



Prediction of imminent fracture risk in Canadian women and men aged 45 years or older: external validation of the Fracture Risk Evaluation Model (FREM)

Sören Möller^{1,2} · Michael K. Skjødt^{2,3} · Lin Yan⁴ · Bo Abrahamsen^{2,3} · Lisa M. Lix⁴ · Eugene V. McCloskey^{5,6} · Helena Johansson^{5,7} · Nicholas C. Harvey^{8,9} · John A. Kanis^{5,7} · Katrine Hass Rubin^{1,2} · William D. Leslie⁴

Received: 11 May 2021 / Accepted: 17 September 2021 / Published online: 1 October 2021

© International Osteoporosis Foundation and National Osteoporosis Foundation 2021

Abstract

Summary The Fracture Risk Evaluation Model (FREM) identifies individuals at high imminent risk of major osteoporotic fractures. We validated FREM on 74,828 individuals from Manitoba, Canada, and found significant fracture risk stratification for all FREM scores. FREM performed better than age alone but not as well as FRAX® with BMD.

Introduction The FREM is a tool developed from Danish public health registers (hospital diagnoses) to identify individuals over age 45 years at high imminent risk of major osteoporotic fractures (MOF) and hip fracture (HF). In this study, our aim was to examine the ability of FREM to identify individuals at high imminent fracture risk in women and men from Manitoba, Canada.

Methods We used the population-based Manitoba Bone Mineral Density (BMD) Program registry, and identified women and men aged 45 years or older undergoing baseline BMD assessment with 2 years of follow-up data. From linked population-based data sources, we constructed FREM scores using up to 10 years of prior healthcare information.

Results The study population comprised 74,828 subjects, and during the 2 years of observation, 1612 incident MOF and 299 incident HF occurred. We found significant fracture risk stratification for all FREM scores, with AUC estimates of 0.63–0.66 for MOF for both sexes and 0.84 for women and 0.65–0.67 for men for HF. FREM performed better than age alone but not as well as FRAX® with BMD. The inclusion of physician claims data gave slightly better performance than hospitalization data alone. Overall calibration for 1-year MOF prediction was reasonable, but HF prediction was overestimated.

Conclusion In conclusion, the FREM algorithm shows significant fracture risk stratification when applied to an independent clinical population from Manitoba, Canada. Overall calibration for MOF prediction was good, but hip fracture risk was systematically overestimated indicating the need for recalibration.

Keywords Automated risk calculation · External validation · Osteoporotic fractures · Prediction models

Introduction

Osteoporotic fractures confer a substantial burden on individual patients and societies, including an increased mortality, a loss of quality-adjusted life years, and a high economic cost. In Europe, for example, an estimated 56.9 billion euro was spent on osteoporosis in 2019, of which 64% was for the treatment of incident fractures [1–3]. In recent years, it

has been demonstrated that some individuals experience a particularly high risk of osteoporotic fractures in the short (imminent) term [4–6]. Timely identification of such individuals would allow further evaluation and fracture risk management, potentially preventing the devastating fractures. While many tools have been developed to evaluate individual fracture risk [7], among others FRAX® [8], QFracture® [9], and Garvan [10, 11], there is an unmet need for validated tools — automated or manual — for the identification of individuals with an increased fracture risk in the short-term window, also termed “imminent risk.” One example of this is patients with a recent fragility fracture.

The Fracture Risk Evaluation Model (FREM) was developed for automated case-finding of individuals age

Katrine Hass Rubin and William D. Leslie share last authorship.

✉ Sören Möller
soren.moller@rsyd.dk

Extended author information available on the last page of the article

45 years or older, who are at high imminent (1-year) risk of major osteoporotic fractures (MOF) and hip fractures (HF) [12]. The tool was developed and subsequently validated using model selection on logistic regression models with routinely collected individual patient data from the Danish Health Registries as possible predictors and consists of odds ratios used for absolute risk prediction for a selected number of comorbidities. Internal validation demonstrated good discrimination (given as the area under the receiver operating curve, AUC) for MOF and excellent discrimination for HF. Using risk cutoffs of 2% for MOF and 0.3% for HF, positive predictive values were low, while negative predictive values were very high (generally > 99% for 1-year risk prediction) [12, 13].

To evaluate the generalizability and applicability of novel fracture risk assessment tools, they need to be tested in other populations and compared with widely deployed risk assessment tools. Hence, the aim of this study was to examine the ability of FREM to identify individuals at high imminent risk of fracture among women and men from Manitoba, Canada. Furthermore, we also set out to compare the performance of FREM with that of FRAX (including BMD) and age alone. The Manitoba DXA database and linked healthcare databases represent a unique opportunity to evaluate FREM in a different setting where appropriate clinical risk factor information and fracture outcomes have already been collected and verified.

Materials and methods

Study population

In the Canadian province of Manitoba (population 1.3 million in 2017), health services are provided to virtually all residents through a public healthcare system. DXA-based BMD testing has been managed as an integrated clinical program; criteria for testing have been published [14]. The program maintains a database of all DXA results which can be linked with other provincial population-based computerized health databases through an anonymous personal identifier. The DXA database has completeness and accuracy in excess of 99% [15].

The study population consisted of all women and men age 45 years or older with baseline DXA scans from January 1, 1996, to March 31, 2016. For those with more than one qualifying examination, only the first was included. We excluded those not registered for healthcare in Manitoba and those without 2 years of coverage after the first DXA (index date).

Estimation of fracture risk from FREM and FRAX

One-year predicted incidence of MOF and HF after index date was calculated with FREM using up to 10 years of prior hospital and physician claims data (implemented using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA)) [12]. Over 93% of the included individuals had at least 10 years of coverage and registry data before their baseline assessment. FREM consists of a weighted score including 38 risk factors for MOF for women and 43 risk factors for men, and 32 risk factors for hip fractures for both sexes. Recency of fracture is not reflected in FREM. These risk factors were determined by a data-driven algorithm based on Danish hospital diagnoses registry data with the aim of maximizing predictive ability without requiring plausible causal effects. Each condition considered by FREM was in this study assessed through a combination of hospital discharge abstracts (diagnoses and procedures coded using the International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] prior to 2004 and International Classification of Diseases, Tenth Revision, Canadian Enhancements [ICD-10-CA] thereafter) and physician billing claims (coded using ICD-9-CM in all years). Crosswalk tables were generated for ICD-9-CM and ICD-10 codes. Each condition was entered as present if there was a single hospitalization diagnosis or multiple physician claims diagnoses (2 or more within a 3-year time period). A single physician claims diagnosis was considered insufficient for registering a condition. Previous work has shown that this approach has good sensitivity and specificity for the diagnosis of several conditions associated with fractures [16–18]. In the evaluation, FREM scores were created based upon hospitalizations (in-patients only) alone (denoted FREM (H) in the following), physician claims alone (FREM (P)), or the combination (FREM (HP)). The coefficients in the Danish FREM score calculation were based on hospital in- and outpatient consults as previously described applying an ICD-9-CM/ICD-10 crosswalk to translate the ICD codes included in FREM [12].

Ten-year probability of a MOF and HF were calculated using the Canadian FRAX tool (FRAX® Desktop Multi-Patient Entry, version 3.8) as recently described [19]. Femoral neck BMD was included in the calculation. Hip DXA scans were performed and analyzed in accordance with manufacturer recommendations. The seven cross-calibrated instruments used for this study (1 DPX, 3 Prodigy and 3 iDXA, GE/Lunar Healthcare, Madison WI) exhibited equivalent phantom and in vivo calibration (T-score differences < 0.1), with stable long-term performance (coefficient of variation < 0.5%). The program's quality assurance is under strict supervision by a medical physicist

(Leslie and Metge [14]). The Canadian FRAX tool was calibrated using nationwide hip fracture and mortality data [20] and independently validated in the general population [21, 22]. In this cohort, 11,025 (14.7%) of the participants had a prior fragility fracture after age 40 included in the FRAX tool.

Moreover, we applied models including age as the only predictor of fracture risk, as it is known that age of the individual is a strong predictor in its own right [23]. These models were included to ascertain to which degree FREM, which includes age as a predictor, outperforms age alone.

Incident fracture ascertainment

Manitoba Health records for the study population were assessed for the presence of fracture diagnosis codes following BMD assessment up to March 31, 2018, using previously validated algorithms [24, 25]. Analyses were based upon hip, clinical vertebral, forearm, and humerus fracture diagnostic codes (collectively designated “major osteoporotic fractures” (MOF)) and hip fracture (HF). To minimize potential misclassification of prior incident fractures, we conservatively required that there be no hospitalization or physician visit(s) with the same fracture type in the 6 months preceding an incident fracture diagnosis. Fractures associated with codes indicating high trauma (external injury) were excluded, as FREM is not tailored to incorporate the uncertainty related to traumatic injuries.

Statistical analysis

Descriptive statistics for demographic and baseline characteristics are presented as mean \pm standard deviation (SD) for

continuous variables or count (proportion) for categorical variables and stratified by sex. Sex differences are tested by *t* test for numerical variables.

The proportions of individuals experiencing a MOF or a HF during 1 and 2 years after risk ascertainment are presented as counts and frequencies. Stratification in 1- and 2-year MOF and HF prediction was assessed from the area under the receiver operating curve (AUC) using each version of FREM, FRAX with BMD (as the upper bound reference standard), and age alone (as the lower bound base case). AUCs for FREM vs FRAX and FREM vs age alone were compared by the Hanley–McNeill method [26]. Calibration for 1-year MOF and HF prediction was assessed by comparing observed fracture risk with FREM (H), FREM (P), and FREM (HP).

Statistical analyses were performed with Statistica (version 13.0, StatSoft Inc., Tulsa, OK) and SPSS for Windows (version 24) for the ROC-AUC analyses.

Results

We included 74,828 individuals, corresponding to 67,930 (91%) women and 6898 (9%) men with a mean age of 65.3 years and a mean period of prior registry coverage of 33.1 years (Table 1).

The mean calculated FREM (HP) score, corresponding to the predicted absolute fracture risk, for 1-year MOF was 1.4% for women and significantly lower at 1.1% for men ($P < 0.001$). Furthermore, the mean FREM (HP) score for 1-year HF was 0.5% for women and 0.8% for men ($P < 0.001$). FREM scores determined only from hospital records (FREM (H)) and, respectively, only from physician

Table 1 Cohort characteristics at baseline

	Overall ($N=74,828$)	Women ($N=67,930$)	Men ($N=6898$)	<i>P</i> value for sex difference
	Mean \pm SD	Mean \pm SD	Mean \pm SD	
Age (years)	65.3 \pm 10.6	65.1 \pm 10.5	67.4 \pm 11.3	<0.001
Femoral neck T-score	-1.4 \pm 1	-1.4 \pm 1	-1.1 \pm 1.1	<0.001
1-year risk scores (% \pm SD)				
FREM MOF (HP)	1.4 \pm 1.3	1.4 \pm 1.3	1.1 \pm 1.6	<0.001
FREM MOF (H)	1.1 \pm 0.8	1.1 \pm 0.8	0.7 \pm 0.7	<0.001
FREM MOF (P)	1.3 \pm 1.1	1.3 \pm 1.1	0.9 \pm 1.1	<0.001
FRAX MOF percent (with BMD)	10.3 \pm 7.1	10.5 \pm 7.3	7.9 \pm 4.8	<0.001
FREM HIP (HP)	0.5 \pm 1	0.5 \pm 0.9	0.8 \pm 1.9	<0.001
FREM HIP (H)	0.4 \pm 0.6	0.4 \pm 0.6	0.4 \pm 0.7	0.392
FREM HIP (P)	0.5 \pm 0.8	0.4 \pm 0.8	0.6 \pm 1.3	<0.001
FRAX HIP percent (with BMD)	2.4 \pm 3.9	2.4 \pm 3.9	2.5 \pm 3.3	0.207

FREM (HP) uses hospitalization diagnoses and physician claims diagnoses; FREM (H) uses hospitalization diagnoses only; FREM (P) uses physician claims diagnoses only

claims (FREM (P)) were slightly lower than those determined using both sources (FREM (HP)) (Table 1).

In total, 902 (1.2%) individuals experienced a MOF during 1-year follow-up, and 1612 (2.2%) experienced a MOF during 2-year follow-up, while 136 (0.2%) experienced a HF during 1-year follow-up and 299 (0.4%) individuals during 2-year follow-up. A higher proportion of men than women experienced a MOF, and MOF as well as HF was more common with increasing age (Table 2).

Performance of FREM to stratify MOF risk gave an overall AUC of 0.652 (95% CI 0.633; 0.671) at 1 year and

0.659 (95% CI 0.645; 0.673) at 2 years for FREM (HP), with slightly lower for FREM (H) and FREM (P). The AUC for FREM (HP) was significantly higher than for age alone but significantly lower than for FRAX ($P < 0.001$ for both 1-year and 2-year risk). AUCs were generally higher for women than for men (Table 3 and Supplementary Figs. 1 and 2). The AUC for FREM (HP) was significantly higher than for age alone in both sexes and for both 1-year and 2-year risk ($P < 0.001$) and significantly lower than the AUC for FRAX in women ($P < 0.001$ for both 1-year and

Table 2 Observed fracture outcomes

	Number	MOF 1 year, <i>N</i> (%)	MOF 2 year, <i>N</i> (%)	Hip 1 year, <i>N</i> (%)	Hip 2 year, <i>N</i> (%)
Overall	74,828	902 (1.2%)	1,612 (2.2%)	136 (0.2%)	299 (0.4%)
Women	67,930	781 (1.1%)	1,445 (2.1%)	118 (0.2%)	271 (0.4%)
Age < 65 years	33,781	248 (0.7%)	447 (1.3%)	10 (0.0%)	33 (0.1%)
Age 65–79 years	27,928	359 (1.3%)	672 (2.4%)	52 (0.2%)	117 (0.4%)
Age 80+ years	6,221	174 (2.8%)	326 (5.2%)	56 (0.9%)	121 (1.9%)
Men	6,898	121 (1.8%)	167 (2.4%)	18 (0.3%)	28 (0.4%)
Age < 65 years	2,875	39 (1.4%)	58 (2.0%)	< 10 (0.1%)	< 10 (0.2%)
Age 65–79 years	2,995	56 (1.9%)	73 (2.4%)	< 10 (0.3%)	15 (0.5%)
Age 80+ years	1,028	26 (2.5%)	36 (3.5%)	< 10 (0.6%)	< 10 (0.6%)

Table 3 Area under the curve (AUC) for incident fracture stratification

	1-year prediction, AUC (95% CI)			2-year prediction, AUC (95% CI)		
	Overall	Women	Men	Overall	Women	Men
Major osteoporotic fractures						
FREM (HP)	0.652 (0.633; 0.671) *†	0.664 (0.644; 0.684) *†	0.640 (0.592; 0.688) *	0.659 (0.645; 0.673) *†	0.666 (0.651; 0.680) *†	0.644 (0.603; 0.685) *†
FREM (H)	0.631 (0.611; 0.650) †	0.656 (0.636; 0.677) *†	0.586 (0.533; 0.639) †	0.646 (0.631; 0.660) †	0.661 (0.646; 0.676) *†	0.596 (0.551; 0.641) †
FREM (P)	0.642 (0.623; 0.661) †	0.656 (0.636; 0.676) *†	0.632 (0.582; 0.681) *	0.652 (0.638; 0.666) *†	0.660 (0.645; 0.674) *†	0.633 (0.591; 0.675) *†
Age	0.631 (0.612; 0.650) †	0.639 (0.618; 0.659) †	0.565 (0.512; 0.618) †	0.638 (0.624; 0.652) †	0.647 (0.632; 0.662) †	0.562 (0.517; 0.607) †
FRAX	0.687 (0.669; 0.704) *	0.703 (0.684; 0.721) *	0.642 (0.595; 0.689) *	0.691 (0.678; 0.704) *	0.700 (0.686; 0.714) *	0.649 (0.610; 0.688) *
Hip fractures						
FREM (HP)	0.826 (0.795; 0.857) †	0.846 (0.816; 0.877) †	0.667 (0.547; 0.787) †	0.817 (0.794; 0.840) *†	0.832 (0.808; 0.856) *†	0.662 (0.568; 0.755) †
FREM (H)	0.823 (0.791; 0.856) †	0.845 (0.816; 0.874) †	0.657 (0.519; 0.795) †	0.806 (0.782; 0.831) *†	0.821 (0.796; 0.846) *†	0.647 (0.549; 0.746) †
FREM (P)	0.826 (0.794; 0.857) †	0.847 (0.816; 0.877) †	0.666 (0.545; 0.787) †	0.813 (0.789; 0.836) *†	0.829 (0.805; 0.853) *†	0.643 (0.546; 0.739) †
Age	0.815 (0.780; 0.850) †	0.835 (0.801; 0.869) †	0.664 (0.529; 0.799) †	0.790 (0.763; 0.816) †	0.807 (0.780; 0.833) †	0.613 (0.512; 0.715) †
FRAX	0.886 (0.866; 0.905) *	0.900 (0.880; 0.919) *	0.774 (0.698; 0.850) *	0.862 (0.844; 0.879) *	0.869 (0.851; 0.887) *	0.774 (0.700; 0.849) *

FREM (HP) uses hospitalization diagnoses and physician claims diagnoses; FREM (H) uses hospitalization diagnoses only; FREM (P) uses physician claims diagnoses only. * $P < 0.05$ versus age alone. † $P < 0.05$ versus FRAX

2-year risk) but not in men ($P=0.957$ for 1-year risk and $P=0.814$ for 2-year risk).

The overall AUC for HF prediction from FREM (HP) was 0.826 (95% CI 0.795; 0.857) at 1-year and 0.817 (95% CI 0.794; 0.840) at 2-year risk, which was again slightly higher than for FREM (H) and FREM (P). The AUC for FREM (HP) was significantly higher than for age (2-year risk) but significantly lower than for FRAX ($P<0.001$ for both 1-year and 2-year risk). AUCs were higher for women than for men (Table 3 and Supplementary Figs. 3 and 4). Again, AUC for

FRAX was higher than for FREM (HP) for both 1-year risk ($P<0.001$ for women and $P=0.034$ for men) and 2-year risk ($P<0.001$ for women and $P=0.005$ for men). On the other hand, FREM (HP) for HF did not significantly outperform age alone for 1-year risk ($P=0.143$ for women and $P=0.938$ for men), as significant difference was seen for 2-year risk in women ($P<0.001$) but not in men ($P=0.100$). Age-specific AUC results are presented in Table 4. Specifically, we note that the AUC values obtained in individuals aged 80+ are relatively low, indicating that FREM underperforms in the

Table 4 Validation of predictive models stratified by sex and age group

	1-year prediction, AUC (95% CI)		2-year prediction, AUC (95% CI)	
	Women	Men	Women	Men
Major osteoporotic fractures				
Age < 65 years				
FREM (HP)	0.608 (0.572; 0.645)	0.637 (0.541; 0.734)	0.600 (0.573; 0.626)	0.660 (0.584; 0.736)
FREM (H)	0.582 (0.544; 0.619)	0.510 (0.417; 0.603)	0.583 (0.556; 0.610)	0.557 (0.478; 0.637)
FREM (P)	0.596 (0.560; 0.633)	0.616 (0.515; 0.718)	0.592 (0.565; 0.618)	0.643 (0.565; 0.720)
Age	0.546 (0.511; 0.582)	0.481 (0.406; 0.556)	0.549 (0.523; 0.574)	0.517 (0.452; 0.582)
FRAX	0.657 (0.622; 0.692)	0.649 (0.560; 0.738)	0.646 (0.621; 0.672)	0.659 (0.590; 0.729)
Age 65–79 years				
FREM (HP)	0.608 (0.576; 0.640)	0.647 (0.577; 0.717)	0.607 (0.584; 0.630)	0.648 (0.589; 0.708)
FREM (H)	0.606 (0.575; 0.638)	0.575 (0.497; 0.653)	0.607 (0.584; 0.630)	0.585 (0.518; 0.652)
FREM (P)	0.593 (0.561; 0.625)	0.634 (0.560; 0.707)	0.596 (0.574; 0.619)	0.642 (0.579; 0.705)
Age	0.575 (0.545; 0.604)	0.525 (0.443; 0.607)	0.584 (0.563; 0.606)	0.542 (0.471; 0.612)
FRAX	0.682 (0.655; 0.710)	0.664 (0.594; 0.735)	0.673 (0.652; 0.693)	0.670 (0.611; 0.729)
Age 80+ years				
FREM (HP)	0.555 (0.510; 0.600)	0.505 (0.390; 0.620)	0.576 (0.543; 0.609)	0.526 (0.430; 0.623)
FREM (H)	0.535 (0.491; 0.579)	0.558 (0.456; 0.659)	0.557 (0.524; 0.590)	0.560 (0.466; 0.653)
FREM (P)	0.538 (0.493; 0.583)	0.482 (0.374; 0.591)	0.558 (0.525; 0.592)	0.495 (0.407; 0.582)
Age	0.561 (0.514; 0.609)	0.557 (0.446; 0.669)	0.570 (0.536; 0.604)	0.539 (0.445; 0.632)
FRAX	0.630 (0.587; 0.673)	0.470 (0.372; 0.568)	0.640 (0.609; 0.671)	0.508 (0.423; 0.593)
Hip fractures				
Age < 65 years				
FREM (HP)	0.797 (0.688; 0.906)	Too few observation	0.710 (0.620; 0.801)	Too few observation
FREM (H)	0.801 (0.686; 0.917)		0.684 (0.598; 0.771)	
FREM (P)	0.812 (0.710; 0.914)		0.710 (0.625; 0.794)	
Age	0.698 (0.575; 0.820)		0.596 (0.517; 0.675)	
FRAX	0.962 (0.938; 0.986)		0.853 (0.803; 0.904)	
Age 65–79 years				
FREM (HP)	0.736 (0.673; 0.800)	Too few observation	0.747 (0.704; 0.790)	0.617 (0.469; 0.766)
FREM (H)	0.719 (0.659; 0.780)		0.718 (0.669; 0.767)	0.602 (0.460; 0.744)
FREM (P)	0.730 (0.662; 0.798)		0.739 (0.696; 0.782)	0.592 (0.438; 0.747)
Age	0.708 (0.640; 0.777)		0.694 (0.647; 0.741)	0.527 (0.362; 0.692)
FRAX	0.834 (0.780; 0.887)		0.813 (0.775; 0.851)	0.746 (0.635; 0.856)
Age 80+ years				
FREM (HP)	0.575 (0.495; 0.654)	Too few observation	0.609 (0.558; 0.659)	Too few observation
FREM (H)	0.555 (0.475; 0.636)		0.575 (0.520; 0.631)	
FREM (P)	0.574 (0.497; 0.651)		0.606 (0.555; 0.658)	
Age	0.616 (0.535; 0.698)		0.642 (0.590; 0.694)	
FRAX	0.698 (0.634; 0.762)		0.700 (0.657; 0.743)	

oldest individuals in which prediction by age alone shows a similar performance.

Investigating the calibration of FREM, we observed generally higher predicted fracture risk obtained from FREM for MOF compared to the observed fracture rate, especially in the oldest age group (85+ years) where 2.80% of the women and 2.53% of the men experienced a fracture during the first year, but FREM (HP) predicted fracture rates of 3.88% in women and 2.99% in men. A similar trend of predicted risk above the observed risk was observed for HF, with predicted risks of 2–3% in the oldest compared to observed risk of 0.5–1% (Table 5 and Fig. 1).

Discussion

In this study, we have applied the FREM algorithm to a large registry-based cohort of individuals from Manitoba, Canada, with the aim of validating its performance in a population that is independent, and in some aspects quite distinct, from the Danish population in which FREM had been developed and validated [12, 13].

We observed that FREM (HP) performed well for HF in women with excellent AUC values for both 1- and 2-year risk prediction yet performed less well for HF in men with moderate AUC values. The predictive values for MOF were

Table 5 Calibration of the prediction models, comparing predicted and observed 1-year fracture risk

	Women			Men		
	Age < 65 years, mean (SD)	Age 65–79 years, mean (SD)	Age 80+ years, mean (SD)	Age < 65 years, mean (SD)	Age 65–79 years, mean (SD)	Age 80+ years, mean (SD)
Major osteoporotic fractures						
Observed	0.73%	1.29%	2.80%	1.36%	1.87%	2.53%
FREM (HP)	0.76% (0.48%)	1.61% (0.91%)	3.88% (2.17%)	0.52% (0.73%)	1.00% (1.22%)	2.99% (2.39%)
FREM (H)	0.64% (0.29%)	1.35% (0.50%)	3.03% (0.95%)	0.33% (0.28%)	0.61% (0.47%)	1.75% (1.01%)
FREM (P)	0.72% (0.40%)	1.54% (0.76%)	3.63% (1.75%)	0.44% (0.46%)	0.87% (0.78%)	2.59% (1.70%)
Hip fractures						
Observed	0.03%	0.19%	0.90%	0.10%	0.30%	0.58%
FREM (HP)	0.11% (0.14%)	0.49% (0.54%)	2.36% (1.78%)	0.17% (0.36%)	0.62% (1.22%)	3.00% (3.57%)
FREM (H)	0.08% (0.07%)	0.40% (0.34%)	1.89% (0.84%)	0.09% (0.13%)	0.31% (0.40%)	1.32% (1.31%)
FREM (P)	0.11% (0.12%)	0.47% (0.44%)	2.21% (1.42%)	0.13% (0.24%)	0.50% (0.72%)	2.36% (2.32%)

As the observed fractures are true observations, no SD are reported for these

FREM (HP) uses hospitalization diagnoses and physician claims diagnoses; FREM (H) uses hospitalization diagnoses only; FREM (P) uses physician claims diagnoses only

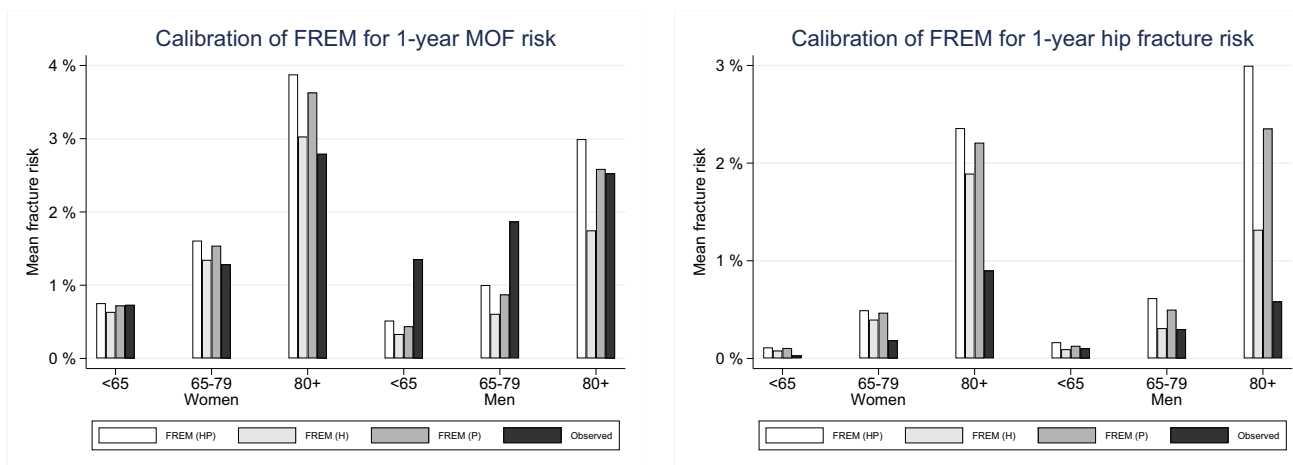


Fig. 1 Calibration of FREM for 1-year MOF risk (left) and 1-year hip fracture risk (right), stratified by sex and age group. Bars show predicted versus observed fracture risk

also somewhat lower than for HF in women, with moderate AUC values in both sexes. The performance was better if applying both hospital diagnosis codes and physician claims data in the model. FREM tended to overestimate absolute fracture risk in this population, as indicated by the higher predicted than observed fracture risk, especially for hip fractures.

In terms of discrimination, FREM generally outperformed age for MOF and for 2-year HF risk prediction in women. FREM was found to be less accurate than FRAX including BMD (except in the prediction of MOF in men, where discrimination was comparable between FRAX and FREM). This was as expected, due to the inclusion of BMD in FRAX but not in FREM. It is worth noting in this context that the intention of FREM is to identify individuals on a population level who are at high imminent risk of fractures and should undergo further risk evaluation, rather than to provide a final risk estimate on which to base treatment decisions. Our results could indicate that FREM may be a useful tool for such first-line screening.

Comparing the validity of FREM in Canada with that observed in Denmark [12, 13], we determined similar levels of predictive power, indicating good generalizability of the FREM algorithm. On the other hand, we in the current study observed larger differences in the calibration of the algorithm observing much lower rates of osteoporotic fractures in the cohort than predicted by the algorithm. This indicates that predictions of absolute risk, contrary to merely identifying high-risk individuals, might require a population- and setting-specific calibration of the FREM algorithm. One possible contributing factor in the calibration difference between this study and the original FREM development is the inclusion of only DXA-scanned individuals in the current cohort, compared to the full population in the Danish cohort, although additional differences caused by differences in population risk between Manitoba and Denmark cannot be excluded. In addition, it is conceivable that the threshold for hospital contact — even on an outpatient basis — marks a somewhat more severe case than a patient seen under the same diagnosis in General Practice. This, combined with the known high osteoporotic fracture risk in the Danish population [27], likely explains the requirement for a somewhat lower calibration of the risk prediction for use in Canada.

In recent years, several studies regarding tools for imminent fracture risk assessment have been published. The studies differ in their eligibility criteria and settings, potentially resulting in rather heterogeneous fracture risk profiles, and base their model fully or partially on risk factors with a biologically plausible link to fractures [28–30]. A conceptually different tool was developed by Almog et al. [31], who — using a US administrative health database — applied deep learning techniques from natural language processing to patient medical history to develop an algorithm to identify

individuals at risk of fracture within a 2-year period [31]. In comparison, FREM included the entire Danish population aged 45 years or older and is based on covariates identified in the analytical process to have a predictive value for fractures, irrespective of whether these fit into the pathophysiological process of fracture development [12]. This difference in covariates applied in the FREM algorithm may cause a lower face validity of FREM — as well as the algorithm developed by Almog et al. [31] — as compared to tools based on established clinical risk factors for osteoporotic fractures.

The AUC among these published tools range from 0.71 to 0.77 for HF and 0.62 to 0.73 for composite fracture outcomes, derived from development or internal validation cohorts [28–30]. Almog et al. [31] demonstrated AUC of 0.79–0.82 for any fractures — depending on the model framework applied — and higher if focusing on subsequent fractures only [31]. While direct comparison of FREM against these tools is not appropriate, among others as this is an external validation of FREM as compared to the internal validation of the abovementioned tools [32], the results do seem to indicate that FREM tends to perform better or on par with regard to discriminative power for the prediction of hip fractures in women. A lower AUC for composite fracture outcomes is similarly noted for FREM, as well as FRAX, in this study as well as in prior FREM studies [12, 13], and has been suggested to be due to an increased variability in predictors across the different fracture types included in such outcomes [29].

Beyond this external validation of FREM, we are only aware of one other imminent fracture risk assessment tool having been validated externally, though these data have yet to be published in full. The available data indicate that the IFRISK tool achieves an AUC for HF in the internal validation of 0.78, while in two external validations, it was 0.70–0.71 [33].

Uncertainties remain about the appropriate time horizon for fracture risk assessments, and it has been argued that a 10-year time frame is most relevant [34]. Nevertheless, the appealing perspective of imminent fracture risk assessment tools is the potential of improved prevention of osteoporotic fractures. Hence, evidence-based pharmacological and non-pharmacological means to manage an increased fracture risk are available [35], and there is an increasing call for action in this setting [36–38]. Despite this, fragility fracture prevention remains insufficient [38–40], indicating a need for other ways of identifying individuals at risk. The FREM tool holds the potential for automated screening, as it only includes diagnoses, which are often available in electronic databases and patient files, and we believe that the incorporation of this or a similar tool may prompt a more proactive approach to fracture risk management among physicians. However, this has yet to be evaluated in clinical trials. As with all

risk assessment tools, perhaps in particular those leveraging administrative health data, there is a risk that they be used by healthcare insurers to screen (potential) subscribers, and this risk should be mitigated through legislative efforts.

The main strength of this study is the validation on a large cohort distinct from the original setting FREM was developed in. Moreover, due to the Manitoba database including both diagnoses from hospitals as well as from general practitioners, we could investigate the dependence of FREM on data source, hereby increasing the generalizability of the algorithm to settings with different regimes of administrative data collection. Furthermore, the availability of BMD measurements for the included participants enabled for the first time a direct comparison between FREM and the much-used FRAX algorithm, enabling us to quantify the degree to which administrative information can approximate the information content of clinical measurements.

Acknowledged limitations to this study include the reliance on a clinical BMD registry, though this captures all clinical BMD testing in the province of Manitoba and therefore reflects a broad range of individuals typically referred for fracture risk assessment. Women referred for testing have mean BMD measurements that are average for age with prevalence of FRAX risk factors that are also similar to the general Canadian population; as expected, men typically show greater referral bias [41, 42]. Another limitation is the use of a 10-year lookback in this study, whereas FREM was originally developed using a 15-year lookback [12]. However, as FREM was subsequently validated in a Danish cohort using a 5-year lookback [13], we do not expect this to be a source of error in the fracture risk prediction in this study.

Moreover, differences between the Canadian sample included in this study, and the Danish sample used to develop FREM, are both a weakness and a strength of the current study. There are differences with respect to ethnic composition, Manitoba being much more ethnically heterogeneous than Denmark with a large proportion of immigrants, which might contribute to the need for recalibration. On the other hand, significant stratification with FREM, in spite of such differences, indicates that FREM might show acceptable performance in a wide range of populations after recalibration.

Future research on FREM should explore the additional predictive value obtained by including registry data concerning the use of medications and potentially biochemistry results obtained as part of daily clinical practice. Furthermore, as has been demonstrated in this study, there is a need to calibrate the tool for use under local conditions, as has also been done for FRAX. Moreover, taking into account the recency of previous fractures as well as of other risk factors might have the possibility to further improve FREM, similar to that which has been observed for FRAX [43]. Hence, a

future development of an extended version of FREM taking recency as well as frequency of predictors into account is desirable. Additionally, studies investigating the cost-effectiveness of implementing FREM in screening programs would be desirable.

In conclusion, this study documents that FREM has good validity for predicting the risk of osteoporotic — and particularly hip — fractures also in populations different from the Danish population in which the algorithm was developed but with the need of population-specific calibration if the aim is absolute risk prediction instead of just risk stratification. FREM performs better than age alone but weaker than FRAX including BMD measurements, and inclusion of physician claims enhanced the performance of FREM.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00198-021-06165-1>.

Acknowledgements The authors acknowledge the Manitoba Centre for Health Policy for use of data contained in the Population Health Research Data Repository (HIPC 2016/2017–29). The results and conclusions are those of the authors and no official endorsement by the Manitoba Centre for Health Policy, Manitoba Health, Seniors and Active Living, or other data providers is intended or should be inferred. This article has been reviewed and approved by the members of the Manitoba Bone Density Program Committee.

Funding LML is supported by a Tier I Canada Research Chair.

Data availability Data sharing is not permitted under the Researcher Agreement with Manitoba Health, Seniors and Active Living (MHSAL). However, researchers may apply for data access through the Health Research Ethics Board of the University of Manitoba and the Health Information and Privacy Committee of MHSAL.

Declarations

Ethics approval The study was approved by the Health Research Ethics Board of the University of Manitoba.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflicts of interest Michael K. Skjødt: educational grant, UCB; institutional research grant, UCB/Amgen. Eugene McCloskey: nothing to declare for the context of this paper, but numerous ad hoc consultancies/speaking honoraria and/or research funding from Amgen, Bayer, General Electric, GSK, Fresenius Kabi, Hologic, Lilly, Merck Research Labs, Novartis, Novo Nordisk, Nycomed, Ono, Pfizer, ProS-trakan, Roche, Sanofi-Aventis, Servier, Tethys, UCB, and Warner-Chilcott. Nicholas Harvey: nothing to declare for the context of this paper, but has received consultancy/lecture fees/honoraria/grant funding from the Alliance for Better Bone Health, Amgen, MSD, Eli Lilly, Servier, Shire, UCB, Consilient Healthcare, Radius Health, Kyowa Kirin, and Internis Pharma. Bo Abrahamsen: institutional research contracts with Novartis, UCB, Kyowa-Kirin. Fees or honoraria from UCB, Amgen, Pharmacosmos. John Kanis: Professor Kanis led the team that developed FRAX as director of the WHO Collaborating Centre for Metabolic Bone Diseases; he has no financial interest in FRAX. Wil-

liam Leslie, Lisa Lix, Lin Yan, Helena Johansson, Sören Möller, and Katrine Hass Rubin: No conflicts of interest.



References

- Kanis JA, Cooper C, Rizzoli R et al (2017) Identification and management of patients at increased risk of osteoporotic fracture: outcomes of an ESCEO expert consensus meeting. *Osteoporos Int* 28:2023–2034
- Kanis J, Norton N, Harvey N, Jacobson T, Johansson H, Lorentzon M, McCloskey E, Willers C, Borgström F (2021) SCOPE 2020: a new scorecard for osteoporosis in Europe. *Arch Osteoporos*
- Kanis JA, Norton N, Harvey NC, Jacobson T, Johansson H, Lorentzon M, McCloskey EV, Willers C, Borgström F (2021) SCOPE 2021: a new scorecard for osteoporosis in Europe. *Arch Osteoporos* 16:82
- Johansson H, Siggeirsdóttir K, Harvey NC, Odén A, Gudnason V, McCloskey E, Sigurdsson G, Kanis JA (2017) Imminent risk of fracture after fracture. *Osteoporos Int* 28:775–780
- Roux C, Briot K (2017) Imminent fracture risk. *Osteoporos Int* 28:1765–1769
- Adachi JD, Berger C, Barron R et al (2019) Predictors of imminent non-vertebral fracture in elderly women with osteoporosis, low bone mass, or a history of fracture, based on data from the population-based Canadian Multicentre Osteoporosis Study (CaMos). *Arch Osteoporos* 14:53
- Rubin KH, Friis-Holmberg T, Hermann AP, Abrahamsen B, Brixen K (2013) Risk assessment tools to identify women with increased risk of osteoporotic fracture: complexity or simplicity? A systematic review. *J Bone Miner Res* 28:1701–1717
- Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E (2008) FRAX™ and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 19:385–397
- Hippisley-Cox J, Coupland C (2009) Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *BMJ* 339:b4229
- Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV (2007) Development of a nomogram for individualizing hip fracture risk in men and women. *Osteoporos Int* 18:1109–1117
- Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV (2008) Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. *Osteoporos Int* 19:1431–1444
- Rubin KH, Möller S, Holmberg T, Bliddal M, Søndergaard J, Abrahamsen B (2018) A New Fracture Risk Assessment Tool (FREM) based on public health registries. *J Bone Miner Res* 33:1967–1979
- Skjold MK, Moller S, Hyldig N, Clausen A, Bliddal M, Søndergaard J, Abrahamsen B, Rubin KH (2021) Validation of the Fracture Risk Evaluation Model (FREM) in predicting major osteoporotic fractures and hip fractures using administrative health data. *Bone* 147:115934
- Leslie WD, Metge C (2003) Establishing a regional bone density program: lessons from the Manitoba experience. *J Clin Densitom* 6:275–282
- Leslie WD, Caetano PA, Macwilliam LR, Finlayson GS (2005) Construction and validation of a population-based bone densitometry database. *J Clin Densitom* 8:25–30
- Lix L, Yogendran M, Shaw S, Burchill C, Metge C, Bond R (2008) Population-based data sources for chronic disease surveillance. *Chronic Dis Can* 29:31–38
- Duerksen DR, Lix LM, Leslie WD (2019) Development of an administrative definition for celiac disease. *BMC Res Notes* 12:661
- Blanchard JF, Ludwig S, Wajda A, Dean H, Anderson K, Kendall O, Depew N (1996) Incidence and prevalence of diabetes in Manitoba, 1986–1991. *Diabetes Care* 19:807–811
- Leslie WD, Schousboe JT, Morin SN, Martineau P, Lix LM, Johansson H, McCloskey EV, Harvey NC, Kanis JA (2020) Measured height loss predicts incident clinical fractures independently from FRAX: a registry-based cohort study. *Osteoporos Int* 31:1079–1087
- Leslie WD, Lix LM, Langsetmo L et al (2011) Construction of a FRAX(R) model for the assessment of fracture probability in Canada and implications for treatment. *Osteoporos Int* 22:817–827
- Fraser LA, Langsetmo L, Berger C et al (2011) Fracture prediction and calibration of a Canadian FRAX® tool: a population-based report from CaMos. *Osteoporos Int* 22:829–837
- Leslie WD, Lix LM, Johansson H, Oden A, McCloskey E, Kanis JA (2010) Independent clinical validation of a Canadian FRAX tool: fracture prediction and model calibration. *J Bone Miner Res* 25:2350–2358
- Rubin KH, Abrahamsen B, Friis-Holmberg T, Hjelmberg JV, Bech M, Hermann AP, Barkmann R, Glüer CC, Brixen K (2013) Comparison of different screening tools (FRAX®, OST, ORAI, OSIRIS, SCORE and age alone) to identify women with increased risk of fracture A population-based prospective study. *Bone* 56:16–22
- Lix LM, Azimaee M, Osman BA, Caetano P, Morin S, Metge C, Goltzman D, Kreiger N, Prior J, Leslie WD (2012) Osteoporosis-related fracture case definitions for population-based administrative data. *BMC Public Health* 12:301
- Epp R, Alhrbi M, Ward L, Leslie WD (2018) Radiological validation of fracture definitions from administrative data. *J Bone Miner Res* 33:S275
- Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843
- Svedbom A, Hernlund E, Ivergård M, Compston J, Cooper C, Stenmark J, McCloskey EV, Jönsson B, Kanis JA (2013) Osteoporosis in the European Union: a compendium of country-specific reports. *Arch Osteoporos* 8:137
- Sheer RL, Barron RL, Sudharshan L, Pasquale MK (2020) Validated prediction of imminent risk of fracture for older adults. *Am J Manag Care* 26:e91–e97
- Weycker D, Edelsberg J, Barron R, Atwood M, Oster G, Crittenden DB, Grauer A (2017) Predictors of near-term fracture in osteoporotic women aged ≥65 years, based on data from the study of osteoporotic fractures. *Osteoporos Int* 28:2565–2571
- Reber KC, König HH, Becker C, Rapp K, Büchele G, Mächler S, Lindlbauer I (2018) Development of a risk assessment tool for osteoporotic fracture prevention: a claims data approach. *Bone* 110:170–176
- Almog YA, Rai A, Zhang P et al (2020) Deep learning with electronic health records for short-term fracture risk identification: crystal bone algorithm development and validation. *J Med Internet Res* 22:e22550
- Kanis JA, Oden A, Johansson H, McCloskey E (2012) Pitfalls in the external validation of FRAX. *Osteoporos Int* 23:423–431
- Khalid S, Pineda-Moncusi M, El-Hussein L, et al. (2021) Predicting imminent fractures in patients with a recent fracture or starting oral bisphosphonate therapy: development and international validation of prognostic models. *J Bone Miner Res*
- Kanis JA, Johansson H, Harvey NC et al (2021) The use of 2-, 5-, and 10-year probabilities to characterize fracture risk after a recent sentinel fracture. *Osteoporos Int* 32:47–54
- Black DM, Rosen CJ (2016) Clinical practice Postmenopausal osteoporosis. *N Engl J Med* 374:254–262
- The Lancet Diabetes E (2018) Osteoporosis: a roadmap to close the treatment gap. *Lancet Diabetes Endocrinol* 6:833

37. Kanis JA, Svedbom A, Harvey N, McCloskey EV (2014) The osteoporosis treatment gap. *J Bone Miner Res* 29:1926–1928
38. Harvey NC, McCloskey EV, Mitchell PJ, Dawson-Hughes B, Pierroz DD, Reginster JY, Rizzoli R, Cooper C, Kanis JA (2017) Mind the (treatment) gap: a global perspective on current and future strategies for prevention of fragility fractures. *Osteoporos Int* 28:1507–1529
39. Skjødtk MK, Khalid S, Ernst M et al (2020) Secular trends in the initiation of therapy in secondary fracture prevention in Europe: a multi-national cohort study including data from Denmark, Catalonia, and the United Kingdom. *Osteoporos Int* 31:1535–1544
40. Abrahamsen B, Skjødtk MK, Vestergaard P (2019) Hip fracture rates and time trends in use of anti-osteoporosis medications in Denmark for the period 2005 to 2015: missed opportunities in fracture prevention. *Bone* 120:476–481
41. Leslie WD, Tsang JF, Caetano PA, Lix LM (2007) Effectiveness of bone density measurement for predicting osteoporotic fractures in clinical practice. *J Clin Endocrinol Metab* 92:77–81
42. Leslie WD, Berger C, Langsetmo L et al (2011) Construction and validation of a simplified fracture risk assessment tool for Canadian women and men: results from the CaMos and Manitoba cohorts. *Osteoporos Int* 22:1873–1883
43. Kanis JA, Johansson H, Harvey NC, Gudnason V, Sigurdsson G, Siggeirsdottir K, Lorentzon M, Liu E, Vandenput L, McCloskey EV (2020) Adjusting conventional FRAX estimates of fracture probability according to the recency of sentinel fractures. *Osteoporos Int* 31:1817–1828

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sören Möller^{1,2}  · Michael K. Skjødtk^{2,3}  · Lin Yan⁴ · Bo Abrahamsen^{2,3}  · Lisa M. Lix⁴  · Eugene V. McCloskey^{5,6}  · Helena Johansson^{5,7}  · Nicholas C. Harvey^{8,9}  · John A. Kanis^{5,7}  · Katrine Hass Rubin^{1,2}  · William D. Leslie⁴ 

Michael K. Skjødtk
miksk@regionsjaelland.dk

Lin Yan
Lin.Yan@umanitoba.ca

Bo Abrahamsen
b.abrahamsen@physician.dk

Lisa M. Lix
Lisa.Lix@umanitoba.ca

Eugene V. McCloskey
e.v.mccloskey@sheffield.ac.uk

Helena Johansson
helena@statiq.se

Nicholas C. Harvey
nch@mrc.soton.ac.uk

John A. Kanis
jakanis@outlook.com

Katrine Hass Rubin
Katrine.Rubin@rsyd.dk

William D. Leslie
bleslie@sbgh.mb.ca

¹ Open Patient data Explorative Network, Odense University Hospital, Odense, Denmark

² Research unit OPEN, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

³ Department of Medicine, Holbæk Hospital, Holbæk, Denmark

⁴ University of Manitoba, Winnipeg, Canada

⁵ Centre for Metabolic Bone Diseases, University of Sheffield Medical School, Sheffield, UK

⁶ Centre for Integrated Research in Musculoskeletal Ageing (CIMA), Mellanby Centre for Bone Research, University of Sheffield, Sheffield, UK

⁷ Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Australia

⁸ MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

⁹ NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK