IJSCIA

International Journal of Scientific Advances
ISSN: 2708-7972
Volume: 7 | Issue: 6 | Nov – Dec 2026 Available Online: www.ijscia.com
DOI: 10.51542/ijscia.v7i6.40

Calibration and Discrimination of FRAX and Diabetes-Specific Adjustments for Major Osteoporotic and Hip Fracture Prediction in Adults with Type 2 Diabetes: A Systematic Review

Muhammad Farrell Rikhad[1*], Rendy Wijaya[1], Zarifa Nurinnisa[1], Ritma Ratri Ayunda Putri[1], Adinda Nabila Azzah[2]

[1]Faculty of Medicine, Brawijaya University, Malang, Indonesia

[2]Master's Program in Biomedical Sciences,
Faculty of Medicine, Brawijaya University, Malang, Indonesia

*Corresponding author: Muhammad Farrell Rikhad; rikhadfm@gmail.com
ORCID ID: https://orcid.org/0000-0003-2012-0758

ABSTRACT

Background: Adults with type 2 diabetes mellitus (T2DM) have increased fragility-fracture risk despite often normal or elevated bone mineral density (BMD). The Fracture Risk Assessment Tool (FRAX) omits T2DM as a direct predictor and may underestimate absolute fracture probability. We evaluated the calibration and discrimination of unadjusted and diabetes-adjusted FRAX for major osteoporotic fracture (MOF) and hip fracture in adults with T2DM. ***Subjects and Method:*** PubMed, Europe PMC, ScienceDirect, OpenAlex, and the Cochrane Library were searched for longitudinal studies from 2016 through September 2026. Eligible studies compared country-specific FRAX predictions with incident fractures over aligned horizons and reported calibration, discrimination, or reclassification. We used CHARMS for extraction, PROBAST for risk-of-bias assessment, and narrative synthesis. ***Results:*** Of 968 database records identified, three reports met the eligibility criteria. The diabetes groups totaled 14,706 participants, with overlap between the two Manitoba reports. In one Manitoba analysis (n=4,136), unadjusted FRAX underestimated MOF (observed-to-predicted [O:P] ratio 1.15, 95% confidence interval [CI] 1.03 to 1.28) and hip fracture (O:P ratio 1.85, 95% CI 1.51 to 2.20). Rheumatoid arthritis proxy coding, trabecular bone score adjustment, lowering the femoral-neck T-score by 0.5 standard deviations, or adding 10 years to input age yielded MOF O:P ratios of 0.91 to 1.02. Hip-fracture ratios remained 1.22 to 1.63. The other Manitoba report found greater MOF underprediction with longer diabetes duration. In Shanghai (n=1,730, 109 MOFs), FRAX without BMD underestimated MOF (O:P ratio 2.62, 95% CI 2.17 to 3.08), with an area under the curve of 0.75 (95% CI 0.70 to 0.79). ***Conclusion:*** Unadjusted FRAX underestimated fracture probability in the studied T2DM cohorts. The adjustments improved MOF calibration in one registry cohort, but residual hip-fracture underprediction and limited external validation preclude recommending a universal adjustment.

keywords: FRAX; T2DM; osteoporotic fracture; hip fracture; prediction model

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is associated with increased fragility-fracture risk, even though areal bone mineral density (BMD) is often normal or higher than in people without diabetes. This apparent mismatch complicates fracture prevention: a patient may have clinically important skeletal fragility without meeting a densitometric definition of osteoporosis. In older adults, fracture risk at a given femoral-neck BMD T-score is higher in those with T2DM than in those without diabetes [1–3]. Meta-analyses of observational studies consistently report an increased risk of hip fracture in T2DM, with pooled relative risks of roughly 1.30 to 1.70, and higher risk with longer diabetes duration and insulin use [4–7].

The excess risk is not confined to the hip: incident vertebral fractures are also more frequent in T2DM, even though prevalent vertebral fractures appear less common [8].

Several mechanisms may contribute. Prolonged hyperglycemia and the accumulation of advanced glycation end products can impair the material properties of bone collagen, while changes in bone turnover and microarchitecture may weaken bone without a proportionate decline in BMD. Diabetes-related neuropathy, visual impairment, reduced physical function, and hypoglycemia can also increase the likelihood of falling. These pathways differ among patients and may change with diabetes duration and complications [2,3,9–11].

Because these mechanisms act largely on bone quality rather than bone quantity, DXA-derived BMD may underestimate fracture risk in T2DM, although this can be improved by applying specific adjustments [11].

The Fracture Risk Assessment Tool (FRAX) estimates 10-year probabilities of major osteoporotic fracture (MOF: hip, spine, forearm, or proximal humerus) and hip fracture. It combines clinical risk factors with optional femoral-neck BMD and uses country-specific fracture and mortality data. Since its launch in 2008, FRAX models have been made available for 64 countries [12], and fracture risk algorithms of this kind are now widely used in clinical practice and guidance to target high-risk individuals for treatment [13,14]. Although FRAX is widely used to guide decisions about further assessment and treatment, conventional FRAX has no direct input for T2DM. A calculated probability may therefore fail to capture the full fracture risk associated with diabetes [15,16].

Evaluating that concern requires attention to two distinct properties of a prediction model. Discrimination describes whether the model assigns higher predicted risk to people who subsequently fracture than to those who do not; it is commonly summarized by a C-statistic or area under the curve (AUC). Calibration describes agreement between predicted probability and observed fracture occurrence. An observed-to-predicted ratio above one indicates underprediction on average. A model may distinguish patients reasonably well yet underestimate their absolute risk, which matters when an intervention threshold is based on a FRAX probability [16,17]. Methodological guidance therefore recommends that discrimination and calibration are both reported for any prediction model, because poorly calibrated models can be misleading and potentially harmful for clinical decision-making [18,19].

Evidence from prospective cohorts shows that adults with T2DM experience more fractures than adults without diabetes at comparable FRAX probabilities or femoral-neck T-scores. Registry studies have likewise reported underprediction of MOF or hip fracture in people with diabetes, with greater discrepancies in some groups with longer diabetes duration. The size of that discrepancy may differ between fracture outcomes and between populations; a single correction cannot be assumed to work equally well for both MOF and hip fracture [1,16,20].

Four practical approaches have been proposed to adjust FRAX in T2DM: selecting rheumatoid arthritis as a proxy risk factor, applying a trabecular bone score (TBS) adjustment, lowering the femoral-neck T-score entered into FRAX by 0.5 standard deviations, or increasing the age entered into FRAX by 10 years. The TBS adjustment is biologically motivated: in a Manitoba registry cohort of women aged 50 years and older, lumbar spine TBS was lower in those with diabetes despite higher BMD and captured more of the diabetes-associated fracture risk than BMD did [21]. These approaches alter predicted probabilities in different ways. Their performance needs to be judged against observed fractures, rather than by the increase in predicted probability alone. Diabetes duration and complications may provide additional prognostic information, but neither is incorporated as a standard direct input in conventional FRAX [17,20,22,23].

The available evaluations differ in setting, country-specific FRAX model, use of BMD, participant characteristics, fracture outcome, and reported performance measures. Some reports also draw on overlapping registry populations. These differences can obscure whether an adjustment improves average calibration, preserves discrimination, or generalizes beyond the cohort in which it was studied. A synthesis organized by model variant and fracture outcome can make the clinical evidence and its limits clearer [20,22,23].

We systematically reviewed longitudinal studies evaluating the calibration and discrimination of unadjusted FRAX and explicit diabetes-related FRAX adjustments for incident MOF and hip fracture in adults with T2DM. We also examined reported reclassification, risk of bias, applicability, and the availability of independent, comparable evidence for quantitative synthesis.

SUBJECTS AND METHOD

1. Study Design

This study is designed as a systematic review that aims to summarize and critically evaluate the scientific evidence available regarding longitudinal evaluations of fracture prediction using FRAX in adults with type 2 diabetes mellitus (T2DM) based on PRISMA 2020 guidelines for review reporting, TRIPOD-SRMA informed prediction-model reporting, and CHARMS informed data extraction [24–27]. This systematic review protocol was registered in PROSPERO (CRD420261532316).

2. Study Population

Eligibility was specified using population, index model, comparator, outcome, timing, setting, and study design (PICOTS). We included adults aged 18 years or older with T2DM. Mixed-diabetes cohorts were eligible only if results for T2DM could be extracted separately. Eligible settings included community, registry, primary-care, specialty-care, hospital, and population-based healthcare cohorts.

Eligible designs were prospective or retrospective longitudinal cohorts and registry cohorts, including case-cohort or nested case-control analyses only when the underlying longitudinal risk set and aligned performance estimate were identifiable. We considered full reports published from 1 January 2016 to the search date. The selection flow records one language-based removal during database filtering. We excluded cross-sectional and conventional case-control studies, reviews, editorials, case reports, conference abstracts without adequate original performance data, simulations without empirical

validation, standalone model-development studies without eligible FRAX validation, and studies of non-FRAX tools lacking separable FRAX performance.

3. Study Variables

The index model was a country-specific Fracture Risk Assessment Tool (FRAX) calculation, with or without femoral-neck bone mineral density (BMD). We considered unadjusted FRAX and explicit diabetes-related modifications: entering rheumatoid arthritis as a proxy risk factor; using trabecular bone score (TBS)-adjusted FRAX; adding 10 years to entered age; lowering femoral-neck T-score by 0.5 standard deviations; combinations of these approaches; or formally recalibrating or extending FRAX using diabetes-related variables. For adjustment studies, unadjusted FRAX in the same cohort was the preferred comparator. A separate comparator was not required to validate unadjusted FRAX.

Eligible fracture outcomes were incident major osteoporotic fracture (MOF: hip, clinical vertebral, forearm, or proximal humerus fracture) and hip fracture, assessed separately. A study had to report an interpretable measure of calibration, discrimination, reclassification, threshold-specific classification, overall prediction accuracy, or clinical utility. We required the prediction horizon and observed follow-up to align or to be reconciled using appropriate time-to-event methods; 10 years was preferred, and other aligned horizons were interpreted separately.

4. Operational Definition of Variables

Calibration and discrimination were the co-primary model-performance domains, evaluated separately for MOF and hip fracture. We preferred an observed-to-predicted fracture ratio (O:P) for calibration, followed by calibration-in-the-large or intercept, slope, and plots were reported. For ratios expressed in the reverse direction (predicted-to-observed), we inverted them only when sufficient information was available. O:P = 1 denotes agreement in average risk; values greater than 1 indicate underprediction and values below 1 indicate overprediction.

Discrimination measures included a time-to-event C-statistic, Harrell C-index, or area under the curve (AUC) at the aligned horizon. Secondary measures included net reclassification improvement, integrated discrimination improvement, threshold-specific classification, Brier score, and decision-curve net benefit. Estimates were interpreted in light of confidence intervals, outcome ascertainment, model formulation, and follow-up; no universal discrimination threshold was imposed.

5. Search Strategy and Study Selection

This systematic review used PubMed, Europe PMC, ScienceDirect, OpenAlex, and the Cochrane Library for reports dated from 1 January 2016 until 14 September 2026. Search concepts combined T2DM and FRAX terms with prediction, validation, calibration, discrimination, reclassification, and diabetes-adjustment terms. Where supported, adjustment terms covered TBS, a rheumatoid arthritis proxy, entered age, femoral-neck T-score, and diabetes duration. Search terms used in these databases are shown in Table 1 below.

After the deduplication process, titles and abstracts were screened as included, excluded, or uncertain against the eligibility criteria. Reports lacking enough information for exclusion were retained for full-text retrieval. We applied the PICOTS criteria for determining eligible studies.

TABLE 1: Search terms used in database searching.

Database	Search Terms Used
PubMed	("Diabetes Mellitus, Type 2"[Mesh] OR "type 2 diabetes"[tiab] OR "type 2 diabetes mellitus"[tiab] OR "type II diabetes"[tiab] OR "type II diabetes mellitus"[tiab] OR T2DM[tiab] OR T2D[tiab] OR NIDDM[tiab] OR "non-insulin-dependent diabetes"[tiab] OR "noninsulin-dependent diabetes"[tiab]) AND `(FRAX[tiab] OR FRAXplus[tiab] OR "FRAX plus"[tiab] OR "fracture risk assessment tool"[tiab] OR "fracture risk assessment score"[tiab] OR "fracture risk assessment algorithm"[tiab])`
Europe PMC	(KW:"Diabetes Mellitus, Type 2" OR TITLE_ABS:"type 2 diabetes" OR TITLE_ABS:"type 2 diabetes mellitus" OR TITLE_ABS:"type II diabetes" OR TITLE_ABS:"type II diabetes mellitus" OR TITLE_ABS:T2DM OR TITLE_ABS:T2D OR TITLE_ABS:NIDDM OR TITLE_ABS:"non-insulin-dependent diabetes" OR TITLE_ABS:"noninsulin-dependent diabetes") AND `(TITLE_ABS:FRAX OR KW:FRAX OR TITLE_ABS:FRAXplus OR TITLE_ABS:"FRAX plus" OR TITLE_ABS:"fracture risk assessment tool" OR TITLE_ABS:"fracture risk assessment score" OR TITLE_ABS:"fracture risk assessment algorithm")`
ScienceDirect	("type 2 diabetes" OR "type 2 diabetes mellitus" OR T2DM OR T2D) AND `(FRAX OR FRAXplus OR "fracture risk assessment tool")`
OpenAlex	*Works searched in the title and abstract* `("type 2 diabetes" OR "type 2 diabetes mellitus" OR "type II diabetes" OR "type II diabetes mellitus" OR T2DM OR T2D OR NIDDM OR "non-insulin-dependent diabetes" OR "noninsulin-dependent diabetes")` AND `(FRAX OR FRAXplus OR "FRAX plus" OR "fracture risk assessment tool" OR "fracture risk assessment score" OR "fracture risk assessment algorithm")`
Cochrane Library	("type 2 diabetes" OR "type 2 diabetes mellitus" OR "type II diabetes" OR "type II diabetes mellitus" OR T2DM OR T2D OR NIDDM OR "non-insulin-dependent diabetes" OR "noninsulin-dependent diabetes") AND `(FRAX OR FRAXplus OR "FRAX plus" OR "fracture risk assessment tool" OR "fracture risk assessment score" OR "fracture risk assessment algorithm")`

6. Data Extraction and Risk of Bias Assessments

CHARMS-informed form were used to extract bibliographic details, study family, setting, design, recruitment period, eligibility criteria, participant characteristics, diabetes definition and duration, sample and event counts, follow-up duration, FRAX country model, use of femoral-neck BMD or TBS, adjustment strategy, fracture definition and ascertainment, prediction horizon, handling of censoring and competing death, missing-data methods, reported performance estimates, confidence intervals, funding, and overlap with companion reports.

When several studies referred to the same cohort, model, outcome, and horizon, we prioritized the largest eligible sample with the most complete follow-up and appropriate handling of censoring and competing mortality. We retained clinically relevant subgroup estimates separately. Any calculation from reported values was documented, and unavailable or unreliable values were recorded as not reported rather than imputed.

We used PROBAST to assess risk of bias across participants, predictors, outcome, and analysis, and applicability across participants, predictors, and outcome. Signaling questions informed domain-level judgments of low, high, or unclear risk and supporting rationales. We examined selection of participants, ascertainment of FRAX predictors and fractures, availability of predictors at the time of prediction, event numbers, missing data, follow-up, censoring, competing death, and any adjustment fitted within its evaluation cohort. One reviewer completed these judgments; independent appraisal and consensus are pending [28].

7. Data Analysis

We organized findings by independent study family, fracture outcome, FRAX country model, BMD inclusion, prediction horizon, adjustment strategy, and performance domain. Adjusted and unadjusted estimates from the same participants were compared within cohorts and were never counted as separate cohorts. We summarized whether each adjustment moved average calibration toward O:P = 1 and reported discrimination and reclassification separately.

We considered risk of bias, indirectness, inconsistency, imprecision, and selective reporting when interpreting each performance domain, as relevant to certainty of prediction-model evidence. Formal outcome-specific certainty ratings have not been completed and should not be implied by the descriptive confidence judgments in this manuscript [29].

RESULTS

1. Study Selection

The databases identified 968 records (PubMed, 113; Cochrane Library, 9; ScienceDirect, 519; OpenAlex, 211; Europe PMC, 116). 688 records were removed using database filters (170 by publication year, 430 by full-text availability, 87 by publication type, and one by journal language). From the 280 remaining records, 91 duplicates records were removed and the rest of 189 underwent title and abstract screening.and 171 records were removed based on eligibility criteria (wrong population (n=35), wrong intervention or index model (n=16), wrong outcome (n=40), and wrong study design or publication type (n=80). None were excluded for the comparator at this stage) the exclusion process can be seen in (Figure 1).

We continued to assess 18 reports at full-text screening and excluded 15: wrong population (n=3), wrong intervention or index model (n=5), wrong comparator (n=1), wrong outcome (n=2), and wrong study design or publication type (n=4). Three reports met the eligibility criteria. Majumdar et al. [20] and Leslie et al. [22] used overlapping participants from the Manitoba BMD Registry, so the three reports represent two independent study families (Figure 1).

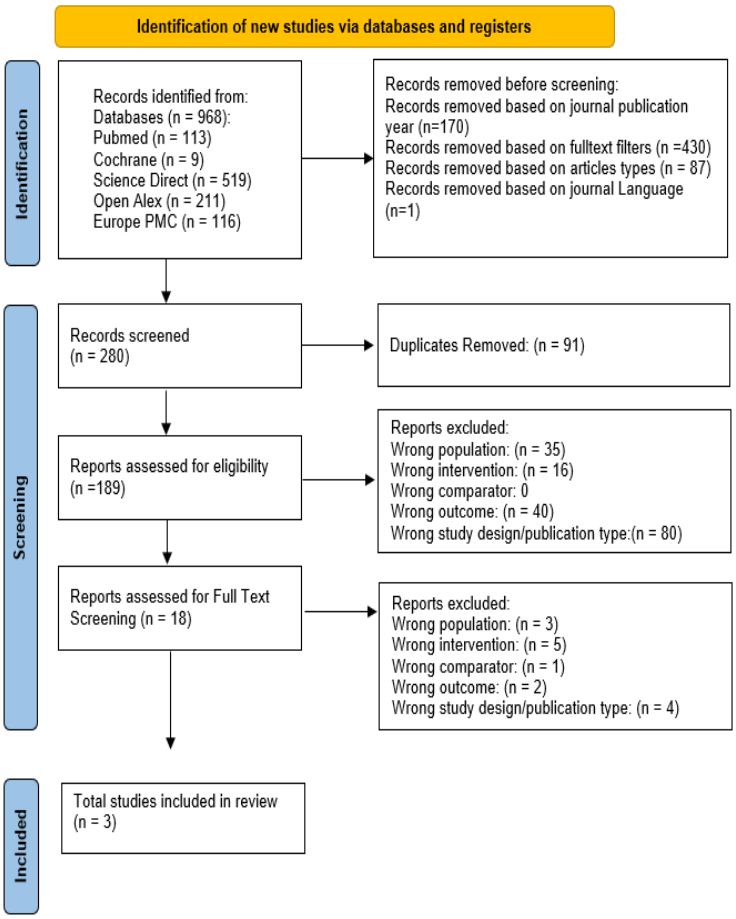

FIGURE 1: PRISMA 2020 flowchart diagram.

2. Sample Characteristics

The three reports were published from 2016 to 2022. Their reported diabetes groups sum to 14,706 report-level participants (range, 1,730–8,840), but this is neither a unique-person count nor exclusively baseline T2DM: the two Manitoba reports overlap, and Majumdar et al. [20] included 2,190 women diagnosed with diabetes after baseline DXA. Both Canadian registry analyses were retrospective longitudinal cohorts; the independent Chinese study

followed a retrospective tertiary-hospital cohort in Shanghai. Mean follow-up was 6.50 years in the Majumdar diabetes group and 8.30 years in Leslie; Kong reported a median of 10 years (Table 1) [20,22,23].

Majumdar et al. [20] classified 8,840 women with diabetes, including 6,650 with T2DM at baseline and 2,190 diagnosed after baseline DXA, and assessed Canadian FRAX with femoral-neck BMD across duration categories. The latter stratum is ancillary to baseline T2DM validation. Leslie et al. [22] studied a broader Manitoba sample of 4,136 adults with T2DM and compared unadjusted FRAX against rheumatoid arthritis (RA) proxy coding, trabecular bone score (TBS) adjustment, a reduction of 0.5 standard deviations in the entered femoral-neck T-score, and an increase of 10 years in entered age. Both reported major osteoporotic fracture (MOF) and hip fracture. Kong et al. [23] included 1,730 hospitalized Chinese adults with T2DM (66.00% men), evaluating China FRAX and selected adjustments for MOF; separate hip-fracture performance was unavailable (Table 2) [20,22,23].

TABLE 2: Characteristics of included study.

Report/ study family	Setting and population	Reported diabetes	Follow-up	Outcomes	FRAX formulation
Majumdar et al. [20] Manitoba Registry	Clinical DXA registry; women ≥40 y; Canada	8,840 (6,650 at baseline)	Mean 6.50 y in T2DM; 10-y probabilities estimated	MOF, hip	Canadian FRAX with femoral-neck BMD; duration strata
Leslie et al. [22] Manitoba Registry	Clinical DXA registry; adults ≥40 y; Canada	4,136	Mean 8.30 y; 10-y cumulative incidence	MOF, hip	FRAX with BMD; RA proxy, TBS, T-score −0.5, age +10
Kong et al. [23] Ruijin Hospital	Hospitalized adults with T2DM; Shanghai, China	1,730	Median 10 y	MOF	China FRAX ± BMD; RA proxy, age +10, T-score −0.5

Abbreviations: BMD, bone mineral density; DXA, dual-energy X-ray absorptiometry; FRAX, Fracture Risk Assessment Tool; MOF, major osteoporotic fracture; RA, rheumatoid arthritis; TBS, trabecular bone score; T2DM, type 2 diabetes mellitus; y, years. Majumdar and Leslie overlap and must not be counted as independent cohorts.

3. Risk of Bias and Applicability

All three reports were rated high overall in the PROBAST risk of bias assessment (Table 3). The Analysis domain was rated high in each report. Both Manitoba reports gave insufficient information on missing FRAX inputs. Kong et al. [23] also received a high rating in the Outcome domain because major osteoporotic fractures were ascertained partly by retrospective telephone reporting. The reported mean imputation primarily concerned biochemical variables for the separate Chinese Diabetes Fracture Risk (CDFR) model. Handling of missing FRAX inputs and the AUC method for variable follow-up were incompletely described. CDFR development was appraised separately. These are single-reviewer judgments pending independent confirmation.

Participant applicability was judged a high concern for all three reports when applying their results to the general adult T2DM population. Both Manitoba reports drew on people referred for clinical dual-energy X-ray absorptiometry, with the earlier report restricted to women. The Shanghai cohort comprised hospitalized patients from one tertiary center and excluded several major comorbidities. Predictor and outcome applicability was otherwise judged acceptable (Table 3).

Overall risk of bias was rated high if at least one domain was high, unclear if at least one domain was unclear and none was high, and low only if all domains were low; overall applicability followed the same rule across the participants, predictors and outcome domains. The Participants domain was rated unclear in Leslie et al. [22] because individuals with a body mass index outside 15–37 kg/m^2 were excluded to meet trabecular bone score requirements, an exclusion that may disproportionately affect obese adults with T2DM and whose size was not reported. The Predictors domain was rated unclear in Kong et al. [23] because ascertainment of several FRAX inputs (glucocorticoid use, rheumatoid arthritis, secondary osteoporosis and alcohol intake) was not described and femoral-neck BMD was available for only 1142 of 1730 participants. The Analysis domain was rated high in all three reports. Majumdar et al. [20] reported calibration only, without any discrimination measure, and 279 hip fractures were distributed across four diabetes-duration strata, leaving at least one stratum with fewer than 70 events. Leslie et al. [22] reported calibration and reclassification but no C-statistic or AUC for the T2DM group, as gradient-of-risk estimates were presented for the whole cohort only. Neither Manitoba report described participant exclusions or the handling of missing FRAX inputs

In Kong et al. [23], 146 of 1876 eligible patients (7.80%) without a completed telephone survey were excluded; deaths during follow-up and competing mortality were not reported, mean imputation was described only for biochemical variables (≤3% missing each), the handling of missing FRAX inputs was not specified, and the AUC method for variable follow-up was not reported. The Outcome domain was rated high in Kong et al. [23] because MOF was ascertained partly by retrospective telephone self-report, supplemented by EHR radiographic records, with possible recall bias and no reported blinding of interviewers. Participant applicability was rated high concern because all three cohorts were selected rather than general adult T2DM populations: Manitoba adults referred for clinical DXA (women aged ≥40 years only in Majumdar et al. [20]; women and men aged ≥40 years with a body mass index of 15–37 kg/m^2 in Leslie et al. [22]), and a single-center hospitalized cohort in Shanghai that excluded hyperparathyroidism, eGFR <30 mL/min/1.73 m^2, and malignancy [23].

TABLE 3: PROBAST risk-of-bias assessment.

Report	Participants	Predictors	Outcome	Analysis	Overall RoB	Overall applicability
Majumdar et al. [20]	Low	Low	Low	High	High	High concern
Leslie et al. [22]	Unclear	Low	Low	High	High	High concern
Kong et al. [23]	Low	Unclear	High	High	High	High concern

Abbreviations: PROBAST, Prediction Model Risk of Bias Assessment Tool; RoB, risk of bias.

4. Calibration of Unadjusted FRAX

Majumdar et al. [20] reported a higher MOF O:P ratio with longer baseline diabetes duration; hip-fracture ratios did not increase monotonically. For MOF, observed-to-predicted (O:P) ratios were 1.07 (95% CI 0.90 to 1.24) for T2DM duration below five years, 1.13 (95% CI 0.94 to 1.33) for five to ten years, and 1.24 (95% CI 1.08 to 1.39) for more than ten years. The corresponding hip-fracture ratios were consistently larger: 1.79 (95% CI 1.30 to 2.27), 1.46 (95% CI 0.98 to 1.95), and 1.93 (95% CI 1.50 to 2.35). After adjustment for FRAX with BMD, T2DM duration of more than ten years remained associated with MOF (hazard ratio [HR] 1.47, 95% CI 1.30 to 1.66) and hip fracture (HR 2.10, 95% CI 1.71 to 2.59) [20].

In the later Manitoba comparison study, 386 MOFs and 139 hip fractures occurred among 4,136 adults with T2DM. The estimated 10-year observed risks were 11.90% and 4.50%, compared with mean unadjusted FRAX predictions of 10.30% and 2.50%. Observed-to-predicted (O:P) ratios were 1.15 (95% confidence interval [CI], 1.03 to 1.28) for MOF and 1.85 (95% CI, 1.51 to 2.20) for hip fracture. Diabetes remained associated with fracture after adjustment for FRAX: hazard ratios were 1.32 (95% CI, 1.19 to 1.46) for MOF and 1.76 (95% CI, 1.47 to 2.10) for hip fracture. These estimates complement, but are not independent of, the earlier Manitoba duration analysis [22].

In Shanghai, 109 of 1,730 participants experienced an MOF over a median of 10 years; the reported observed risk was 6.30% (95% CI, 5.20% to 7.40%). Unadjusted China FRAX without BMD underestimated MOF, with an observed-to-predicted (O:P) ratio of 2.62 (95% CI, 2.17 to 3.08) and a calibration χ^2 statistic of 134.50 ($p<0.001$). The study's supplementary Table S3 reported a BMD-based subgroup (n=1,142): unadjusted FRAX also underestimated MOF in this subgroup (O:P ratio, 2.42; 95% CI, 2.08 to 2.96; calibration χ^2=37.12, $p<0.001$; Table 3) [23].

5. Discrimination of Unadjusted FRAX

Only Kong et al. [23] reported discrimination for eligible FRAX evaluations. For 10-year MOF, China FRAX without BMD had an area under the receiver operating characteristic curve (AUC) of 0.75 (95% CI, 0.70 to 0.79; n=1,730). In the smaller BMD-based subgroup, unadjusted FRAX had an AUC of 0.75 (95% CI, 0.70 to 0.81; n=1,142). These different sample sizes preclude interpreting their close AUC values as a direct within-person comparison of BMD addition. Neither Manitoba report provided an AUC or C-statistic (Table 3) [23].

6. Calibration After Diabetes-Specific Adjustment

In Leslie et al. [22], the four modifications shifted average MOF calibration toward an O:P ratio of 1.00. Relative to unadjusted FRAX (1.15; 95% CI, 1.03 to 1.28), the ratio was 0.91 (95% CI, 0.81 to 1.01) with the RA proxy, 1.02 (95% CI, 0.91 to 1.12) with TBS adjustment, 0.96 (95% CI, 0.86 to 1.06) after lowering the entered femoral-neck T-score by 0.5 standard deviations, and 0.93 (95% CI, 0.83 to 1.03) after adding 10 years to entered age. The observed 10-year MOF risk was 11.90%; corresponding predicted risks were 13.10%, 11.70%, 12.40%, and 12.90%, respectively [22].

Hip-fracture underprediction decreased but persisted at the point-estimate level. The unadjusted O:P ratio of 1.85 became 1.35 (95% CI, 1.10 to 1.60) with the RA proxy, 1.63 (95% CI, 1.33 to 1.93) with TBS adjustment, and 1.22 (95% CI, 0.99 to 1.44) with either the 0.5-standard-deviation T-score reduction or the 10-year age increase. The last two confidence intervals included 1.00. Residual diabetes-associated hip-fracture hazard ratios were 1.30 (95% CI, 1.09

to 1.55), 1.48 (95% CI, 1.24 to 1.77), 1.15 (95% CI, 0.96 to 1.38), and 1.12 (95% CI, 0.94 to 1.34), respectively [22].

In the Shanghai cohort, the rheumatoid arthritis (RA) proxy and 10-year age increase each reduced the MOF O:P ratio for FRAX without BMD from 2.62 to 2.10 (95% CI, 1.73 to 2.47); calibration χ^2 values were 63.30 and 63.70, respectively (both p<0.001). In the BMD subgroup (n=1,142), the unadjusted O:P ratio of 2.42 decreased to 1.80 (95% CI, 1.49 to 2.11) with the RA proxy, 2.17 (95% CI, 1.79 to 2.55) with age increased by 10 years, and 1.91 (95% CI, 1.58 to 2.24) after lowering the entered femoral-neck T-score by 0.5 standard deviations. All point estimates and confidence intervals indicated persisting average underprediction (Supplementary Table S3; Table 4) [23].

7. Discrimination and Reclassification After Adjustment

The Manitoba comparison did not provide AUC or C-statistic estimates. Relative to unadjusted FRAX, categorical net reclassification improvement (NRI) for MOF was 5.60% with the RA proxy, 3.90% with TBS adjustment, 5.60% with the T-score reduction (p<0.001), and 4.40% with the age increase. The study reported significance at the 1% level for the RA proxy and TBS adjustments and at the 5% level for the age increase. Gains among fracture cases were partly offset by downward classification among noncases. Hip-fracture NRI ranged from 1.50% to 4.70% and was not statistically significant for any modification [22].

In Shanghai, among participants assessed without BMD, adjusted AUCs were 0.75 (95% CI, 0.71 to 0.80) with the RA proxy and 0.76 (95% CI, 0.70 to 0.79) with the age increase, compared with 0.75 for unadjusted FRAX; reported comparisons were not statistically significant (p=0.365 and p=0.264). In the BMD subgroup, corresponding AUCs were 0.76 (95% CI, 0.70 to 0.81; p=0.532), 0.77 (95% CI, 0.72 to 0.82; p=0.197), and 0.76 (95% CI, 0.71 to 0.81; p=0.504) for the RA, age, and T-score adjustments, versus 0.75 unadjusted. For the full cohort without BMD, the RA proxy yielded a net reclassification improvement (NRI) of 4.88% (p=0.046) and integrated discrimination improvement (IDI) of 0.63% (p<0.001); age increase yielded an NRI of 4.21% (p=0.063) and IDI of 0.61% (p<0.001). These small changes coexisted with residual underprediction (Supplementary Table S2; Table 3) [23].

Kong et al. [23] additionally developed the CDFR model. Its apparent AUC was 0.80 (95% CI, 0.76 to 0.84); bootstrap optimism correction gave a C-statistic of 0.79. Within the development cohort, the reported O:P ratio was 0.93 (95% CI, 0.76 to 1.09), with calibration χ^2=4.63 (p=0.860). These are model-development and internal-validation findings, not independent external validation of a FRAX adjustment (Table 4) [23].

TABLE 4: Principal calibration and discrimination estimates.

Report / outcome	Model	BMD	O:P ratio (95% CI)	AUC (95% CI)	Interpretation
Majumdar et al. [20], MOF	Unadjusted; T2DM >10 y	Yes	1.24 (1.08 to 1.39)	NR	Underprediction with long duration
Majumdar et al. [20], hip	Unadjusted; T2DM >10 y	Yes	1.93 (1.50 to 2.35)	NR	Marked underprediction
Leslie et al. [22], MOF	Unadjusted	Yes	1.15 (1.03 to 1.28)	NR	Mild underprediction
Leslie et al. [22], hip	Unadjusted	Yes	1.85 (1.51 to 2.20)	NR	Marked underprediction
Leslie et al. [22], MOF	RA / TBS / T-score / age	Yes	RA 0.91 (0.81 to 1.01) TBS 1.02(0.91 to 1.12) T-score 0.96 (0.86 to 1.06) Age 0.93 (0.83 to 1.03)	NR	All moved calibration toward 1.00
Leslie et al. [22], hip	RA / TBS / T-score / age	Yes	RA 1.35 (1.10 to 1.60) TBS 1.63 (1.33 to 1.93) T-score 1.22 (0.99 to 1.44) Age 1.22 (0.99 to 1.44)	NR	Residual underprediction; least with T-score or age
Kong et al. [23], MOF	Unadjusted	No	2.62 (2.17 to 3.08)	0.75 (0.70 to 0.79)	Severe underprediction
Kong et al. [23], MOF	RA / age +10	No	RA 2.10 (1.73 to 2.47) Age 2.10 (1.73 to 2.47)	RA 0.75 (0.71 to 0.80) Age 0.76 (0.70 to 0.79)	Partial calibration improvement; little AUC change

Report / outcome	Model	BMD	O:P ratio (95% CI)	AUC (95% CI)	Interpretation
Kong et al. [23], MOF (n=1,142)	Unadjusted	Yes	2.42 (2.08 to 2.96)	0.75 (0.70 to 0.81)	Underprediction with BMD
Kong et al. [23], MOF (n=1,142)	RA proxy	Yes	1.80 (1.49 to 2.11)	0.76 (0.70 to 0.81)	Underprediction persists
Kong et al. [23], MOF (n=1,142)	Age +10 years	Yes	2.17 (1.79 to 2.55)	0.77 (0.72 to 0.82)	Underprediction persists
Kong et al. [23], MOF (n=1,142)	T-score −0.5 SD	Yes	1.91 (1.58 to 2.24)	0.76 (0.71 to 0.81)	Underprediction persists

AUC, area under the receiver-operating-characteristic curve; BMD, bone mineral density; CI, confidence interval; FRAX, Fracture Risk Assessment Tool; MOF, major osteoporotic fracture; NR, not reported; O:P, observed-to-predicted; RA, rheumatoid arthritis proxy; SD, standard deviation; TBS, trabecular bone score; T2DM, type 2 diabetes mellitus; y, years. Kong BMD results are derived from the 1,142-person subgroup in Supplementary Tables S2–S3 [23].

DISCUSSION

Principal findings

Three reports from two independent study families met the eligibility criteria. Unadjusted FRAX underestimated average fracture probability in the populations evaluated, although the magnitude depended on outcome and setting. In the Manitoba comparison, the observed-to-predicted (O:P) ratio was 1.15 for major osteoporotic fracture (MOF) and 1.85 for hip fracture; the earlier, overlapping registry report showed that the MOF ratio rose with longer diabetes duration. In the Shanghai hospital cohort, FRAX calculated without bone mineral density (BMD) yielded a MOF O:P ratio of 2.62. These results identify underprediction in the studied cohorts, not a single correction suitable for all adults with type 2 diabetes mellitus (T2DM) [20,22,23].

All four approaches assessed in Manitoba moved average MOF calibration closer to an O:P ratio of 1.00. Hip-fracture ratios remained above 1.00 at the point-estimate level, although the intervals for age and T-score adjustments included 1.00. In Shanghai, coding rheumatoid arthritis or increasing entered age reduced MOF underprediction without eliminating it. Only the Shanghai study reported an AUC for eligible FRAX evaluations: 0.75 for unadjusted FRAX, with no statistically significant difference after either tested adjustment. Evidence for discrimination and performance outside the original settings is consequently much thinner than the calibration evidence.

Interpretation of calibration and discrimination

Calibration and discrimination serve different clinical purposes. A model discriminates when it ranks people who later fracture above those who do not. Calibration asks whether its stated probabilities correspond to observed risk. Because treatment decisions often use an absolute FRAX probability, a model may rank patients reasonably well and still place some below a decision threshold if it underpredicts average risk. The Shanghai cohort illustrates this distinction: the proxies increased predicted MOF risk and reduced the O:P ratio from 2.62 to 2.10, while AUC changed little.

An O:P ratio assesses average calibration across a cohort. It does not show whether the model is accurate at low risk, at high risk, or near a treatment threshold, and a cohort average near 1.00 can conceal opposing errors in patient subgroups. The included reports did not provide a complete set of calibration intercepts, slopes, and risk-stratified plots suitable for assessing those patterns. Differences in fracture ascertainment, loss to follow-up, and treatment of death as a competing event can also affect observed risk. We therefore interpret changes in O:P as evidence about population-average agreement, not proof that an adjusted probability is accurate for an individual patient.

The difference between the Manitoba and Shanghai findings also raises the question of whether an adjustment should correct the overall level of risk or the relationship between individual predictors and outcome. A systematic shift in baseline risk may be addressed by recalibration, whereas errors that vary across predicted risk require closer examination of the calibration slope and curve. Van Calster et al. [19] emphasize that acceptable discrimination does not guarantee reliable probabilities when a model is transferred to a new population. Applied to this review, the different O:P ratios support testing local recalibration rather than assuming that one fixed diabetes proxy will transport unchanged. However, the available aggregate estimates cannot identify the appropriate updating method, and any updated model would require evaluation in data separate from those used to estimate the correction.

Diabetes-related FRAX adjustments

The four Manitoba approaches modify different parts of the risk calculation. Rheumatoid arthritis serves as a proxy clinical risk factor; it does not mean the patient has rheumatoid arthritis, and this option cannot represent both conditions in someone who has them simultaneously. Adding 10 years to entered age or lowering the femoral-neck T-score by 0.5 standard deviations changes existing predictors. TBS-adjusted FRAX uses an additional measure of trabecular texture. These approaches cannot be assumed to deliver equivalent corrections even when their average MOF ratios approach 1.00. In Manitoba,

MOF O:P ratios ranged from 0.91 to 1.02 across adjustments, whereas hip-fracture ratios ranged from 1.22 to 1.63 [22].

External consistency was limited. In Shanghai, the RA and age proxies each yielded an MOF O:P ratio of 2.10 without BMD. The supplementary BMD-based results also showed underprediction after all tested modifications (O:P, 1.80 to 2.17). This differs from the near-unit average MOF ratios in Manitoba. Country-specific FRAX models, BMD availability, underlying fracture rates, and DXA registry versus hospital selection could contribute; the two cohorts cannot isolate any one source of heterogeneity. Within the Shanghai BMD subgroup, RA produced the smallest O:P ratio, but this does not establish a generally preferred correction [22,23].

Diabetes duration offers another possible source of variation. The earlier Manitoba report found MOF O:P ratios of 1.07 for T2DM of less than five years and 1.24 for more than ten years; hip-fracture ratios were 1.79 and 1.93 in those respective groups. The intermediate-duration hip estimate was lower, so the hip pattern was not monotonic. Duration may represent cumulative metabolic exposure, complications, treatment history, frailty, or fall risk rather than a single biological mechanism. Its incremental predictive value and the size of any duration-specific adjustment require validation outside Manitoba [20].

Kong et al. [23] also developed a Chinese Diabetes Fracture Risk (CDFR) model. Its apparent AUC was 0.80, declining to an optimism-corrected C-statistic of 0.79 after internal validation. The reported calibration was measured within the development cohort. Because CDFR is a separate model, these figures should not be presented as externally validated evidence for a FRAX adjustment; independent testing is needed before clinical comparison [23].

Earlier evidence helps explain the rationale for including TBS. In a Manitoba registry study of 29,407 women, including 2,356 with diabetes, diabetes was associated with higher BMD but lower lumbar spine TBS; TBS predicted major osteoporotic fractures independently of BMD in women with diabetes [21]. This supports the possibility that lumbar spine texture captures fracture-related information omitted by areal BMD. It does not establish that TBS-adjusted FRAX is fully calibrated for both MOF and hip fracture. The residual hip-fracture underprediction in the included adjustment comparison illustrates that distinction [22]. The earlier study provides contextual evidence from the same registry setting, rather than an additional independent validation cohort, and its publication preceded this review's eligibility window.

Outcome and setting differences

In Manitoba, hip-fracture underprediction was larger than MOF underprediction in both overlapping reports. The Shanghai study did not report separate hip-fracture model performance, so the pattern has not been independently reproduced for hip fracture. Separate reporting matters because composite MOF and hip fracture differ in event rates, ascertainment, the influence of falls, and competing mortality. An adjustment that corrects average MOF risk should not be assumed to correct hip-fracture risk.

The cohorts also address different clinical populations. Both Manitoba reports relied on DXA referrals and BMD-based Canadian FRAX, with women exclusively in the earlier analysis. The Shanghai study included mostly men admitted to a tertiary hospital and reported comparisons both without BMD (n=1,730) and in a smaller BMD subgroup (n=1,142). These features restrict transportability. Agreement between the Manitoba papers provides complementary analyses of one registry, not replication in an independent cohort. The larger Shanghai O:P ratios cannot be attributed specifically to country, BMD use, sex, or care setting with these data.

Relation to previous evidence

Earlier cohort studies found that adults with T2DM have higher fracture risk than people without diabetes at comparable femoral-neck T-scores or FRAX probabilities [1,16]. Our results extend that observation by showing how reported model performance changes across fracture outcomes and candidate corrections. Preserved or relatively high areal BMD may coexist with impaired bone material properties, altered microarchitecture, and greater fall risk, although the included validation studies cannot establish which pathways explain their calibration errors [2,3,9,17].

Direct measurements of bone material properties offer a further explanation for the mismatch between BMD and fracture risk. Farr et al. [30] compared 30 postmenopausal women with T2DM lasting more than ten years with 30 age-matched controls. Tibial microindentation showed lower bone material strength in the diabetes group, and the difference persisted after adjustment for BMI and other measured covariates. This finding supports the biological plausibility of skeletal fragility that is incompletely represented by BMD-based risk assessment. However, this small cross-sectional study neither evaluated prospective FRAX performance nor established the magnitude of a probability correction. Measurements at the tibia also cannot directly explain the different MOF and hip-fracture calibration errors in our included cohorts. Its contribution is therefore mechanistic context rather than validation evidence.

The present synthesis also shows why several estimates from the same registry cannot substitute for independent replication. We treated the Manitoba duration analysis and adjustment comparison as related reports, and we distinguished the newly developed CDFR model from FRAX validation. Both choices limit the amount of independent evidence but make the interpretation of adjustment performance more reliable.

Clinical interpretation

Clinicians should interpret an unadjusted FRAX probability cautiously when evaluating fracture risk in a patient with T2DM, particularly if the estimate lies near a decision threshold or if long diabetes duration, falls, complications, or prior fractures suggest additional risk. The model and outcome being considered matter: the Manitoba findings for FRAX with BMD and hip fracture cannot simply be transferred to MOF prediction without BMD in a different population. FRAX can inform a broader clinical assessment alongside BMD, falls, diabetes history, and applicable local treatment guidance [3,17].

A proxy adjustment may be useful for exploring whether a plausible increase in predicted risk would change clinical discussion. However, an adjusted probability should not be treated as a validated estimate for every setting. Using rheumatoid arthritis as a proxy also requires care if rheumatoid arthritis is present. No included study tested whether choosing treatment through an adjusted FRAX score improves treatment uptake, avoids overtreatment, or reduces fractures. Nor did these studies establish a T2DM-specific treatment threshold. Those decisions require clinical judgment until comparative validation and impact evidence become available.

The clinical importance of identifying underestimated risk is supported by evidence that osteoporosis treatment can remain effective in people with diabetes. An individual-participant analysis of 15 randomized trials included 96,385 participants, of whom 6,539 had T2DM. Pooled analyses found no statistically significant interaction between diabetes status and the fracture-reduction effects of antiresorptive therapy; results were similar when restricted to bisphosphonate trials [31]. These findings support considering fracture prevention when otherwise indicated, but do not establish which FRAX adjustment should determine treatment eligibility. The analysis was post hoc, participants with diabetes were almost entirely women, and some individual-drug findings differed from the pooled results. Treatment efficacy, the accuracy of an adjusted risk estimate, and the choice of a treatment threshold therefore require separate evaluation.

Study limitations and Future Directions

Across three reports representing two independent country registry database, unadjusted FRAX underestimated average fracture probability in the studied populations with type 2 diabetes mellitus. Four adjustment strategies improved MOF calibration in the Manitoba registry, but hip-fracture ratios remained above 1.00 at the point-estimate level. In the independent Shanghai cohort, the tested rheumatoid arthritis and age proxies reduced, but did not resolve, major-fracture underprediction. Discrimination was reported for only the Shanghai cohort, where the adjustments did not significantly change the area under the curve.

These data do not establish a single adjustment that is reliable across fracture outcomes, FRAX formulations, or clinical settings. Clinicians should interpret FRAX alongside diabetes duration, falls, complications, bone mineral density, and other relevant patient factors rather than assume that an adjusted probability is exact. Independent external validations should report major osteoporotic and hip fractures separately, align prediction with follow-up, and assess calibration and discrimination across the risk range. The small, partly overlapping evidence base limits stronger recommendations. Future adequate, multicentre randomized trials are needed to directly compare the best FRAX adjustment to predict major osteoporotic and hip fracture incidence in adults, especially with type 2 diabetes.

Meta-analysis required at least three independent, clinically comparable cohorts with extractable estimates for the same outcome, model formulation, and aligned prediction horizon. No analysis set met this prespecified minimum because of the two Manitoba reports overlap, and the independent Chinese cohort differs in setting and has no separate hip-fracture model performance. We therefore did not calculate pooled performance, heterogeneity statistics, prediction intervals, or small-study-effect tests.

Independent validation should cover community and clinical populations, women and men, different country-specific FRAX models, and patients across diabetes duration and complication profiles. Investigators should report FRAX with and without BMD separately and provide distinct MOF and hip-fracture outcomes. Prediction and observation horizons must be aligned, and the method used to account for competing death should be explicit.

Future reports should present observed and predicted risk, an O:P ratio or calibration-in-the-large with uncertainty, calibration intercept and slope where appropriate, and plots across the range of predicted risk. AUC or time-dependent C-statistics should include confidence intervals and prespecified comparisons between unadjusted and adjusted models. Reporting missing predictor values and the method used to handle them is essential. Decision-curve analyses could test whether improved calibration offers net benefit at locally relevant thresholds [24,27].

Where data permit, local recalibration or model updating using duration, complications, falls, or treatment exposures may be more useful than a single fixed proxy. Development studies must avoid relying on apparent performance alone and should be followed by external validation. Prospective impact studies can then test whether any improvement in model performance changes treatment decisions and patient outcomes.

Clinical utility should also be tested directly rather than inferred from improved O:P ratios or reclassification alone. Decision-curve analysis evaluates net benefit across threshold probabilities

by weighting false-positive classifications against true-positive classifications according to the relative consequences of unnecessary intervention and missed events [32]. For diabetes-related FRAX adjustments, a useful validation would compare unadjusted FRAX, each prespecified adjustment, and strategies of treating everyone or no one across thresholds relevant to the intended healthcare setting. Such analyses would need methods appropriate to the prediction horizon, censoring, and competing mortality. They could show whether correcting average underprediction improves treatment selection enough to justify additional intervention. Even a favorable decision curve would still require subsequent impact evaluation to establish changes in care and fracture outcomes.

DECLARATIONS

Availability of data and materials: All data analyzed in this review were extracted from published studies.

Use of artificial intelligence: ChatGPT (OpenAI) assisted with identifying relevant literature, helping determining search term for database literature searching, and formatting the journal manuscript. The authors take responsibility for the content of this journal manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: MFR and RW. Methodology: MFR and RRAP. Data extraction: MFR, RW, ZN, and RRAP. Risk of bias assessment: MFR, RRAP, and ANA. Writing-original draft: MFR and ZN. Writing-review and editing: MFR and ANA. Supervision: ANA and RRAP.

REFERENCES

[1] Schwartz AV, Vittinghoff E, Bauer DC, Hillier TA, Strotmeyer ES, Ensrud KE, et al. Association of BMD and FRAX score with risk of fracture in older adults with type 2 diabetes. JAMA. 2011;305(21):2184-92. doi:10.1001/jama.2011.715

[2] Walsh JS, Vilaca T. Obesity, type 2 diabetes and bone in adults. Calcif Tissue Int. 2017;100 (5):528-35. doi:10.1007/s00223-016-0229-0

[3] Ferrari SL, Abrahamsen B, Napoli N, Akesson K, Chandran M, Eastell R, et al. Diagnosis and management of bone fragility in diabetes: an emerging challenge. Osteoporos Int. 2018;29 (12):2585-96. doi:10.1007/s00198-018-4650-2

[4] Janghorbani M, Van Dam RM, Willett WC, Hu FB. Systematic review of type 1 and type 2 diabetes mellitus and risk of fracture. Am J Epidemiol. 2007;166(5):495-505. doi:10.1093/aje/kwm106

[5] Vestergaard P. Discrepancies in bone mineral density and fracture risk in patients with type 1 and type 2 diabetes - a meta-analysis. Osteoporos Int. 2007;18(4):427-44. doi:10.1007/s00198-006-0253-4

[6] Moayeri A, Mohamadpour M, Mousavi SF, Shirzadpour E, Mohamadpour S, Amraei M. Fracture risk in patients with type 2 diabetes mellitus and possible risk factors: a systematic review and meta-analysis. Ther Clin Risk Manag. 2017;13:455-68. doi:10.2147/TCRM.S131945

[7] Vilaca T, Schini M, Harnan S, Sutton A, Poku E, Allen IE, et al. The risk of hip and non-vertebral fractures in type 1 and type 2 diabetes: a systematic review and meta-analysis update. Bone. 2020;137:115457. doi:10.1016/j.bone.2020.115457

[8] Koromani F, Oei L, Shevroja E, Trajanoska K, Schoufour J, Muka T, et al. Vertebral fractures in individuals with type 2 diabetes: more than skeletal complications alone. Diabetes Care. 2020;43(1):137-44. doi:10.2337/dc19-0925

[9] Dede AD, Tournis S, Dontas I, Trovas G. Type 2 diabetes mellitus and fracture risk. Metabolism. 2014;63(12):1480-90. doi:10.1016/j.metabol.2014.09.002

[10] Napoli N, Chandran M, Pierroz DD, Abrahamsen B, Schwartz AV, Ferrari SL; IOF Bone and Diabetes Working Group. Mechanisms of diabetes mellitus-induced bone fragility. Nat Rev Endocrinol. 2017;13 (4):208-19. doi:10.1038/nrendo.2016.153

[11] Hofbauer LC, Busse B, Eastell R, Ferrari S, Frost M, Müller R, et al. Bone fragility in diabetes: novel concepts and clinical implications. Lancet Diabetes Endocrinol. 2022;10(3):207-20. doi:10.1016/S2213-8587(21)00347-8

[12] Kanis JA, Johansson H, Harvey NC, McCloskey EV. A brief history of FRAX. Arch Osteoporos. 2018;13(1):118. doi:10.1007/s11657-018-0510-0

[13] Compston JE, McClung MR, Leslie WD. Osteoporosis. Lancet. 2019;393(10169):364-76. doi:10.1016/S0140-6736(18)32112-3

[14] Kanis JA, Cooper C, Rizzoli R, Reginster JY; Scientific Advisory Board of the European Society for Clinical and Economic Aspects of Osteoporosis (ESCEO) and the Committees of Scientific Advisors and National Societies of the International Osteoporosis Foundation (IOF). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int. 2019;30(1):3-44. doi:10.1007/s00198-018-4704-5

[15] 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):385-97. doi:10.1007/s00198-007-0543-5

[16] Giangregorio LM, Leslie WD, Lix LM, Johansson H, Oden A, McCloskey E, et al. FRAX underestimates fracture risk in patients with diabetes. J Bone Miner Res. 2012;27(2):301-8. doi:10.1002/jbmr.556

[17] Agarwal A, Leslie WD. Fracture prediction tools in diabetes. Curr Opin Endocrinol Diabetes Obes. 2022;29(4):326-32. doi:10.1097/MED.0000000000000734

[18] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38. doi:10.1097/EDE.0b013e3181c30fb2

[19] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7

[20] Majumdar SR, Leslie WD, Lix LM, Morin SN, Johansson H, Oden A, et al. Longer duration of diabetes strongly impacts fracture risk assessment: the Manitoba BMD cohort. J Clin Endocrinol Metab. 2016;101(11):4489-96. doi:10.1210/jc.2016-2569

[21] Leslie WD, Aubry-Rozier B, Lamy O, Hans D; Manitoba Bone Density Program. TBS (trabecular bone score) and diabetes-related fracture risk. J Clin Endocrinol Metab. 2013;98(2):602-9. doi:10.1210/jc.2012-3118

[22] Leslie WD, Johansson H, McCloskey EV, Harvey NC, Kanis JA, Hans D. Comparison of methods for improving fracture risk assessment in diabetes: the Manitoba BMD Registry. J Bone Miner Res. 2018;33(11): 1923-30. doi:10.1002/jbmr.3538

[23] Kong XK, Zhao ZY, Zhang D, Xie R, Sun LH, Zhao HY, et al. Major osteoporosis fracture prediction in type 2 diabetes: a derivation and comparison study. Osteoporos Int. 2022;33 (9):1957-67. doi:10.1007/s00198-022-06425-8

[24] Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744

[25] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71. doi:10.1136/bmj.n71

[26] Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z

[27] Snell KIE, Levis B, Damen JAA, Dhiman P, Debray TPA, Hooft L, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ. 2023;381: e073538. doi:10.1136/bmj-2022-073538

[28] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8. doi:10.7326/M18-1376

[29] Foroutan F, Guyatt G, Trivella M, Kreuzberger N, Skoetz N, Riley RD, et al. GRADE concept paper 2: concepts for judging certainty on the calibration of prognostic models in a body of validation studies. J Clin Epidemiol. 2022;143: 202-11. doi:10.1016/j.jclinepi.2021.11.024

[30] Farr JN, Drake MT, Amin S, Melton LJ 3rd, McCready LK, Khosla S. In vivo assessment of bone quality in postmenopausal women with type 2 diabetes. J Bone Miner Res. 2014;29 (4):787-95. doi:10.1002/jbmr.2106

[31] Eastell R, Vittinghoff E, Lui LY, Ewing SK, Schwartz AV, Bauer DC, et al. Diabetes mellitus and the benefit of antiresorptive therapy on fracture risk. J Bone Miner Res. 2022;37(11):2121-31. doi:10.1002/jbmr.4697

[32] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74. doi:10.1177/0272989X06295361