and Gamble Pharmaceuticals Canada, Dairy Farmers of Canada and the Arthritis Society. These funding sources had no role in the conception of this analysis, in the statistical methods used or in the interpretation of the data.

Acknowledgement: The authors thank all those participants in CaMos whose careful responses and attendance made this analysis possible.

CaMos Research Group: David Goltzman (co-principal investigator, McGill University, Montréal, Que.), Nancy Kreiger (co-principal investigator, Toronto, Ont.), Alan Tenenhouse (principal investigator emeritus, Toronto, Ont.). CaMos Coordinating Centre, McGill University, Montreal, Que.: Suzette Poliquin (national coordinator), Suzanne Godmaire (research assistant), Claudie Berger (study statistician). Memorial University, St. John's, NL: Carol Joyce (director), Christopher Kovacs (co-director), Emma Sheppard (coordinator). Dalhousie University, Halifax, NS: Susan Kirkland, Stephanie Kaiser (co-directors), Barbara Stanfield (coordinator). Laval University, Québec City, Que.: Jacques P. Brown (director), Louis Bessette (co-director), Marc Gendreau (coordinator). Queen's University, Kingston, Ont.: Tassos Anastassiades (director), Tanveer Towheed (co-director), Barbara Matthews (coordinator). University of Toronto, Toronto, Ont.: Bob Josse (director), Sophie A. Jamal (co-director), Tim Murray (past director), Barbara Gardner-Bray (coordinator). McMaster University, Hamilton, Ont.: Jonathan D. Adachi (director), Alexandra Papaioannou (co-director), Laura Pickard (coordinator). University of Saskatchewan, Saskatoon, Sask.: Wojciech P. Olszynski (director), K. Shawn Davison (co-director), Jola Thingvold (coordinator). University of Calgary, Calgary, Alta.: David A. Hanley (director), Jane Allan (coordinator). University British Columbia, Vancouver, BC: Jerilynn C. Prior (director), Millan Patel (co-director), Yvette Vigna (coordinator); Brian C. Lentle (radiologist).