Journal of Bone and Mineral Research, 2025, **00**, 1–8
https://doi.org/10.1093/jbmr/zjaf113
Advance access publication: August 19, 2025
Research Article

Fracture risk scores using output from an opportunistic screen of low bone density from conventional X-ray

Catriona A. Syme[1,*], Mark D. Cicero[1,2], Jonathan D. Adachi[3], Claudie Berger[4], Suzanne N. Morin[4,5], David Goltzman[4,5], Alexander Bilbily[1,6]

[1]Department of Research, 16 Bit Inc., Toronto, ON, M5J 2N8 Canada
[2]Diagnostic Radiology, True North Imaging, Toronto, ON, L4J 7Y7 Canada
[3]Division of Rheumatology, Department of Medicine, McMaster University, Hamilton, ON, L8N 3Z5 Canada
[4]Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, QC, H3H 2R9 Canada
[5]Department of Medicine, McGill University, Montreal, QC, H4A 3J1 Canada
[6]Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, M4N 3M5 Canada

*Corresponding author: Catriona A. Syme, Department of Research, 16 Bit Inc., 20 Bay Street, 11th Floor, Toronto, ON M5J 2N8, Canada (catriona@16bit.ai).

Abstract

Fracture risk is commonly assessed by FRAX, a tool that estimates 10-yr risk for major osteoporotic fracture (MOF) and hip fracture. FRAX scores are often refined by including FN BMD measured by DXA as an input. Rho, a novel AI-powered software, estimates FN BMD T-Scores from conventional X-rays, even when FN is not in the image. Whether a FRAX score using this estimate (FRAX-Rho) can improve a FRAX score without a T-Score input (FRAX-NoT) has not been studied. We conducted a retrospective analysis of Canadian Multicentre Osteoporosis Study participants who had X-rays of the lumbar and/or thoracic spine, FRAX risk factors, and DXA T-Scores acquired at the same time point, and follow-up fracture outcomes over 9 yr. In 1361 participants with lumbar X-rays, FRAX-Rho and FRAX with DXA FN T-Scores (FRAX-DXA) had very good agreement in categorizing participants by MOF risk (Cohen's weighted kappa $\kappa = 0.80$ [0.77-0.82]), which tended to be better than that between FRAX-NoT and FRAX-DXA (0.76 [0.73-0.79]). Agreement in categorizing participants by hip fracture risk was significantly greater between FRAX-Rho and FRAX-DXA (0.67 [0.63-0.71]) than FRAX-NoT and FRAX-DXA (0.52 [0.48-0.56]). In predicting true incident MOF, FRAX-Rho and FRAX-DXA did not differ in their discriminative power (c-index) (0.76 and 0.77; $p = .36$); both were significantly greater than that of FRAX-NoT (0.73; $p < .004$). The accuracy of FRAX-Rho for predicting MOF (Brier Score) was better than FRAX-NoT ($p < .05$) but not as good as FRAX-DXA. Similar results were observed in participants with thoracic X-rays. In conclusion, FN T-Scores estimated by Rho from lumbar and thoracic X-rays add value to FRAX-NoT estimates and may be useful for risk assessment when DXA is not available.

Keywords: osteoporosis, fracture risk, artificial intelligence, X-ray, DXA

Lay Summary

Risk of a future fracture is commonly calculated to identify patients who might benefit from lifestyle modification or treatment to optimize bone health. Other studies have demonstrated that fracture risk calculations are better when they include the patient's BMD, measured by a special X-ray system called DXA. A novel software powered by artificial intelligence, Rho, can estimate the patient's BMD from a regular X-ray. This study shows that fracture risk calculations with no estimate of BMD are improved when they include the Rho-estimated BMD. Rho-estimated BMD may be useful for risk assessment when DXA is not available.

Introduction

Osteoporosis is a common, chronic disease characterized by low BMD and deterioration of bone tissue that increases an individual's risk of fracture. In adults over 50 yr of age, 6% of men and 21% of women have osteoporosis[1] and 1 in 5 men and 1 in 3 women will sustain an osteoporotic fracture in their lifetime.[2] These preventable fractures lead to significant morbidity and mortality, and health-care costs.[3]

Among several fracture risk prediction tools available, FRAX[4] has the largest number of externally validated and independent studies. Since its release, 71 models have been made available for 66 countries covering more than 80% of the global population.[5]

FRAX uses a number of clinical risk factors to estimate a 10-yr probability of a hip fracture and of major osteoporotic fracture (MOF; ie, osteoporotic fracture of the spine, hip, forearm, or humerus). It requires as input: age, sex, and clinical fracture risk factors (eg, BMI, previous fragility fracture, current smoking, and alcohol use) and, optionally, a BMD T-Score from the FN. Inclusion of a T-Score (acquired by DXA) increases the FRAX tool's accuracy.[6,7]

Rho (16 Bit Inc.) is a software-as-a-medical device licensed by Health Canada and authorized by the U.S. Food and Drug Administration that serves as an opportunistic pre-screen for low BMD from conventional X-ray. Rho automatically analyzes X-rays of the chest, LS, thoracic spine, pelvis, hand

Received: November 12, 2024. **Revised:** February 10, 2025. **Accepted:** August 6, 2025

or knee, acquired for any clinical indication in patients aged 50 yr and older. The Rho algorithm estimates BMD at L1-L4 and at the FN, even if those regions are not in the analyzed X-ray. It uses a locked machine algorithm, trained with over 67 000 pairs of X-rays (of the chest, LS, thoracic spine, pelvis, hand/wrist, or knee) and DXA (which included the ground truth BMD at L1-L4 and at the FN) that learned to predict BMD at L1-L4 and at the FN from features of the X-ray.[8] It uses these estimates to output an alert to a healthcare provider if a patient may have low BMD. Rho is not DXA-equivalent. Currently, clinical implementation of a screening tool like Rho, which takes advantage of already-available X-ray scans (acquired for any reason), is intended to prompt a clinical fracture risk assessment to help improve osteoporosis management.

This study aimed to determine whether a Rho-derived FN T-Score estimate, not currently a regulatory-approved output of the software, could add clinical value to a FRAX score calculated without a DXA FN T-Score. If so, it could improve fracture risk prediction in the absence of DXA.

Materials and methods

This is a retrospective study that utilized data collected from a large prospective study, the Canadian Multicentre Osteoporosis Study (CaMos).[9] CaMos was a nationwide prospective cohort study of 9423 non-institutionalized, randomly selected women ($n = 6539$) and men ($n = 2884$) aged 25 years or older. Recruitment was conducted between 1995 and 1997 and participants were followed for 19 yr. Informed consent was obtained from each participant, and the study was approved by the Institutional Review Board of each participating institution. In order to take advantage of more recent X-ray technology, the present study analyzed a subset of participants from year 10 ($n = 4921$), collected between 2005 and 2008. As part of CaMos, participants aged 50 and above had X-rays of the thoracic and LS, and a DXA scan from which T-Scores of L1-L4 and FN were available. BMD was measured using Hologic QDR 1000 or 2000 or Lunar DPX densitometers at each institution that had been cross-calibrated at the start of the study and yearly thereafter. BMD values from Lunar systems were converted to a Hologic base. Of these 4921 participants, 2701 unique individuals had X-rays. Rho has a built-in quality control (QC) system that classifies adequate frontal projections of candidate X-rays; 1798 unique individuals had X-rays that passed Rho's QC. The high rate of failed QC X-rays was due largely to X-ray acquisition parameters because some X-rays were lateral projections and Rho only accepts frontal projections.

The following FRAX inputs for fracture risk calculations were collected by a study coordinator at an in-person study visit: age, sex, BMI, previous fracture, parent hip fracture, current smoking, glucocorticoid use, rheumatoid arthritis, secondary osteoporosis, and alcohol. Of the 1798 unique individuals with eligible X-rays, 11 did not have complete FRAX inputs, thus the final dataset for analysis thus included 1787 unique individuals. Optionally, the FRAX score can be refined by inputting a DXA-derived FN T-Score. FRAX scores for MOF and hip fracture were calculated using no T-Scores (FRAX-NoT), DXA T-Scores (FRAX-DXA) and Rho-derived FN T-Scores (FRAX-Rho) from either a lumbar or thoracic X-ray.

Participants were followed longitudinally with yearly questionnaires that included information on incident fracture (hip or MOF). Surveys were reviewed to determine the number of years from time of data acquisition (year 10 of CaMos) that incident fractures occurred. Incident fractures were categorized as either major osteoporotic (hip, spine, forearm, or humerus) and/or hip if the mechanism was reported to be low trauma.

Sample size

For sample size calculations for Cohen's Kappa,[10] an alpha of .05 was Bonferroni corrected by a factor of 8 (2 types of fracture concordance tests [hip and MOF] × 2 pairs of agreement [FRAX-NoT with FRAX-DXA and FRAX-Rho with FRAX-DXA] × 2 types of X-rays [lumbar and thoracic]).

To identify concordance between FRAX estimates to categorize participants into 3 MOF risk categories of low (<10.0%), moderate (10.0%-19.9%), or high (≥20.0%) requires an estimate of the frequency of participants in each category. From CaMos data, the percentage of community-based participants in each risk category (low, moderate, and high) was 69%, 23%, and 8%, respectively.[11] A sample size calculation using these probabilities was performed for a null hypothesis of $\kappa = 0.7$, an alternative hypothesis of $\kappa = 0.8$, with 80% power and alpha = .05/8. The minimum sample size to assess concordance for categorizing MOF is $n = 676$.

To identify concordance between FRAX estimates to categorize participants into 2 hip fracture risk categories: low (<3%) or high (>3%), we used the frequency of participants in each category from a community-based study from Manitoba, of approximate mean age 68 yr, that found 16% of participants had a high risk of hip fracture.[12] A sample size calculation using this probability was performed for a null hypothesis of $\kappa = 0.7$, an alternative hypothesis of $\kappa = 0.8$, with 80% power and alpha = .05/8. The minimum sample size to assess concordance for categorizing hip fracture is $n = 1196$.

Statistical analysis

Descriptive statistics stratified by sex were used to describe the analytic cohort. Continuous variables were described with mean and SD, with between-group comparisons analyzed using Kruskal–Wallis tests. Categorical variables were described with the number and percent of each attribute, with between-group comparisons analyzed using Pearson's Chi-squared test.

A quadratic weighted kappa and 95% CI were calculated to estimate the chance-corrected agreement between FRAX estimates.[13] Tables (3 × 3) were constructed to measure concordance between FRAX-DXA and FRAX-Rho for categorizing participants in three MOF risk categories. Tables (2 × 2) were constructed for hip fracture risk categories. A Cohen's kappa coefficient $\kappa > 0.75$ is considered very good,[14] and $\kappa > 0.80$ is considered excellent agreement.[13]

Questionnaire data for each year following the X-ray and DXA collection were reviewed for the occurrence of an event, either MOF or death, and time to event, in years, was recorded. If no event occurred, the time to event was the last completed questionnaire and the event was censored. We calculated C-index, the discriminative power of the risk predictive model, using the concordance function on Cox proportional hazards regression models using the R package *survival*. The response variable (time to MOF or hip fracture, and whether or not the fracture occurred) was modeled against each FRAX estimate (ie, FRAX-DXA, FRAX-RHO, or FRAX-NoT). Discrimination performance

Table 1. Participant characteristics.

	Males ($N = 517$)	Females ($N = 1270$)	*p*-value
Age, years	68.6 (9.4)	70.0 (9.3)	.005
BMI	27.6 (4.1)	27.4 (5.1)	.061
Current smoking	64 (12.4)	114 (9.0)	.011
Alcohol	17 (1.2)	7 (0.6)	<.001
Corticosteroids	6 (1.2)	35 (2.8)	.041
Parental hip fracture	55 (10.6)	145 (11.4)	.64
Prior fracture	79 (15.3)	287 (22.6)	<.001
Secondary osteoporosis	16 (3.1)	69 (5.4)	.035
Rheumatoid arthritis diagnosis/treatment	2 (0.4)	14 (1.1)	.145
DXA T-score FN	−0.528 (1.103)	−1.310 (1.054)	<.001
FRAX-NoT MOF	6.9 (3.9)	14.9 (9.9)	<.001
FRAX-DXA MOF	6.2 (4.0)	12.3 (8.3)	<.001
FRAX-Rho MOF	7.0 (4.1)	13.3 (8.3)	<.001
FRAX-NoT hip	2.3 (3.0)	5.2 (6.5)	<.001
FRAX-DXA hip	1.7 (2.9)	3.2 (5.2)	<.001
FRAX-Rho hip	2.1 (3.2)	3.7 (5.1)	<.001

Data are mean ± SD or *N* (%).

of FRAX refers to its ability to predict who would fracture earlier and who would fracture later. Higher values of the C-index indicate better discrimination performance of the model. We compared correlated C-indices (eg, FRAX-DXA vs FRAX-RHO) using the R package compareC. Brier Score, the average squared distances between the observed survival status and the predicted survival probability, was calculated using incident fracture as the ground truth outcome (yes/no) for MOF. The 95% CIs were calculated by bootstrapping (B = 1000). It is a measure of overall performance of the model.

Statistical analyses were performed by C.S., in R, version 4.2.2.

Research ethics board approval

Participants of the CaMos study provided informed consent and the study was approved by the Institutional Review Board of each participating institution. For the current study analyzing CaMos data, research ethics board approval was obtained from an independent review board (Veritas) with a waiver of informed consent.

Results

Participant characteristics

The dataset included unique participants with at least one frontal X-ray available and FRAX inputs available ($n = 1787$). The dataset included more females than males (Table 1). Females vs males were older (70.0 [9.3] vs 68.6 [9.4]), had a higher rate of prior fractures (22.6% vs 15.3%), lower T-Score estimates (−1.310 [1.054] vs −0.528 [1.103]), and higher FRAX scores by all methods of calculation.

Of these participants, some participants had only a lumbar X-ray, only a thoracic X-ray, or both X-rays available. About 1361 unique individuals had a lumbar X-ray with complete FRAX input (excluding DXA FN T-Score); and 1324 of those had DXA FN T-Scores available. A total of 1507 unique individuals had thoracic X-rays and FRAX inputs; 1467 of those had DXA T-Scores available. Follow-up data were available to ascertain if participants sustained an MOF in the 9 yr following data collection (Table 2). A flowchart is provided in the Supplementary Material (Figure S1). Females vs males had a higher rate of incident MOF (7% vs 3%).

The following results are for FRAX-Rho derived from lumbar X-rays, as, in the clinical experience of the authors, lumbar X-rays are more commonly performed in routine clinical practice than thoracic X-rays. Results for FRAX-Rho derived from thoracic X-rays are presented in the Supplementary Material.

Agreement between FRAX indices at categorizing MOF risk

Participants were classified as being at low-, moderate-, or high-MOF risk according to each of the FRAX indices, and the agreement of FRAX pairs was evaluated using a 3 × 3 confusion matrix. FRAX-NoT vs FRAX-DXA had very good agreement (0.76 [0.73-0.78]; Table S1), and the agreement between FRAX-Rho and FRAX-DXA was slightly greater (0.79 [0.77-0.82]; Table S3).

Incident MOF rate in FRAX-NoT and FRAX-Rho risk categories

In the 1361 participants with lumbar X-rays, over the course of a 9-yr follow-up, 81 (6%) sustained an incident MOF within the follow-up period of 9-yr and 148 (11%) participants died. To compare FRAX-Rho and FRAX-NoT, we constructed a confusion matrix of the MOF risk categories (Table 3). Each cell shows the number of participants who sustained incident MOF (numerator), total number of participants (denominator), and fracture rate in parentheses. Overall, MOF rate was similar between FRAX-Rho and FRAX-NoT in the low-risk groups (2% and 3%, respectively), moderate risk groups (8% and 6%, respectively), and high-risk groups (14% in both). FRAX-Rho upgraded the risk category from low by FRAX-NoT to moderate ($n = 56$) or high ($n = 2$) and downgraded the risk category to low from moderate ($n = 98$) or high ($n = 2$).

Incident MOF rate in discordant FRAX-Rho and FRAX-NoT risk categories

Participants were next grouped by whether their risk would have been upgraded from low by FRAX-NoT to moderate or high by FRAX-Rho, or downgraded from moderate or high by FRAX-NoT to low by FRAX-Rho (Table 4). Of downgraded participants, 51/100 did not sustain an MOF

Table 2. Available X-rays and 9-yr incident fracture.

	Lumbar ($N = 1361$)		Thoracic ($N = 1507$)	
	Male ($N = 401$)	Female ($N = 960$)	Male ($N = 443$)	Female ($N = 1064$)
Major Osteoporotic Fracture within 9 yr				
No (%)	340 (84.8)	792 (82.5)	380 (85.8)	875 (82.2)
Yes (%)	13 (3.2)	68 (7.1)[a]	15 (3.4)	77 (7.2)[b]
Deceased (%)	48 (12.0)	100 (10.4)	48 (10.8)	112 (10.5)
Hip Fracture within 9 yr				
No (%)	347 (86.5)	831 (86.6)	388 (87.6)	926 (87.0)
Yes (%)	5 (1.2)	20 (2.1)	7 (1.6)	18 (1.7)
Deceased (%)	49 (12.2)	109 (11.4)	48 (10.8)	120 (11.3)

Females vs males had higher percentage of incident MOFs, [a]$p = .020$, [b]$p = .017$.

Table 3. MOF rate according to FRAX-NoT and FRAX-Rho from participants with lumbar X-rays.

		FRAX-NoT MOF Category (*n* [%])			
		Low	Moderate	High	Total
FRAX-Rho MOF Category (*n* [%])	**Low**	13/641 (2%)	3/98 (3%)	0/2 (0%)	16/741 (2%)
	Moderate	9/56 (16%)	18/263 (7%)	6/76 (8%)	33/395 (8%)
	High	0/2 (0%)	2/35 (6%)	30/188 (16%)	32/225 (14%)
	Total	22/699 (3%)	23/396 (6%)	36/266 (14%)	81/1361 (6%)

Table 4. Follow-up events in participants initially stratified with FRAX-NoT with modified risk by FRAX-Rho (from lumbar X-rays).

Event	Risk downgraded to low by FRAX-Rho ($n = 100$)	Risk upgraded from low to moderate or high by FRAX-Rho ($n = 58$)
No MOF (*n* [%])	51 (51)	32 (55)
MOF (*n* [%])	3 (3)	9 (16)
Death (*n* [%])	17 (17)	6 (10)
Lost to follow-up (*n* [%])	29 (29)	11 (19)

within 9 yr; these could be considered "true negatives" for Rho as they were considered lower risk and truly did not fracture. 3/100 did sustain an MOF; these could be considered "false negatives" for Rho as they were considered lower risk but this classification was not correct, as they did go on to fracture. Of upgraded participants, 32/58 did not sustain an MOF; these could be considered "false positives" for Rho as they were considered at higher risk but this classification was not correct as they did not fracture. 9/58 did sustain an MOF; these could be considered "true positives" for Rho as they were considered higher risk and truly did fracture. Death and loss to follow-up events are also shown as competing risks. The median (first quartile—third quartile) time to loss of follow-up was 5 (5-6) yr and 5 (5-8) yr in downgraded and upgraded participants, respectively, indicating that on average, participants lost to follow-up had not sustained an MOF in a 5-yr period following initial data collection. A X^2 test comparing rates of events in participants who would have had their risk category upgraded vs downgraded by FRAX-Rho (vs FRAX-NoT) was significant ($p = .013$). The proportion of participants who went on to have an MOF in those upgraded by Rho (9/58; 16%) was greater than in those downgraded by Rho (3/100; 3%; $p = .011$), and the proportion of participants who went on to have an MOF in the downgraded by Rho group was similar to the proportion in the category assigned low risk by both FRAX-NoT and FRAX-Rho (13/641; 2%).

Assuming a moderate- or high-risk classification would lead to screening or therapy, FRAX-Rho vs FRAX-NoT would have decreased the number of participants screened or treated by $100 - 58 = 42$ while correctly identifying an additional 6 participants (9-3) who, because they went on to have an MOF, would have benefited from that screening or therapy.

Incident MOF rate in FRAX-NoT vs FRAX-DXA risk categories

Similar comparisons of incident MOF rate in MOF risk categories were made for FRAX-NoT vs FRAX-DXA (Tables S1 and S2). FRAX-DXA upgraded the risk category from low by FRAX-NoT to moderate ($n = 36$) and downgraded the risk category to low from moderate ($n = 158$) or high ($n = 7$). As expected, those participants categorized as low-risk by FRAX-NoT that would be upgraded to higher risk by including DXA in the FRAX calculation had a significantly higher rate of true incident MOFs (6/36; 17%) vs those participants with moderate- or high-risk by FRAX-NoT that would be downgraded to low risk by including DXA in the FRAX calculation (6/165; 4%; $p = .009$). Overall, FRAX-DXA vs FRAX-NoT would have decreased the number of participants screened or treated by $165\text{-}36 = 129$ and would have (correctly) upgraded the risk category of 6 participants who went on to have an MOF, but simultaneously (incorrectly) downgraded the

risk category of 6 participants who went on to have an incident MOF.

Incident MOF rate in FRAX-Rho vs FRAX-DXA risk categories

FRAX-DXA vs FRAX-Rho upgraded the risk category from low by FRAX-Rho to moderate ($n = 47$; Table S3) and downgraded the risk category to low from moderate ($n = 127$) or high ($n = 5$). Those participants low-risk by FRAX-Rho that would be upgraded to higher risk by including DXA in the FRAX calculation (instead of Rho FN) had a similar rate of true incident MOFs (4/47; 9%; Table S4) vs those participants with moderate- or high-risk by FRAX-Rho that were downgraded to low risk by including DXA in the FRAX calculation (10/132; 8%; $p = 1$). Overall, FRAX-DXA vs FRAX-Rho would have decreased the number of participants screened or treated by $132 - 47 = 85$, and would have (correctly) upgraded the risk category of 4 participants who went on to have an MOF, but simultaneously (incorrectly) downgraded the risk category of 10 participants who went on to have an incident MOF.

Agreement between FRAX indices at categorizing hip fracture risk

Participants were classified as being at low- or high-hip fracture risk according to each of the FRAX indices, and the agreement of FRAX pairs was evaluated using a 2 × 2 confusion matrix. Agreement between FRAX-Rho and FRAX-DXA (0.67 [0.63-0.71]) was significantly greater than the agreement between FRAX-NoT and FRAX-DXA (0.52 [0.47-0.56]). When FRAX included the Rho FN T-Score (vs no T-Score), accuracy and specificity were both improved (0.86 [0.83-0.87] vs 0.77 [0.75-0.79] and 0.84 vs 0.72, respectively), and sensitivity was similar (0.91 vs 0.90; Table S5).

Incident hip fracture rate in FRAX-NoT and FRAX-Rho risk categories

In these 1361 participants with lumbar X-rays, over the course of a 9-yr follow-up, 25 (2%) sustained an incident hip fracture within the follow-up period of 9-yr; and 158 (12%) participants died. Note that in the breakdown of events for MOF, 148 participants died; this means 10 participants in that dataset suffered a non-hip MOF before their death, thus were counted as MOF events in that dataset as opposed to death events, while in this dataset they are counted as death events, bringing the total death events to 158. To compare FRAX-Rho and FRAX-NoT, we constructed a confusion matrix of the hip fracture risk categories (Table S6). Hip fracture rate was similar between FRAX-Rho and FRAX-NoT in the low-risk groups (0.6% and 0.4%, respectively) and high-risk groups (4% in both). FRAX-Rho upgraded the risk category from low by FRAX-NoT to high in 53 participants, and downgraded the risk category to low from high by FRAX-NoT in 161 participants. A X^2 test comparing rates of events in participants who would have had their risk category upgraded vs downgraded by FRAX-Rho (vs FRAX-NoT) was not significant ($p = .36$; Table S7). The proportion of participants who went on to have a hip fracture did not differ in those upgraded by Rho (0/53; 0%) vs in those downgraded by Rho (2/161; 1%; $p = 1$).

Incident hip fracture rate in FRAX-NoT vs FRAX-DXA risk categories

Similar comparisons of incident hip fracture rate in hip fracture risk categories were made for FRAX-NoT vs FRAX-DXA (Tables S8 and S9). FRAX-DXA upgraded the risk category from low by FRAX-NoT to high ($n = 35$), and downgraded the risk category to low from high ($n = 269$). Those participants low-risk by FRAX-NoT that would be upgraded to moderate or high risk by including DXA in the FRAX calculation had a similar rate of true incident hip fractures (0/35; 0%) vs those participants high-risk by FRAX-NoT that would be downgraded to low risk by including DXA in the FRAX calculation (3/269; 1%).

Incident hip fracture rate in FRAX-Rho vs FRAX-DXA risk categories

FRAX-DXA vs FRAX-Rho upgraded the risk category from low by FRAX-Rho to high ($n = 33$; Table S10), and downgraded the risk category to low from high ($n = 159$). Those participants low-risk by FRAX-Rho that would be upgraded to high risk by including DXA in the FRAX calculation (instead of Rho FN) had a similar fraction of hip fractures (0/33; 0%; Table S11) vs those participants high-risk by FRAX-Rho that were downgraded to low risk by including DXA in the FRAX calculation (1/159; 0.6%).

Agreement between FRAX indices at categorizing MOF and hip fracture risk in participants with thoracic X-rays are presented in the Supplementary Material (Tables S12 -S24).

Predicting incident fractures

The C-index, which is a measure of the discriminative power of the risk predictive model, defined as the proportion of concordant pairs divided by the total number of possible evaluation pairs, was calculated using true incident MOF (within 9 yr). Values near 1 indicate high performance, and a value of 0.5 indicates that the discrimination performance of the model is the same as a coin flip (random concordance) in predicting which patient will fracture.

Of participants with lumbar X-rays *and FRAX-DXA available* ($n = 1324$), 77 sustained an incident MOF (24 were hip fractures) and 143 died. Results of all cox proportional hazards model are presented in the Supplementary Material. The best (highest c-score) discriminating risk factor for MOF was FRAX-DXA (0.77; Table 5). Discrimination did not differ statistically between FRAX-Rho (0.76) and FRAX-DXA ($p = .36$), and both FRAX-Rho and FRAX-DXA were statistically significantly higher than that of FRAX-NoT. The highest discriminating risk factor for hip fracture was FRAX-DXA (0.85), which was statistically significantly higher than that of FRAX-NoT. FRAX-Rho did not differ significantly from either FRAX-DXA or FRAX-NoT. In predicting true incident MOF, DXA-FN T and Rho-FN T alone (ie, no other FRAX inputs) did not differ in their discriminative power (0.23 and 0.25; $p = .46$) and both were significantly lower than FRAX-NoT that considers only the other risk factors (0.73; $p < 2.2E\text{-}16$ for both). For true incident hip fracture, their c-indices were lower than for MOF (0.16 and 0.20; $p = .29$) and both were lower than FRAX-NoT ($p < 1E\text{-}14$ for both).

The accuracy for predicting incident MOF was best for FRAX-DXA (lowest Brier score; Table 6). FRAX-DXA had a better Brier Score than FRAX-Rho or FRAX-NoT for both

Table 5. C-index of FRAX risk for incident MOF and hip fractures in participants with lumbar X-rays.

Fracture	FRAX	C-index	Concordant Pairs	Comparable Pairs	Tied Risk	*p*-val vs FRAX-NoT	*p*-val vs FRAX-DXA
MOF	FRAX-NoT	0.73 (0.67-0.79)	50 009	689 150	0	N/A	0.001
	FRAX-DXA	0.77 (0.72-0.83)	52 716	689 150	0	0.001	N/A
	FRAX-Rho	0.76 (0.71-0.82)	52 032	689 150	0	0.004	0.36
Hip	FRAX-NoT	0.80 (0.72-0.88)	18 048	22 536	0	N/A	0.004
	FRAX-DXA	0.85 (0.79-0.91)	19 146	22 536	0	0.004	N/A
	FRAX-Rho	0.82 (0.73-0.90)	18 321	22 536	0	0.60	0.14

Table 6. Brier score for predicting incident MOF in participants with lumbar X-rays.

Fracture	FRAX	Brier score	*t*-stat vs FRAX-NoT	*t*-stat vs FRAX-DXA
MOF	FRAX-NoT	0.060 (0.052-0.069)	N/A	6.27[a]
	FRAX-DXA	0.055 (0.047-0.064)	6.27[a]	N/A
	FRAX-Rho	0.058 (0.049-0.067)	2.73[a]	4.62[a]
Hip	FRAX-NoT	0.019 (0.013-0.025)	N/A	2.56[a]
	FRAX-DXA	0.018 (0.012-0.024)	2.56[a]	N/A
	FRAX-Rho	0.019 (0.013-0.025)	0.77	2.19[a]

[a] indicates $p < .05$.

MOF and hip fractures ($p < .05$ for both). FRAX-Rho had a better Brier Score than FRAX-NoT for MOF ($p < .05$) but not for hip fracture.

Results were similar when FRAX-Rho was derived from thoracic X-rays instead of from lumbar X-rays. Of participants with thoracic x-rays and FRAX-DXA available ($n = 1467$), 88 sustained an incident MOF (24 were hip fractures) and 156 died. Discrimination (C-index) for MOF and hip fracture was greatest for FRAX-DXA and weakest for FRAX-NoT, though the differences were not statistically significant (Table S25). FRAX-DXA had a better Brier Score than FRAX-NoT for both MOF and hip fractures ($p < .05$ for both; Table S26). FRAX-Rho had a better Brier Score than FRAX-NoT for MOF and hip fracture ($p < .05$ for both).

Discussion

This study suggests that a FN T-Score estimated by Rho, whether from lumbar or thoracic X-ray, adds value to a FRAX score calculated without DXA BMD. In the analyzed dataset, FRAX-Rho has the same or better agreement with FRAX-DXA than FRAX-NoT, for both MOF and hip risk categorization. At the time of FRAX score calculation, participants at low-MOF risk by FRAX-NoT who were moderate- or high-risk by FRAX-Rho had a higher rate of true incident MOF than vice versa, suggesting FRAX-Rho correctly upgraded the risk category in these participants.

Assuming moderate- or high-risk for MOF would trigger monitoring and/or treatment, FRAX score using FRAX-Rho rather than FRAX-NoT would have led to fewer participants overall being flagged for monitoring and/or treatment, and a greater proportion of participants who went on to incur an MOF would have been flagged. Further assuming efficacy of monitoring and/or treatment at fracture prevention, using FRAX-Rho rather than FRAX-NoT could lead to a greater number of participants avoiding an MOF. Together, these findings suggest that using FRAX-Rho could offer a cost benefit over using FRAX-NoT by triggering treatment or monitoring in fewer participants overall, while flagging more participants who would go on to fracture and, potentially, preventing that fracture. Given the importance of appropriate fracture risk stratification, additional studies with FRAX-Rho and fracture outcomes are needed. If results support that Rho can refine patient selection for DXA monitoring and/or treatment, a formal cost-effectiveness analysis would be needed to ascertain potential cost savings and utility in a healthcare jurisdiction.

In the same participants, FRAX-DXA rather than FRAX-NoT would also have led to fewer participants overall being flagged for monitoring and/or treatment, and a greater proportion of participants who went on to incur an MOF would have been flagged. When comparing FRAX-DXA and FRAX-Rho, FRAX-DXA would have led to fewer participants overall being flagged for monitoring and/or treatment, but the proportion of participants who went on to incur an MOF did not differ between those with risk upgraded or downgraded by FRAX-DXA. Of note, we reviewed 16 X-rays with discordant categorization of MOF or hip fracture risk when calculated by FRAX-DXA vs FRAX-Rho. Of the X-rays selected, both sexes were represented, there were no clear differences in age, BMI values, or the presence of degenerative changes between patients with discordant risk results and the remainder of the dataset.

In terms of fracture prediction, FRAX-DXA had the most discriminative power of the various FRAX indices to rank order patients according to their risk of MOF or hip fracture (c-index) and the most accurate probabilistic prediction of MOF and hip fracture (Brier score). FRAX-Rho from lumbar X-rays had greater discriminative power and was more accurate than FRAX-NoT for MOF and did not differ statistically in either measure for hip fracture. FRAX-Rho from thoracic X-rays did not differ statistically from FRAX-NoT in terms of discriminative power for MOF or hip fracture but was more accurate than FRAX-NoT for both types of fracture.

Strengths of the current study include the large, community-based dataset with fracture outcome data, and demonstration of the value of a Rho-derived FN estimate derived from two

different X-ray types. Limitations include that both lumbar and thoracic X-rays will have inherent degenerative change, thus other Rho-eligible X-ray body parts (eg, chest, pelvis) would be valuable to study, and that the observation period was 9 yr while FRAX prediction is for 10-yr fracture risk. Although this limitation could influence the calculated results, it is unlikely to alter the conclusions. Finally, fracture questionnaires were designed to collect information on fracture site, fracture number, circumstance, and treatment, but radiology or medical reports to further adjudicate the fractures (eg, hip vs non-hip leg) were only obtained for 78% of all incident fractures. When relevant records could not be obtained, CaMos investigators confirmed self-reported fractures in a telephone interview.

FRAX-Rho, when derived from lumbar or thoracic spine X-rays, offers an improved fracture risk estimate over FRAX-NoT in the CaMos study population. Additional evidence to support its potential utility is required in additional populations, ideally with fracture outcomes. With such evidence, FRAX risk score estimated using information derived from Rho could be used in situations when DXA is not available or has not yet been performed.

Acknowledgments

Amgen Canada Inc. and the INOVAIT FOCUS Fund provided direct funding for this study. The Canadian Multicentre Osteoporosis Study (CaMos) has received support from the Canadian Institutes of Health Research (CIHR); Amgen Canada Inc.; Actavis Pharma Inc. (previously Warner Chilcott Canada Co.); Dairy Farmers of Canada; Eli Lilly Canada Inc.: Eli Lilly and Company; GE Lunar; Hologic Inc.; Merck Frosst Canada Ltd.; Novartis Pharmaceuticals Canada Inc.; P&G Pharmaceuticals Canada Inc.; Pfizer Canada Inc.; Roche (F. Hoffmann-La Roche Ltd.); Sanofi-Aventis Canada Inc. (previously Aventis Pharma Inc.); Servier Canada Inc.; and The Arthritis Society.

CaMos Research Group:

David Goltzman (co-principal investigator, McGill University), Nancy Kreiger (co-principal investigator, University of Toronto), and Alan Tenenhouse (principal investigator emeritus, Toronto).

McGill University, Montreal, Quebec: Elham Rahme (biostatistician), J. Brent Richards (investigator), Suzanne N. Morin (investigator); Data Analysis Centre: Claudie Berger (study statistician).

Memorial University, *St. John's Newfoundland*: Carol Joyce (director), Christopher S. Kovacs (co-director).

Dalhousie University, *Halifax, Nova Scotia*: Susan Kirkland, Stephanie M. Kaiser (co-directors).

Laval University, *Quebec City, Quebec*: Jacques P. Brown (director), Louis Bessette (co-director), *GRMO*.

Queen's University, *Kingston, Ontario*: Tassos P. Anastassiades (director), Tanveer Towheed (co-director), Wilma M. Hopman (investigator).

University of Toronto, *Toronto, Ontario*: Angela M. Cheung (director), Robert G. Josse (co-director), Andy Kin On Wong (co-director).

McMaster University, *Hamilton, Ontario*: Jonathan D. Adachi (director), Alexandra Papaioannou (co-director).

University of Saskatchewan, *Saskatoon, Saskatchewan*: Wojciech P. Olszynski (director), K. Shawn Davison (co-director).

University of Calgary, *Calgary, Alberta*: David A. Hanley (director), Steven K. Boyd (co-director).

University of British Columbia, *Vancouver, British Columbia*: Jerilynn C. Prior (director), Shirin Kalyan (co-director), Brian Lentle (investigator/radiologist), Millan S. Patel (investigator).

University of Alberta, *Edmonton, Alberta*: Stuart D. Jackson (medical physicist).

University of Manitoba, *Winnipeg, Manitoba*: William D. Leslie (investigator/nuclear medicine physician).

Author contributions

Catriona A. Syme (Conceptualization, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review & editing), Mark D. Cicero (Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—review & editing), Jonathan D. Adachi (Conceptualization, Data curation, Writing—review & editing), Claudie Berger (Conceptualization, Data curation, Writing—review & editing), Suzanne N. Morin (Conceptualization, Data curation, Writing—review & editing), David Goltzman (Conceptualization, Data curation, Writing—review & editing), and Alexander Bilbily (Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—review & editing)

Supplementary material

Supplementary material is available at *Journal of Bone and Mineral Research* online.

Funding

Amgen Canada Inc. and the INOVAIT FOCUS Program provided direct funding for this study. The Canadian Multicentre Osteoporosis Study (CaMos) has received support from the Canadian Institutes of Health Research (CIHR); Amgen Canada Inc.; Actavis Pharma Inc. (previously Warner Chilcott Canada Co.); Dairy Farmers of Canada; Eli Lilly Canada Inc.: Eli Lilly and Company; GE Lunar; Hologic Inc.; Merck Frosst Canada Ltd.; Novartis Pharmaceuticals Canada Inc.; P&G Pharmaceuticals Canada Inc.; Pfizer Canada Inc.; Roche (F. Hoffmann-La Roche Ltd.); Sanofi-Aventis Canada Inc. (previously Aventis Pharma Inc.); Servier Canada Inc.; and The Arthritis Society.

Conflicts of interest

M.D.C. reports financial support provided by Amgen Canada Inc. M.D.C. reports financial support provided by INOVAIT FOCUS Program. C.A.S., A.B., and M.D.C. report a relationship with 16 Bit Inc. that includes employment and equity or stocks. M.D.C. and A.B. have a patent for Systems and Methods for Approximating Bone Mineral Density and Fracture RIsk Using Single Energy X-Rays pending to 16 Bit. J.D.A. has received grants from Amgen and payment/honoraria and consulting fees from Alexion, Amgen, and Sandoz.

Data availability

Access to CaMos data can be requested at https://camos.org/data-access/.

Ethics approval statement

Participants of the CaMos study provided informed consent and the study was approved by the Institutional Review Board of each participating institution. For the current study analyzing CaMos data, research ethics board approval was obtained from an independent review board (Veritas, Montreal, Canada) with a waiver of informed consent.

References

1. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int*. 2008;19(4):385-397.
2. Kanis JA, Johnell O, Oden A, et al. Long-term risk of osteoporotic fracture in Malmö. *Osteoporos Int*. 2000;11(8):669-674.
3. GBD 2019 Fracture Collaborators. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the global burden of disease study 2019. *Lancet Healthy Longev*. 2021;2(9):e580-e592.

4. Kanis JA on behalf of the World Health Organization Scientific Group (2007). Assessment of osteoporosis at the primary health care level. In: *Technical Report*. World Health Organization Collaborating Centre for Metabolic Bone Diseases,University of Sheffield, UK. 2007: Printed by the University of Sheffield.
5. Kanis JA, Harvey NC, Johansson H, et al. A decade of FRAX: how has it changed the management of osteoporosis. *Aging Clin Exp Res*. 2020;32(2):187-196.
6. Hillier TA, Cauley JA, Rizzo JH, et al. WHO absolute fracture risk models (FRAX): do clinical risk factors improve fracture prediction in older women without osteoporosis. *J Bone Miner Res*. 2011;26(8):1774-1782.
7. Premaor M, Parker RA, Cummings S, et al. Predictive value of FRAX for fracture in obese older women. *J Bone Miner Res*. 2013;28(1):188-195.
8. Bilbily A, Syme CA, Adachi JD, et al. Opportunistic screening of low bone mineral density from standard X-rays. *J Am Coll Radiol*. 2024;21(4):633-639.
9. Kreiger N, T A, Joseph L, et al. Research notes: the Canadian multicentre osteoporosis study (CaMos): background, rationale, methods. *Canadian Journal on Aging / La Revue canadienne du vieillissement*. 1999;18(3):376-387.
10. Flack VF, Afifi AA, Lachenbruch PA, et al. Sample size determinations for the two rater kappa statistic. *Psychometrika*. 1988;53(3):321-325.
11. Leslie WD, Berger C, Langsetmo L, et al. Construction and validation of a simplified fracture risk assessment tool for Canadian women and men: results from the CaMos and Manitoba cohorts. *Osteoporos Int*. 2011;22(6):1873-1883.
12. Leslie WD, Morin SN, Lix LM, et al. Reassessment intervals for transition from low to high fracture risk among adults older than 50 years. *JAMA Netw Open*. 2020;3(1):e1918954.
13. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-268.
14. Fleiss JL, Levin B, Cho Paik M. *Statistical Methods for Rates and Proportions*. Wiley-Interscience; 2003.

Journal of Bone and Mineral Research, 2025, 00, 1–8
https://doi.org/10.1093/jbmr/zjaf113
Research Article